Scalable Computing:
Practice and Experience

Scientific International Journal
for Parallel and Distributed Computing

ISSN: 1895-1767

Universitatea de Vest din Timişoara

Volume 23(4) December 2022

SUBSCRIPTION INFORMATION: please visit `http://www.scpe.org`

Scalable Computing: Practice and Experience

Volume 23, Number 4, December 2022

TABLE OF CONTENTS

Special Issue:

Regular papers:

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 129–145, DOI 10.12694/scpe.v23i4.1987

MAP-REDUCE BASED DISTANCE WEIGHTED K-NEAREST NEIGHBOR MACHINE LEARNING ALGORITHM FOR BIG DATA APPLICATIONS

E. GOTHAI[*], V. MUTHUKUMARAN[†], K. VALARMATHI[‡], SATHISHKUMAR V E[§], THILLAIARASU.N[¶] AND KARTHIKEYAN P[‖]

Abstract. With the evolution of Internet standards and advancements in various Internet and mobile technologies, especially since web 4.0, more and more web and mobile applications emerge such as e-commerce, social networks, online gaming applications and Internet of Things based applications. Due to the deployment and concurrent access of these applications on the Internet and mobile devices, the amount of data and the kind of data generated increases exponentially and the new era of Big Data has come into existence. Presently available data structures and data analyzing algorithms are not capable to handle such Big Data. Hence, there is a need for scalable, flexible, parallel and intelligent data analyzing algorithms to handle and analyze the complex massive data. In this article, we have proposed a novel distributed supervised machine learning algorithm based on the MapReduce programming model and Distance Weighted k-Nearest Neighbor algorithm called MR-DWkNN to process and analyze the Big Data in the Hadoop cluster environment. The proposed distributed algorithm is based on supervised learning performs both regression tasks as well as classification tasks on large-volume of Big Data applications. Three performance metrics, such as Root Mean Squared Error (RMSE), Determination coefficient (R2) for regression task, and Accuracy for classification tasks are utilized for the performance measure of the proposed MR-DWkNN algorithm. The extensive experimental results shows that there is an average increase of 3% to 4.5% prediction and classification performances as compared to standard distributed k-NN algorithm and a considerable decrease of Root Mean Squared Error (RMSE) with good parallelism characteristics of scalability and speedup thus, proves its effectiveness in Big Data predictive and classification applications.

Key words: Machine Learning, Big Data Analytics, MapReduce Programming, k-Nearest Neighbour, Classification, prediction

AMS subject classifications. 94A16, 68T05

1. Introduction. Now a days, all recent applications, like credit rankings, pattern recognition, location oriented Geographical Information Systems (GIS), all facilities management, recommendation systems, computer vision, smart-cities services and so on are in need of well-organized processing of queries using bigdata methodologies. One of the best query retrieval method is Nearest Neighbor Query. When we consider two points are given, A as Query and B as Training, NN finds out the nearest neighbors of B for every position of A. This form of query is very much helpful in reality.

For example, kNN can be used in recommendation systems since it helps to find public with analogous qualities. It can also be used in an online video streaming platform, for example, to suggest a content that a user is more possible to outlook based on what other users look at. For image categorization, the KNN algorithm can also be used. It is important in a variety of computer vision applications since it can group similar data points jointly, such as cats and dogs in split classes. Additional applications of the kNN include classification, graph-based computational learning etc. when the datasets concerned are moderately small, kNN be able to

[*]Associate Professor, Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, Erode 638060, Tamilnadu, India.(egothai@kongu.ac.in)

[†]Assistant professor, Department of mathematics, School of Applied Sciences, REVA University, Bangalore-560064.(muthu.v2404@gmail.com)

[‡]Professor, Panimalar Engineering College, Bangalore Trunk Road, Varadharajapuram, Poonamallee, Chennai-600123.(valarmathi_1970@yahoo.co.in)

[§]Post-Doctoral Researcher, Department of Industrial Engineering, Hanyang University, Seoul, Republic of Korea.(srisathishkumarve@gmail.com)

[¶]Assistant professor, School of Computing and Information Technology, REVA University, Bangalore-560064.(thillai888@gmail.com)

[‖]Associate Professor, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.(mailbox7576@gmail.com)

be capably answer in a central background. This is commonly not the reason when we are working with huge quantity of data, considering that we live in the epoch of WWW and in the mobile technology. There is huge quantity of data coming out through GIS devices, sensor networks, geotagged tweets, scientific devices, etc. and their handing out is very challenging, thus we come out with the mapping of parallel and distributed computing environment is very much essential to get results in a reasonable amount of time.

The k-Nearest Neighbors (kNN) algorithm [1, 2] is one of the top 10 supervised machine learning algorithms that perform classification as well as regression analysis in Big data applications. Due to its non-parametric nature, easy implementation, and effectiveness, it becomes an important part of the machine learning domain and also it has an inherent feature of parallel implementation in Map Reduce environment. Due to this, the kNN and Map reduce framework found applications in various diverse fields like pattern classification [3], image classification [4, 5], big data classification [6], automated world wide web usage mining [7] and document classification [8, 9, 10, 11, 12].

The contributions of this paper include:

- A distributed and parallel Distance Weighted k-Nearest Neighbor model has been proposed to analyze big data
- Hadoop MapReduce framework cluster has been established to improve execution efficiency and scalability of the proposed system
- Different block size of data in the underlying HDFS chosen to handle large datasets with high efficiency rate.
- Several experiments have been designed and executed to show the speedup and scale-up of the proposed distributed algorithm
- Classification accuracy of the proposed system measured and analyzed for different size of k-Nearest neighbors
- The performance of the proposed system is evaluated on the authentic, standard and reliable dataset.

The rest of this paper is organized as follows. Section two provides Literature survey and Section three provides an introduction to the distance weighted k-NN algorithm; we discussed the big data technology Hadoop framework and the architecture of the proposed MR-DWkNN model with the Hadoop implementation algorithm in section four. The experimental setup such as Hadoop computational cluster and its components, various benchmark datasets, and performance metrics for evaluation are described in section five. The conduct of experiments, performance analyses, and results in comparison with standard k-NN are presented in section six while section seven draws conclusions from the experimental study and suggests the directions for future work.

2. Literature Survey. There are several variants of the kNN algorithm have been proposed to handle various kinds of data and are shown to be very effective in its performance. Keller et al. [13] proposed a modified version of the kNN algorithm based on fuzzy concepts called fuzzy-kNN with three techniques to assign fuzzy memberships to the data points. Denoeux [14] proposed a method D-SkNN based on Dempster-Shafer theory to address the problem of unseen pattern classification in a dataset based on the kNN algorithm. The author demonstrated the performance of the new method D-SkNN with a real-time dataset as well as a simulated dataset and compared it with the standard kNN and majority voting methods. Kuncheva [15] specified an intuitionistic fuzzy version of the kNN rule called IF-kNN and incorporated the voting rule with assigned weights based on the membership and non-membership to a certain category of class. Based on the threshold value, the vote is categorized as positive or negative. Yang et al. [16] developed an enhanced version of the kNN algorithm with a fuzzy editing rule and incorporated a few asymptotic properties into kNN. The experimental results conducted on various datasets conformed that this approach outperforms the standard kNN algorithm. Huang et al. [17] developed a modified version of kNN called DCT-kNN which is based on feature weighting and class distribution. The experiments on UCI datasets had shown a considerable classification accuracy improvement.

Liu et al. [18] introduced the kNN method for Multi-class classification problems and demonstrated its performance in the classification of Multiclass datasets. Liu and Zhang [19] designed the variant of the kNN algorithm termed as mutual nearest neighbors (MkNN) to remove the noisy data and improved the data classification accuracy. Zhang [20] incorporated the certainty factor measure to the kNN algorithm to apply it over the imbalanced class distribution dataset and the variant is called kNN-CF. The authors also demonstrated

that kNN-CF algorithm accuracy is better than the standard kNN algorithm. Shichao et al. [21] proposed a novel method called self-reconstruction to determine the k-value of the kNN algorithm for each training sample and applied on real datasets for data classification. The experimental results show that this method outperformed the standard data classification methods in terms of classification accuracy.

In recent years, the map-reduce based distributed and parallel machine learning algorithms are the focus of the research community. The MapReduce framework has emerged as a powerful, robust, and distributed parallel programming model [22, 23, 24, 25] provides a solution with good performance and efficient execution to large-scale data analytic applications including data mining, web page access ranking, graph analysis, image classification and bioinformatics [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44]. Kolb et al. [45] investigated the use of the MapReduce programming model for parallel entity resolution with automated data partitioning on real-world datasets.

The application of the kNN algorithm and its variants in the context of big data for classification and regression has been already considered. Triguero et al. [46] suggested a method for handling big data in kNN classification. They proposed a novel partitioning method based on the Map-Reduce framework and distributed the functioning of algorithms to multiple nodes in the cluster environment without any loss of classification accuracy. The performance of this method was tested with 5.7 million instances on Poker hand dataset, obtained an accuracy of 0.5171 for k=3 and the results show that kNN is a suitable algorithm to handle big data.

Ding et al. [47] proposed a clustering-based approach for processing large high-dimensional datasets. The authors added the Principal Component Analysis for dimensionality reduction in addition to the kNN classification algorithm in the processing of large datasets. Deng et al. [48] introduced the kNN algorithm in big data applications for classifications. The authors applied the k-means clustering algorithm on a large size dataset, as a result, it is split into several subsets. Finally, they applied the kNN method, its variants RC-kNN and LC-kNN, classified the samples in each subset of the dataset with an accuracy of 72.21%, 83.89% and 86.35% on MNIST dataset. In [49] the authors propose a new distributed and parallel kNN join operation on large real and synthetic datasets in the Map-Reduce platform. They demonstrated the scalability and efficiency of the method with hundreds of millions of records. In [50], the authors developed a cost-effective MapReduce-based k-Nearest Neighbor (MR-kNN) algorithm for Big Data classification and tested for different k-values against the Pokerhand dataset. The classification accuracy was about 0.5386 for k=7. In [51], an iterative Hadoop MapReduce method called iHMR-kNN was developed for kNN based classification and analysis of the image dataset. Singh, A.P. et al [52] employed kNN in deep learning architectures for image classification which shows a higher classification accuracy. In [53], a label driven latent subspace learning for multi-view image classification model was developed which produces an improved classification result.

3. Distance Weighted k-Nearest Neighbor (DWkNN) Algorithm. k-Nearest Neighbor algorithm is a lazy learner and nonparametric method which does not rely on building a model during the training phase, and whose classification rule is based on a given similarity function between the training samples and the query (testing) sample to be classified. Unlike the existing model-based classification algorithms (building a model with a given training dataset and predicting any test samples with the built model), the kNN algorithm needs to keep all the training examples in memory, to search for all the K nearest neighbors for a test sample.

Let $TR = x_1, \cdots, x_n$ be the training dataset with n instances and m attributes. All the instances are class labeled data points $(x_i, c_j), j = 1, \cdots, n_c$ the corresponding class labels of the instances. Here $n > n_c$. In the k-NN algorithm, the learned target function may be either categorical (discrete-valued) or continuous-valued (real-valued) function. Given a query instance , from the collections of query instances dataset TQ, its unknown class $c^{'}$ is determined as follows.

Step 1: Compute the Euclidean distance between the query instances to all instances in TR. Let the instance x_i in TR is described as a feature vector $(a_1(x_i), a_2(x_i), \cdots, a_m(x_i))$ where $a_z(x_i)$ denotes the z^{th} attribute value of instance x_i.

Let $d(x_i, x_q)$ is the Euclidean distance between the query instance x_q and the instance x_i in the dataset TR and is computed as per equation (3.1)

$$d(x_i, x_q) = \sqrt{\sum_{z=1}^{m} (a_z(x_i) - a_z(x_q))^2} \tag{3.1}$$

Step 2: Sort the instances in the dataset TR in ascending order of their Euclidean distances as given in equation (3.2), n is the total data points.

$$sort(TR < (d(x_i, x_q) >)), 1 \leq i \leq n \tag{3.2}$$

Step 3(a): For discrete-valued learning target function:
The general form of a discrete-valued learning function is $f(x) = TR^m \rightarrow V$, where V is the finite set $v_1, \cdots, v_s$, here s is the number of classes.
Let $x_1, \cdots, x_k$ denote the k instances from TR that are nearest to the query instance x_q.
Return the class label for the query instance x_q as given in equation (3.3).

$$f^{'}(x_q) \leftarrow \sum_{i=1}^{k} \delta(\upsilon, f(x_q)) \tag{3.3}$$

where $\delta(x, y) = 1$ if $x = y$ and where $(x, y) = 0$ otherwise.

Step 3(b): For real-valued target function:
The general form of the function is $f : TR^m -> TR$
The target attribute value for the query instance $x_q$is computed as given in equation (3.4)

$$f^{'}(x_q) \leftarrow \frac{\sum_{i=1}^{k} f(x_i)}{k} \tag{3.4}$$

The above steps illustrate the standard kNN algorithm and this can be enhanced with the inclusion of weight function as given in equation (5),

Step 4(a): Select the k-nearest neighbors from the dataset TR Let $TR^{'} = x_1, \cdots, x_k$ be the k nearest instances. Assign a weight w_i to i^{th} nearest neighbor of the query x_q using the distance-weighted function as given in equation (3.5)

$$w_i = \left\{ \begin{array}{ll} \frac{d(x_q,x_k)-d(x_q,x_1)}{d(x_q,x_k)-d(x_q,x_1)}, & if d(xq, xk) \neq d(x_q, x_1) \\ 1 & if d(x_q, x_k) \neq d(x_q, x_1) \end{array} \right\} \tag{3.5}$$

Based on the majority voting, assign the class label c_j to the query instance x_q of discrete-valued function as given in equation

$$f^{'}(x_q) \leftarrow argmax_{\upsilon \epsilon V} \sum_{i=1}^{k} W_i(\upsilon, f(x_q)) \tag{3.6}$$

Step 4(b): For real-valued target functions, the target attribute value for the query instance x_q is computed as given in equation

$$f^{'}(x_q) = \frac{\sum_{i=1}^{k} w_i f(x_i)}{\sum_{i=1}^{k} w_i} \tag{3.7}$$

4. Hadoop Implementation of our proposed algorithm.

4.1. Hadoop Distributed Framework. Hadoop is an Apache project, initially introduced in 2007 as a free and open-source framework with two major components Hadoop Distributed File System (HDFS) and MapReduce (MR) programming model. It becomes very popular in recent years because of its simplicity, built-in fault tolerance, scalability on data-intensive tasks, parallel execution of tasks, and capable of running on a Hadoop cluster with 1000s of commodity hardware machines. Hadoop framework itself takes care of job scheduling, job execution, controlling of all underlying tasks in the cluster, and other runtime management tasks.

MapReduce is a software framework that enables the application developers to explore all the options of parallel procedures with Map and Reduce functions. The MapReduce programming model supports the developers to write and execute the applications upon a cluster consists of a few hundred to several thousand commodity hardware machines. The three major classes of the MapReduce program are Master/Driver, Mapper, and Reducer. The Master class is responsible for setting various execution parameters to the MapReduce job to run in the Hadoop cluster. The parameters are names of Mapper and Reducer classes, data types, and job names to be executed. The MapReduce framework operates on (key, value) pairs. Each Map task process an input split (block) generating intermediate data of (key, value) format. Then, they are sorted and partitioned by key, so later at the Reduce phase, pairs of the same key will be aggregated to the same reducer for further processing [54, 55].

5. MapReduce Architecture of proposed Distance Weighted kNN Algorithm. The main purpose of the supervised machine learning technique is to facilitate an algorithm to learn from the previous historical data/events and extracts the knowledge from the events. The extracted knowledge is represented as an intelligent mathematical model that can be used to make predictions or classifications on the given new scenario/future events. In most general terms, the machine learning model consists of two phases, the training phase, and the testing phase. Fig. 5.1 depicts the Map-Reduce architecture of the proposed Distance weighted kNN algorithm and its implementation.

Algorithm 1 provides the details of the Map function operation of Distance weighted kNN. The input training dataset TR with m samples and n features is split into s partitions/blocks as $p1, p2, \cdots, ps$ in the distributed file system of the Hadoop cluster. Each partition takes one block storage size of HDFS (64MB/128MB/256MB) as set initially in the Hadoop file system. Each partition contains m/s samples, distributed uniformly with a default replication factor of 3. The dataset TQ contains the query instances and since the K-Nearest Neighbor is a lazy learner, the pattern matching process initiated only on the submission of query instance xq to the system. For each block of the input file, a separate map task is created and each map task computes the Euclidean distance between the query instance xq and all the instances in the block. The Euclidean distance that is computed for all the blocks is converted into <key, value> pair as <instance_id, (Euclidean distance, attributes)> and these are stored in HDFS as intermediate results.

The reducer part of the MapReduce program collects the intermediate results generated by all the map tasks and sorts the instances on Euclidean distance from smallest to largest. After sorting instances, k-nearest samples are selected and weights are assigned to the samples based on Euclidean distance using the distance weighted function as in equation (3.5). Finally, the reducer finds the class label for the query instance xq if the target attribute is a discrete-valued function as given in equation (3.6) and computes the target attribute value for the real-valued target function as given in equation (3.7). The detailed operation of the Reduce function is given in algorithm 2. The Map-reduce architecture of standard k-NN is the same as Fig 5.1 with the exclusion of distance weighted function as given in equation (3.5) and is called an MR-SDkNN algorithm.

Algorithm 1 Map function of Distance Weighted kNN

```
Procedure DWkNN-MAP (TR, TQ, k))
Input: TR-> Training dataset with n instances and m attributes
TQ-> Dataset with query instances
k-> number of neighbors
key-> instance_id,
value -> <Euclidean instance, instance features>

Output: Class label/target attribute value for the query instance x_q and performance
metrics of the proposed methodology.
for a given query instance x_q in TQ
for t=1 to size(TR) do
ED(x_q,x_i )<-compute Euclidean distance (TR(x_i ),x_q )
result<-(<key:instance_id(t) ;value:(ED(x_q,x_i ),instance)>
```

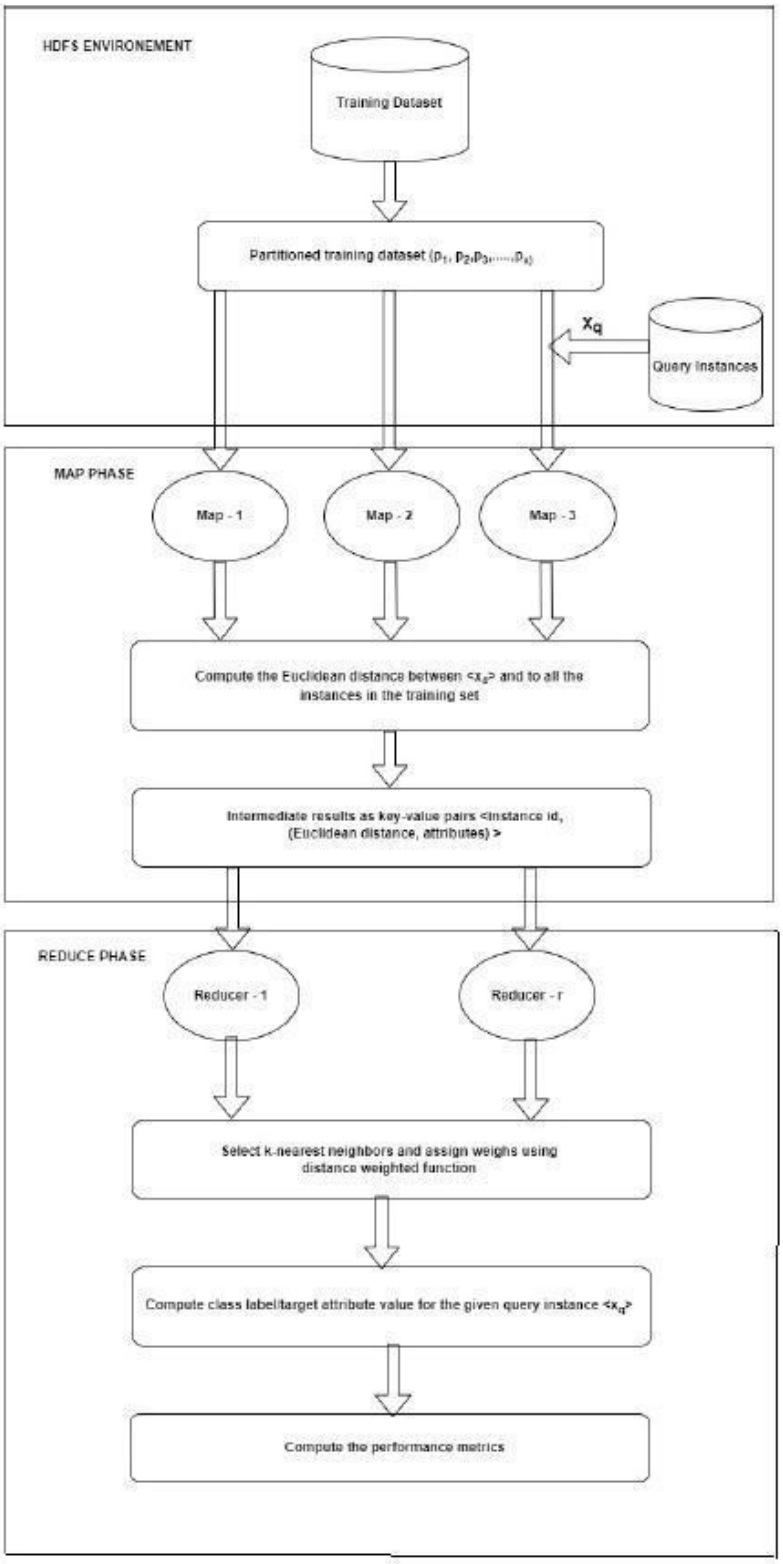


Fig. 5.1: Proposed Distance Weighted kNN Algorithm (MR-DWkNN)

```
context.write(t,ED(x_q,x_i ))
end for
end procedure
-----------------------------------------------------------------------
-----------------------------------------------------------------------
Algorithm 2 Reduce function of Distance Weighted kNN
-----------------------------------------------------------------------
Procedure DWkNN-REDUCE (key, value)
Key: instance_id
Value: Euclidean distance, Attributes
For all k-nearest instances of TR do the following
Compute w_i as given in equation (5)
Assign the distance-weighted value to all k-instances
If the target function is discrete-valued then
Find the class-label of x_q using equation (6)
Compute the performance metric classification Accuracy
Context.write(x_q,  <class label, accuracy>)
else
Find the target attribute value of x_q using equation (7)
Compute the performance metrics RMSE and Determination coefficient (R2)
Context.write(x_q, (class label, RMSE, Determination coefficient (R2)))
endif
end for
```

6. Experimental Setup. This section describes the Multi-node Hadoop cluster configured for this research work (Sec 5.1), specifications of the datasets chosen (Sec 5.2), and the various performance metrics employed to measure the classification and regression task performances. (Sec 5.3).

6.1. Hadoop cluster environment. The various experiments designed for this proposed research work are executed on a Multi-node Hadoop cluster established with 25 physical machines. One machine is designated as a master node (Name node) and configured to run Hadoop services and the remaining machines are configured as the worker nodes (data node). The configuration of all the physical machines is Core i5 four-core Processor, 2.1GHz clock speed, 16 GB RAM, 6MB Cache, 1 TB HDD with 1 Gbps network card.

The specific details of the software used in the cluster and Hadoop environment configuration parameters are the following:

Hadoop framework version 2.9.0
Operating system: Ubuntu Linux
18.04.03 LTS 64-bit
Replication factor: 3
HDFS size: 64 MB/128MB
Virtual memory for the map and reduce task: 8 GB

The total cores in the cluster are 50, enabling the hardware level hyper-threading features, the cores in the cluster become 100.

6.2. Datasets. In this research work, we will use six benchmark large-size datasets from the UCI machine learning repository. Among the six datasets, three datasets with real-valued target attributes used for the regression task (Table 6.1), and the remaining three datasets with discrete-valued target attributes are used for the classification task (Table 6.2). We tabulate the number of attributes (attributes), Data type of attributes (Data types), number of instances (instances), file size (File size), and the year of publication (year). In our experimental work, all the datasets are partitioned into training and test dataset using an n^{th} fold cross-validation technique.

Table 6.1: Summary description of the datasets with discrete-valued target attributes

S.No	Dataset	#Attributes	#DataTypes	# instances	#File size	#year
1	Super conductivity Data	81	Multivariate	21,263	26.8MB	2018
2	Year Prediction MSD (Subset of Million song Dataset)	90	Multivariate	515,345	433MB	2011
3	Wave Energy Converters	49	Multivariate	288,000	123MB	2019

Table 6.2: Summary description of the datasets with discrete-valued target attributes

S.No	Dataset	#Attributes	#DataTypes	# instances	#File size	#classes	#year
1	Higgs	28	Multivariate	11,000,000	7.4GB	2	2014
2	Susy	18	Multivariate	5,000,000	2.22GB	2	2014
3	Pokerhand	11	Multivariate	1,025,000	23.4 MB	10	2007

6.3. Performance measures. In our research work, the performance of the proposed MR-DWkNN algorithm is assessed with the following four metrics.

Root Mean Square Error (RMSE): The error value of the regression model is measured in terms of RMSE. This metric shows the square root of the quadratic mean of the differences between the predicted and expected values of the target attribute. RMSE is computed as given in equation (6.1)

$$RootMeanSquareError(RMSE) = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{6.1}$$

Determination coefficient (R2): This metric analyzes how the differences in one variable (x) can be explained by a difference in another variable (y). This measure evaluates how a model approximates the real data points. The higher R2, the more efficient is the prediction model and its value usually between 0 and 1 is computed using equation (6.2).

$$Determinationcoefficient(R2) = 1 - \frac{\sum_{i=1}^{N}(x_i - yi)^2}{\sum_{i=1}^{N}(x_i - x)^2} \tag{6.2}$$

Accuracy: The performance of classification model is measured using the metric accuracy. It is defined as the ratio of the number of correctly predicted samples to the total number of input samples in the dataset as given in equation (6.3).

$$Accuracy = \frac{Numberofcorrectlypredictedsamples}{TotalNumberofinputsamples} \tag{6.3}$$

These are three commonly used performances metric for measuring the performance of the standard predictors/classifiers

Scalability: This defines the capability of an algorithm in a parallel computing cluster environment to enhance both in terms of the number of processing cores/nodes in the cluster and the number of instances in the dataset. It can also be defined as the capability of an s-times larger computing system in the cluster environment to execute an s-times larger computational task in the same given job execution time as the original system and this can be expressed as in equation (6.4).

$$Scalability(s, TD) = \frac{JET(1, TD)}{JET(s, sTD)} \tag{6.4}$$

Table 7.1: Performance metrics of a Classification task

Dataset	No. of Instances	No. of Blocks	k=3		k=5	
			MR-SDkNN	MR-DWkNN	MR-SDkNN	MR-DWkNN
Higgs	1,100,000	12/12	0.67382	0.69782	0.68688	0.71065
Susy	1,000,000	8/8	0.76361	0.79416	0.78754	0.80965
Pokerhand	1,025,000	1/1	0.54784	0.55329	0.55064	0.56548

Tasks

where 's' is the number of cores/nodes in the cluster environment, JET(1, TD) is the job execution time on one core/node with data size of TD, JET(s, sTD) is the job execution time of the parallel tasks with 's' cores/nodes in the cluster environment with data size s times of TD. Ideal parallelism shows a constant scale up with an increasing number of computing cores in the cluster and dataset size [56-57]

Speedup: Speedup is also one of the measures employed in evaluating the performance of the parallel algorithms and is used to enhance the job execution time in the cluster. It is defined as the ratio of the time of sequential execution to the time of parallel execution. Speedup can be expressed as in equation (6.5).

$$Speedup(s, TD) = \frac{JET(1)}{JET(s)} \tag{6.5}$$

where 's' is the number of cores in the cluster environment, JET(1) is the job execution time on one core with data size of TD, JET(s) is the job execution time of the parallel tasks with 's' cores in the cluster environment with the same data size TD.

7. Conduct of experiments and discussion of results. In this section, we evaluate the proposed supervised MR-DWkNN algorithm on six public datasets in the UCI Machine Learning repository. Here we described the four experiments conducted and compare the results collected from these experiments with the MR-SDkNN algorithm.

1. In the first two experiments, the MR-DWkNN method was executed against the three datasets under the regression category and another three datasets under the classification category. The performance metrics are compared with the standard MR-SDkNN method as given in section 5.1 and 5.2 respectively.
2. Third, the scalability performance of the MR-DWkNN model on regression and classification datasets are analyzed, reported in section 5.3
3. Finally, the performance of our proposed method for different k-values was tested and the performance metrics are described in section 5.4.

7.1. Performance of MR-DWkNN model with MR-SDkNN model. In this experiment, the HDFS block size is set as 64MB, and 10% instances from the Higgs dataset, 20% from Susy dataset, and the entire Pokerhand dataset are chosen. Initially, we run the parallel version of the standard kNN algorithm MR-SDkNN over all the six datasets which is used for comparison with its variants. To do this, 20% of the instances are chosen randomly from each dataset as query instances (TQ) and the remaining 80% of the instances are considered as Training instances (TR). The performance metrics root mean square error (RMSE) and Determination coefficient (R2) are recorded for the regression task and accuracy is recorded for the classification task.

Afterward, we executed the MR-DWkNN algorithm over all the three datasets under the classification category. In this, the entire input dataset file is split into HDFS block size of equal-sized files and contains an equal number of samples. The samples in each file have been distributed evenly and studied the performance of the proposed MR-DWkNN algorithm. Initially, the Map-Reduce model is trained with 80% of the training samples. Table 7.1 shows the classification accuracy of MR-SDkNN and MR-DWkNN on three benchmark datasets for two different k-values. Similarly, the regression performance of MR-SDkNN and MR-DWkNN is tabulated in Table 7.2. From the above two tables, we can conclude that,

1. In the case of the classification task, the proposed MR-DWkNN algorithm produces an increase of classification accuracy in the range of 1.5% to 3.5% for all three datasets as compared with the standard MR-SDkNN.

Table 7.2: Performance metrics of a Regression task

Dataset	No. of Blocks /Map	MR-SDkNN (k=3)		MR-DWkNN (k=3)		MR-SDkNN (k=5)		MR-DWkNN (k=5)	
		RMSE	Det.Coe (R2)	RMSE	Det.Coe (R2)	RMSE	Det.Coe (R2)	RMSE	Det.Coe (R2)
Super conductivity Data	Tasks 1/1	63.65784	0.89327	52.63257	0.90676	55.75436	0.89878	52.63257	0.91564
Year Prediction MSD	7/7	364.73621	0.86213	325.63850	0.88126	343.65785	0.86899	307.64545	0.89675
Wave Energy Converters	2/2	245.71917	0.90234	232.67719	0.91452	241.43650	0.90768	212.65432	0.92456

Table 7.3: Scalability performance of MR-SDkNN and MR-DWkNN

Dataset	#instances	#cores used	#Map Tasks	Scalability	
				MR-SDkNN	MR-DWkNN
Higgs	1,100,000	12	12	1.00000	1.00000
	2,200,000	24	24	0.98650	0.97964
	4,400,000	48	48	0.93256	0.92546
	8,800,000	96	96	0.89756	0.87659
Poker Hand	1,025,000	1	1	1.00000	1.00000
	2,050,000	2	2	0.98082	0.96702
	4,100,000	4	4	0.97125	0.94576
	8,200,000	8	8	0.94366	0.91342
	16,400,000	16	16	0.92453	0.87905
	32,800,000	32	32	0.89786	0.84704
Year Prediction MSD	515,345	7	7	1.00000	1.00000
	1,030,690	14	14	0.97908	0.95409
	2,061,380	28	28	0.96880	0.94378
	4,122,760	56	56	0.94788	0.91876
Wave Energy Converters	288,000	1	1	1.00000	1.00000
	576,000	2	2	0.98761	0.97699
	1,152,000	4	4	0.96034	0.95578
	2,304,000	8	8	0.95979	0.93456
	4,608,000	16	16	0.93435	0.90671
	9,216,000	32	32	0.87692	0.85674

2. In the case of regression task, the MR-DWkNN algorithm produces an increase of Determination coefficient (R2) in the range of 1.5% to 3.2%, and consequently, there is a fall of RMSE in the range of 1.5% to 2.5% for all three datasets as compared with the standard MR-SDkNN.

7.2. Scalability, dataset split and distribution. To demonstrate how well both the MR-SDkNN and MR-DWkNN scales up, two datasets under each category were chosen. The scalability experiments were performed where the instances of the dataset/size of the dataset were increased in proportion to the number of cores. Table 7.3 summarizes the datasets used for the scalability experiment, number of instances, computing processor cores, the number of map tasks, and the scalability results obtained for both MR-SDkNN and MR-DWkNN algorithms. From the Table 7.3, it is observed that the scalability of both MR-SDkNN and MR-DWkNN decreases slowly when the size of the dataset increases, and the number of processor cores used for the computation increases.

Fig 7.1 shows the performance results of our proposed model on Higgs datasets. The experiment was conducted with a size of 1.1 million, 2.2 million, 4.4 million, and million 8.8 million instances on 12, 24, 48, and

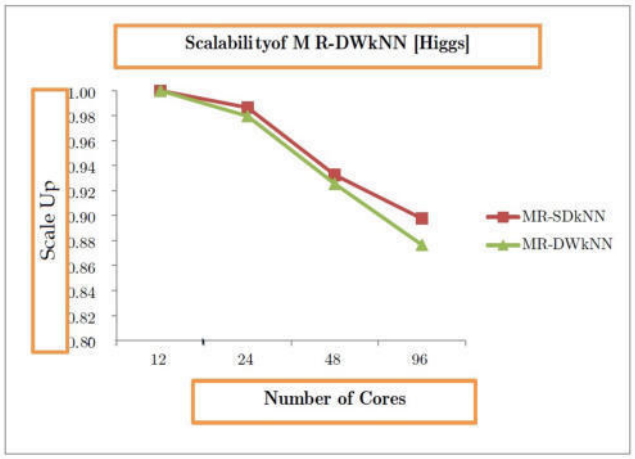


Fig. 7.1: Scalability of MR-SDkNN and MR-DWkNN [Higgs Dataset]

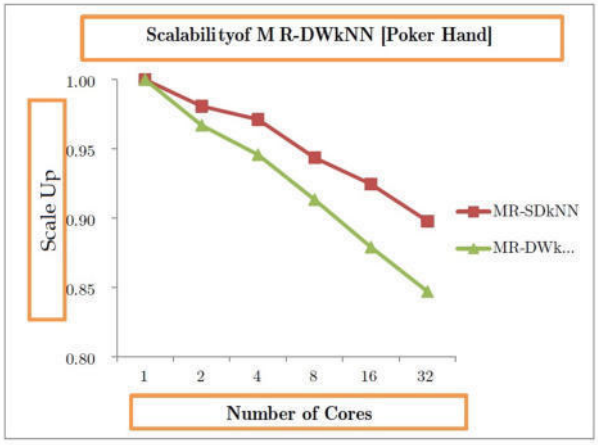


Fig. 7.2: Scalability of MR-SDkNN and MR-DWkNN [Poker Hand Dataset]

96 cores respectively. Since the initial size of the dataset with 1.1 million instances is 740 MB, it occupies about 12 HDFS blocks, hence 12 cores have been chosen initially for its execution. Afterward, the size of the cluster has been doubled. It is observed that the scalability of both MR-SDkNN and MR-DWkNN decreases slowly when the size of the dataset increases and it maintains a value of scale-up higher than 89.76% for MR-SDkNN and 87.65% for MR-DWkNN.

The scalability performance of MR-SDkNN and MR-DWkNN algorithms on the Poker Hand dataset is shown in Fig. 7.2. The initial size of the dataset is 23.4 MB with 1.025 million instances occupies one HDFS block of size 64MB. Initially, one core in the cluster is used to execute the proposed model, and subsequently, the experiment was conducted with a dataset size of 2.05 million, 4.1 million, 8.2 million, 16.4 million, and million 32.8 million instances on 2, 4, 8,16, and 32 cores respectively. From this experiment, it is observed that the scalability of both MR-SDkNN and MR-DWkNN decreases slowly when the size of the dataset increases, and it maintains a value of scale-up higher than 89.76% and 84.7% respectively.

Fig. 7.3 shows the scalability performance results of MR-SDkNN and MR-DWkNN models on year prediction MSD datasets. Initially, the experiment was conducted with 0.515 million instances and the dataset size is 433MB. The dataset is split into 7 HDFS blocks and for each block, one map task is created. Hence, a total

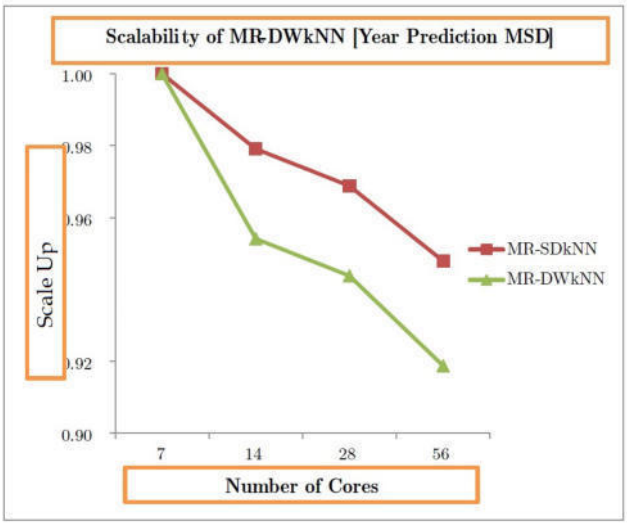


Fig. 7.3: Scalability of MR-SDkNN and MR-DWkNN [Year Prediction MSDDataset]

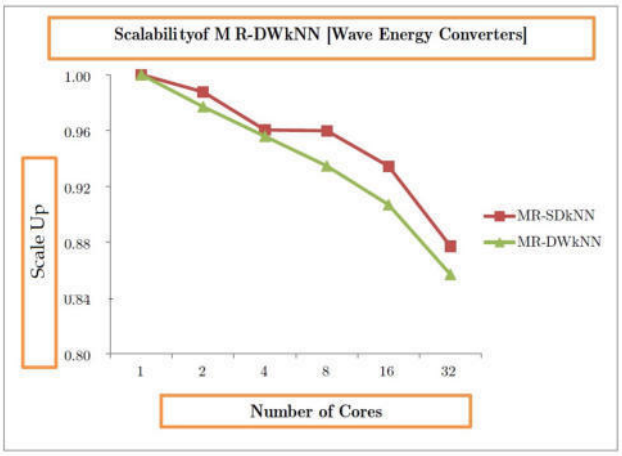


Fig. 7.4: Scalability of MR-SDkNN and MR-DWkNN [Wave Energy Converters Dataset]

of 7 map tasks are created and a cluster of size 7 cores are used for its execution. Afterward, the size of the dataset and the number of cores in the cluster has been increased proportionately. The results show that both MR-SDkNN and MR-DWkNN scale up to higher than 94.78% and 91.87% respectively.

Fig. 7.4 shows the scalability performances of our proposed MapReduce-based kNN versions on the Wave energy converters dataset. The initial size of the dataset is 123 MB with 0.288 million instances occupies one HDFS block of size 128 MB. For this experiment, the HDFS block size is configured as 128 MB. Initially, one core in the cluster is used to execute the proposed model, and subsequently, the experiment was conducted with a dataset size of 0.576 million, 1.152 million, 2.304 million, 4.608 million, and 9.216 million instances on 2, 4, 8, 16 and 32 cores respectively. From this experiment, it is observed that the scalability of both MR-SDkNN and MR-DWkNN decreases gradually when the size of the dataset and number of cores for its execution increases, and it maintains a value of scale-up higher than 87.69% and 85.67% respectively. From these experiments, we can conclude that,

Table 7.4: Speedup performance of MR-SDkNN and MR-DWkNN

Dataset	#instances	#cores used	#Dataset Splits	Speedup (x times)	
				MR-SDkNN	MR-DWkNN
Higgs	1,100,000	5	5	3.66	3.56
		10	10	7.22	6.92
		15	15	10.55	10.31
		20	20	13.64	13.42
Poker Hand	2,050,000	1	1	1.00	1.00
		2	2	1.56	1.50
		4	4	3.01	2.97
		8	8	5.86	5.54
		16	16	10.91	10.59
		32	32	21.70	20.54
Year Prediction MSD	515,345	5	5	3.66	3.56
		10	10	7.12	6.62
		15	15	10.35	9.86
		20	20	13.24	12.82
Wave Energy Converters	576,000	1	1	1.00	1.00
		2	2	1.56	1.48
		4	4	2.97	2.89
		8	8	5.78	5.38
		16	16	10.59	10.11
		32	32	20.22	19.58

- Hadoop clusters can handle large volumes of datasets and provide as many processor cores as required for the execution of our proposed algorithm.
- Both MR-SDkNN and MR-DWkNN able to scale up when the size of the dataset and computing processor cores increases.

7.3. Speedup. To measure the speedup performance of both the MR-SDkNN and MR-DWkNN algorithms, two datasets under each category were chosen. The speedup experiments were performed where the number of dataset splits was increased in proportion to the number of computing cores in the cluster with a fixed size of the dataset. Table 7.4 summarizes the datasets used for the speedup experiments, number of instances, computing processor cores, the number of map tasks, and the speedup (x times) results obtained for both MR-SDkNN and MR-DWkNN algorithms. It is observed that the speedup of both MR-SDkNN and MR-DWkNN increases in proportion to the number of computing processor cores enabled in the cluster. From these experiments, we can conclude that,

- Both MR-SDkNN and MR-DWkNN able to speedup when the computing processor cores increases for a fixed size dataset
- The execution efficiency of the proposed algorithms in the cluster improves proportionally with the size of the cluster

7.4. Influence of neighborhood size k on performance metrics and comparison. To investigate the influence of neighborhood size k on performance metrics, we have chosen all the six datasets under two different categories. The performance of the proposed distributed algorithms is measured in a cluster with an HDFS block size of 64MB and the interval of neighborhood size k ranges from 3 to 31. Table 7.5 shows the influences of neighborhood size k on the classification accuracies on three benchmark classification datasets. It is observed that the classification performance of the MR-DWkNN classifier is better than the MR-SDkNN classifier with increasing neighborhood size k. However, after a certain value of k, the classification accuracy starts decreasing on all three datasets. The best classification accuracy of each classifier on the benchmark data sets is shown in bold-faces against the corresponding k-value. The values in the parenthesis represent the corresponding dataset. Table 7.6 shows the influences of neighborhood size k on the regression task and its performance metric determination coefficient (R^2). The Determination coefficient (R^2) shows that how close

Table 7.5: Influence of neighborhood size k on the classification task

K-Value	Accuracy (Higgs)		Accuracy (Susy)		Accuracy (Poker Hand)	
	MR-SDkNN	MR-DWkNN	MR-SDkNN	MR-DWkNN	MR-SDkNN	MR-DWkNN
3	0.67382	0.69782	0.76361	0.79416	0.54784	0.55329
7	0.68688	0.71065	0.78754	0.80965	0.55064	0.56548
11	0.70214	0.74654	0.82435	0.83324	0.56462	0.57900
15	0.71124	0.76428	0.85864	0.87987	0.57886	0.59809
19	0.70892	0.76002	0.87425	0.90231	0.57031	0.58971
23	0.69745	0.75462	0.86548	0.89489	0.56112	0.57602
27	0.69126	0.74532	0.84387	0.87002	0.54387	0.56043
31	0.68542	0.73457	0.80563	0.85743	0.52301	0.54387

Table 7.6: Influence of neighborhood size k on the regression task

K-Value	Determination Coefficient -R2 (Superconductivity Data)		Determination Coefficient -R2 (Year Prediction MSD)		Determination Coefficient -R2 (Wave Energy Converters)	
	MR-SDkNN	MR-DWkNN	MR-SDkNN	MR-DWkNN	MR-SDkNN	MR-DWkNN
3	0.89327	0.90676	0.86213	0.88126	0.90234	0.91452
7	0.89878	0.91564	0.86899	0.89675	0.90768	0.92456
11	0.91004	0.92764	0.88034	0.91502	0.91241	0.92650
15	0.91678	0.93013	0.89764	0.93013	0.92134	0.93613
19	0.90452	0.92648	0.91245	0.94065	0.92876	0.94261
23	0.88099	0.91438	0.90657	0.92344	0.91834	0.93744
27	0.87239	0.89723	0.89542	0.90723	0.90878	0.92372
31	0.86573	0.87564	0.87421	0.89756	0.89743	0.91076

the data instances fit with the regression line and also it measures the strength of the relationship between the distributed machine learning model constructed in the training and the response variable. The three benchmark datasets Superconductivity, Year prediction MSD, and Wave Energy converters dataset under the regression category are chosen. From the results, it is observed that the prediction performance of the MR-DWkNN method is better than the MR-SDkNN method with increasing neighborhood size k. However, after a certain value of k, the R^2 value starts decreasing on all three datasets. The best R^2 value of each classifier on the benchmark data sets is shown in bold-faces against the corresponding k-value. The values in the parenthesis represent the corresponding dataset.

Table 7.7 shows the classification accuracy of our proposed methods and compared with MR-kNN ((Maillo et al.) and MRPR – FCNN ((Triguero et al.)) methods. It is concluded that our model MR-DWkNN classifies instances with good accuracy rate for $k = 1, 3, 5$ and 7 as compared other models on Poker hand dataset.

8. Conclusions and further work. In this research work, we have developed a MapReduce based on two different versions of the kNN algorithm called MR-SDkNN based on standard k-NN algorithm and MR-DWkNN based on distance weighted k-NN algorithm. The distributed learning model is constructed with 80% of training instances and the remaining 20% instances are used as test (query) instances. Various experiments are carried out on recently published six benchmark large volume datasets from the UCI repository. The predictive and classification performance of MR-DWkNN is evaluated in terms of three metrics as Root Mean Squared Error (RMSE), Determination Coefficient (R2), and Accuracy. In addition to these, the scalability performance of the proposed algorithm was also tested on Hadoop Multi-node cluster. The results obtained from these experiments have shown that the main accomplishments of MR-DWkNN are the following:

1. There is an increase in performance such as classification accuracy and determination coefficient (R2) of the proposed MR-DWkNN as compared to the MR-SDkNN

Table 7.7: Comparison of classification accuracy

k-Value	Accuracy (Poker Hand)			
	MR-SDkNN	MR-DWkNN	MR-kNN (Maillo et al.)	MRPR – FCNN (Triguero et al.)
1	0.52318	0.53786	0.5019	0.5013
3	0.54784	0.55329	0.4959	0.5171
5	0.55098	0.55971	0.5280	-
7	0.55064	0.56548	0.5386	-

2. MR-DWkNN is a scalable approach in a multi-node cluster environment and a proven parallel approach with promising performance metrics achievement in Big Data applications.

The future work considered is the improvement of runtime execution of Map and Reduce tasks through the use of other big data handling frameworks such as Spark and Flink. In addition, the kNN may be integrated with various deep learning architectures for image analytics of multi-class classification.

REFERENCES

[1] T. Cover, P. Hart *nearest neighbor pattern classification*, IEEE Trans. Inf. Theory 13(1) (1967) 21–27.
[2] A. Mucherino, P.J. Papajorgji, P.M. Pardalos, K-nearest neighbor classification, Data Mining in Agriculture, Springer, 2009, pp.83–106.
[3] B. Yang, M. Xiang, Y. Zhang, Multi-manifold discriminant ISO map for visualization and classification, Pattern Recognit. 55 (2016) 215–230.
[4] J. Zou, W. Li, Q. Du, Sparse representation-based nearest neighbor classifiers for hyperspectral imagery, IEEE Geosci. Remote Sens. Lett. 12 (12) (2015) 2418–2422.
[5] W. Yang, C. Sun, L. Zhang, A multi-manifold discriminant analysis method for image feature extraction, Pattern Recognit. 44 (8) (2011) 1649–1657.
[6] J. Maillo, I. Triguero, F. Herrera, A MapReduce-based k-nearest neighbor approach for big data classification, IEEE TrustCom/BigDataSE/ISPA, Helsinki, Finland, Volume 2, 2015, pp. 167–172.
[7] D. Adeniyi, Z. Wei, Y. Yongquan, Automated web usage data mining, and recommendation system using k-nearest neighbor (KNN) classification method, Appl. Comput. Inform. 12 (1) (2016) 90–108.
[8] Bijalwan, V., Kumar, V., Kumari, P. & Pascual, J. (2014). KNN based machine learning approach for text and document mining. International Journal of Database Theory and Application, 7(1): 61-70.
[9] Zhang, W., Yoshida, T. & Tang, X.J. (2008). Text classification is based on multi-word with support vector machine. Knowledge-Based Systems, 21(8): 879-886.
[10] Zhang, W., Yoshida, T. & Tang, X.J. (2011). A comparative study of TF* IDF, LSI, and multi-words for text classification. Expert Systems with Applications, 38(3): 2758-2765.
[11] Chen, J.D. & Tang, X.J. (2017). The distributed representation for societal risk classification toward BBS posts. Journal of Systems Science & Complexity. DOI: 10.1007/s11424-016-5099-z.
[12] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. & Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12:2461-2505.
[13] M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nearest neighbor algorithm, IEEE Trans. Syst. Man Cybern. 4 (1985) 580–585.
[14] T. Denoeux, A k-nearest neighbor classification rule based on Dempster-Shafer theory, IEEE Trans. Syst. Man Cybern. 25(5) (1995) 804–813.
[15] L.I. Kucnehva, An intuitionistic fuzzy k-nearest neighbors rule, Notes IFS 1 (1995) 56–60.
[16] M.-S. Yang, C.-H. Chen, on the edited fuzzy k-nearest neighbor rule, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 28(3) (1998) 461–466.
[17] J. Huang, Y. Wei, J. Yi, and M. Liu, "An Improved kNN Based on Class Contribution and Feature Weighting," 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, 2018, pp. 313-316, DOI: 10.1109/ICMTMA.2018.00083.
[18] H. Liu, X. Li, S. Zhang, Learning instance correlation functions for multi-label classification, IEEE Trans. Cybern. 47 (2) (2016) 499–510.
[19] H. Liu, S. Zhang, Noisy data elimination using mutual k-nearest neighbor for classification mining, J. Syst. Softw. 85 (2012) (2012) 1067–1074.
[20] S. Zhang, KNN-CF approach: incorporating certainty factor to KNN classification, IEEE Intell. Inf. Bull. 11 (1) (2010) 24–33.
[21] Shichao Zhang, Debo Cheng, Ming Zong, Lianli Gao, Self-representation nearest neighbor search for classification, Neurocomputing, Volume 195, 2016, Pages 137-142, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2015.08.115.
[22] Hadoop. in: http://Hadoop.apache.org (Accessed on 13.01.16).
[23] Jimmy Lin, Alek Kolcz, Large-scale machine learning at Twitter, Proceedings of the 2012 ACM SIGMOD International

Conference on Management of Data, Scottsdale, AZ, USA, 2012, pp. 793–804.
[24] Fan Zhang, Junwei Cao, Samee U. Khan, Keqin. Li, Kai Hwang, A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications, Future Gener. Comput. Syst. 43–44 (2015) 149–160.
[25] Panagiotis Moutafis, George Mavrommatis, Michael Vassilakopoulos, Spyros Sioutas, Efficient processing of all-k-nearest-neighbor queries in the MapReduce programming framework,Data & Knowledge Engineering, Volume 121,2019,Pages 42-70,ISSN 0169-023X,https://doi.org/10. 1016/j.datak. 2019.04.003.
[26] Jianping Gou, Wenmo Qiu, Zhang Yi, Xiangjun Shen, Yongzhao Zhan, Weihua Ou, Locality constrained representation-based K-nearest neighbor classification, Knowledge-Based Systems,Volume 167,2019,Pages 38-52,ISSN 0950-7051, https://doi.org/10.1016 /j.knosys.2019.01.016.
[27] Niloofar Rastin, Mansoor Zolghadri Jahromi, Mohammad Taheri,A Generalized Weighted Distance k-Nearest Neighbor for Multi-label Problems, Pattern Recognition,2020, 107526,ISSN 0031-3203,https://doi.org/10.1016/j.patcog.2020.107526.
[28] Ching-Hsue Cheng, Chia-Pang Chan, Yu-Jheng Sheu, A novel purity-based k nearest neighbors imputation method and its application in financial distress prediction, Engineering Applications of Artificial Intelligence, Volume 81,2019,Pages 283-299, ISSN 0952-1976,https://doi.org/ 10.1016/j.engappai.2019.03.003.
[29] Sarah M. Ayyad, Ahmed I. Saleh, Labib M. Labib,Gene expression cancer classification using modified K-Nearest Neighbors technique,Biosystems,Volume 176,2019,Pages 41-51, ISSN 0303-2647,https://doi.org/10.1016/j.biosystems.2018.12.009.
[30] Yanhui Guo, Siming Han, Ying Li, Cuifen Zhang, Yu Bai,K-Nearest Neighbor combined with guided filter for hyperspectral image classification,Procedia Computer Science, Volume 129,2018,Pages 159-165,ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.03.066.
[31] Aparicio-Ruiz, P., Barbadilla-Martín, E., Guadix, J. et al. KNN and adaptive comfort applied in decision making for HVAC systems. Annals of Operations Research (2019). https://doi.org/ 10.1007 /s10479-019-03489-4
[32] Patel, D., Shah, Y., Thakkar, N. et al. Implementation of Artificial Intelligence Techniques for Cancer Detection. Augment Hum Res 5, 6 (2020). https://doi.org/10.1007/s41133-019-0024-3
[33] Yadav, D.C., Pal, S. Thyroid prediction using ensemble data mining techniques. Int. j. inf. tecnol. (2019). https://doi.org/10.1007/s41870-019-00395-7
[34] Owen O'Malley, Arun C. Murthy, Yahoo! Winning a 60 Second Dash with a Yellow Elephant. http://sortbenchmark.org/Yahoo2009.pdf (April 2009) (Accessed on 22.02.16).
[35] Eric Anderson, Joseph Tucek, Efficiency matters!, ACM SIGPOS Oper. Syst. Rev. 44 (1) (2010) 40–45.
[36] F.N. Aftari, J.D. Ullman, Optimizing joins in a MapReduce environment, Proceedings of the 13th International Conference on Extending DatabaseTechnology, EDBT, 2010, pp. 99–110.
[37] Low, Y., Bickson, D., Gonzalez, J., Guestrin, C., Kyrola, A., Hellerstein, J.M, Distributed GraphLab: a framework for machine learning and data mining in the cloud. Proc. VLDB Endow. 5, 716–727 (2012).
[38] Cormack, G.V., Smucker, M.D., Clarke, C.L., Efficient and effective spam filtering and re-ranking for large web datasets. Inf. Retr. 14, 441–465 (2011).
[39] Lin, J.: Brute force and indexed approaches to pairwise document similarity comparisons with MapReduce, Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.155–162 (2009).
[40] Zhao, W., Ma, H., He, Q.: Parallel k-means clustering based on MapReduce, Cloud Computing, pp. 674–679. Springer, Berlin (2009).
[41] Baraglia, R., De Francisci Morales, G., Lucchese, C., Document similarity self-join with MapReduce. In: 2010 IEEE 10th International Conference on Data Mining (ICDM), pp. 731–736 (2010).
[42] Caruana, G., Li, M., Liu, Y.: An ontology enhanced parallel SVM for scalable spam filter training. Neurocomputing 108, 45–57 (2013).
[43] Liao, R., Zhang, Y., Guan, J., Zhou, S.: CloudNMF: a MapReduce implementation of nonnegative matrix factorization for large-scale biological datasets. Genomics Proteomics Bioinforma. 12, 48–51 (2014).
[44] Svendsen, M., Tirthapura, S.: Mining maximal cliques from a large graph using MapReduce: tackling highly uneven sub-problem sizes. J. Parallel Distrib. Comput. 79, 104–114 (2012).
[45] Kolb, L., Thor, A. & Rahm, E. Multi-pass sorted neighborhood blocking with MapReduce. Comput Sci Res Dev 27, 45–63 (2012). https://doi.org/10.1007/s00450-011-0177-x.
[46] I. Triguero, D. Peralta, J. Bacardit, S. García, F. Herrera, MRPR: a mapreduce so- lution for prototype reduction in big data classification, Neurocomputing 150, Part A (0) (2015) 331–345, doi: 10.1016/j.neucom.2014.04.078.
[47] S. Ding, M. Du, H. Jia, Study on density peaks clustering based on k-nearest neighbors and principal component analysis, knowledge-based Systems. 99 (2016) 135–145, DOI: 10.1016/j.knosys.2016.02.00.
[48] Z. Deng, X. Zhu, D. Cheng, M. Zong, S. Zhang, Efficient KNN classification algorithm for big data, Neurocomputing 195 (2016) 143–148.
[49] C. Zhang, F. Li, J. Jestes, Efficient parallel KNN joins for large data in MapReduce, Proceedings of the 15th International Conference on Extending Database Technology, EDBT '12, ACM, New York, NY, USA, 2012, pp. 38–49, DOI: 10. 1145/2247596.2247602.
[50] J. Maillo, I. Triguero, F. Herrera, A MapReduce-based k-nearest neighbor approach for big data classification, 9th International Conference on Big Data Science and Engineering (IEEE BigDataSE-15), 2015, pp. 167–172.
[51] K. Sun, H. Kang, H.-H. Park, Tagging, and classifying facial images in cloud environments based on kNN using MapReduce, Optik 126 (21) (2015) 3227–3233, DOI: 10.1016/j.ijleo.2015.07.080.
[52] Chug, A., Bhatia, A., Singh, A.P. et al. A novel framework for image-based plant disease detection using hybrid deep learning approach. Soft Comput (2022). https://doi.org/10.1007/s00500-022-07177-7
[53] Liu, W., Yuan, J., Lyu, G. et al. Label driven latent subspace learning for multi-view multi-label classification. Appl Intell

(2022). https://doi.org/10.1007/s10489-022-03600-6
[54] Dobre, C., Xhafa, F. Parallel Programming Paradigms and Frameworks in Big Data Era. Int J Parallel Prog 42, 710–738 (2014). https://doi.org/10.1007/s10766-013-0272-7.
[55] Ahmad, A., Paul, A., Din, S. et al. Multilevel Data Processing Using Parallel Algorithms for Analyzing Big Data in High-Performance Computing. Int J Parallel Prog 46, 508–527 (2018). https://doi.org/10.1007/s10766-017-0498-x.
[56] Jin Qian, Duoqian Miao, Zehua Zhang, Xiaodong Yue, "Parallel attribute reduction algorithms using MapReduce", Journal of Information sciences, Elsevier, Vol. 279, pp. 671–690,2014.
[57] Weaver, Jesse. "A scalability metric for parallel computations on large, growing datasets (like the web)." In Proceedings of the Joint Workshop on Scalable and High-Performance Semantic Web Systems. 2012.
[58] Gopi, R., Sathiyamoorthi, V., Selvakumar, S., Manikandan, R., Chatterjee, P., Jhanjhi, N. Z., & Luhach, A. K. (2021). Enhanced method of ANN based model for detection of DDoS attacks on multimedia internet of things. Multimedia Tools and Applications, 1-19.
[59] Sahu, S. K., Mohapatra, D. P., Rout, J. K., Sahoo, K. S., & Luhach, A. K. (2021). An Ensemble-Based Scalable Approach for Intrusion Detection Using Big Data Framework. Big Data, 9(4), 303-321.
[60] Sai, K. B. K., Subbareddy, S. R., & Luhach, A. K. (2019). IOT based Air Quality Monitoring System Using MQ135 and MQ7 with Machine Learning Analysis. Scalable Computing: Practice and Experience, 20(4), 599-606.
[61] Kumar, M. S., & Prabhu, J. (2021). Recent development in big data analytics: research perspective. Research Anthology on Artificial Intelligence Applications in Security, 1640-1663.

Edited by: Vinoth Kumar
Received: Mar 11, 2022
Accepted: Jul 14, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 147–157, DOI 10.12694/scpe.v23i4.2011

COGNITIVE PERCEPTION FOR SCHOLASTIC PURPOSES USING INNOVATIVE TEACHING STRATEGIES

S. ARUNA *AND SWARNA KUCHIBHOTLA †

Abstract. The influence of emotion on attention is particularly strong, changing its selectivity in particular and motivating behavior and action. The degree to which a student participates in class determines their level of conceptual knowledge. Various teaching techniques have been developed over time to improve not only the attention of a student but also their engagement of a student. The level of engagement of a student can help us decide the amount of understanding a student can attain throughout the session. Though these techniques have been developed over time, the basic tests to determine the authenticity of these activities have been done mainly by the use of assessment-based methods. According to research in the field of neuroscience, a person's emotions can assist us to determine a student's level of participation. We also have the affective circumplex model to show us the correlation between emotions and the level of engagement of a person. Taking this into account, we developed an attentivity model with the help of an emotion recognition model (made with the help of VGG-16 architecture in CNN) and the eye tracking system to analyze the amount of engagement being displayed by the student in the class. This model applied to the students on the various teaching models helps us in deciding the effectiveness of various teaching methodologies for the primitive methods of teaching.

Key words: Emotion Recognition, Convolutional Neural Network, Eye tracking system, Attententivity, teaching methodologies.

AMS subject classifications. 68T05

1. Introduction.

Human cognitive functions like perception, attention, learning, memory, reasoning, and problem-solving are significantly impacted by emotion [1]. There exist various methods to increase the amount of engagement of a student in a class. These strategies primarily recognize students' gaining knowledge of capacities for higher expertise and engagement. The present situation requires students to attend their lectures in the online mode but the experiment done can still be relevant for classroom teaching online or offline with a few differences due to the change of environment. Learning is the process of acquiring knowledge, behaviours, skills, standards, or preferences, and learning research has been intimately connected to the growth of psychology as a field of study [2]. Because it demonstrates students' involvement in the in-depth processing of course material and exposes the amount of time spent on the task, the degree of students' (dis)engagement in learning activities can be regarded as an important measure of both cognitive activation and classroom management [3][4].The quantity of information a student can learn depends on how involved they are in the class. There exists a fine line between engagement and attention, while attention is merely to focus, Understanding of the idea is what is meant by participation. The effective circumplex model, developed in the field of neurology, can be used to determine a student's level of engagement. The model indicates different emotional states, such as active positive, activated negative, deactivated positive, and deactivated negative. The best way to understand negative and positive emotions is to use the examples of sadness and happiness. Sadness is a negative emotion, and happiness is a positive emotion. Emotions are complicated phenomena that involve numerous connected subsystems that determine the activated section. Now as we have encountered these, further going into concept the positive activated emotions and the negative activated emotions can lead to better engagement performance in a person, such emotions can be happy, surprise, angry, etc, and vice versa as we can see in Figure 1.1. Taking this model into account we are capable of making a model that can help us in understanding the engagement of

*Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India, & Associate Professor, Department of Information Technology, Vasavi College of Engineering, Hyderabad, India (s.aruna@staff.vce.ac.in).

†CSE Department Koneru Lakshmaiah Education Foundation Guntur, Andhra Pradesh, India (drkswarna@kluniversity.in)

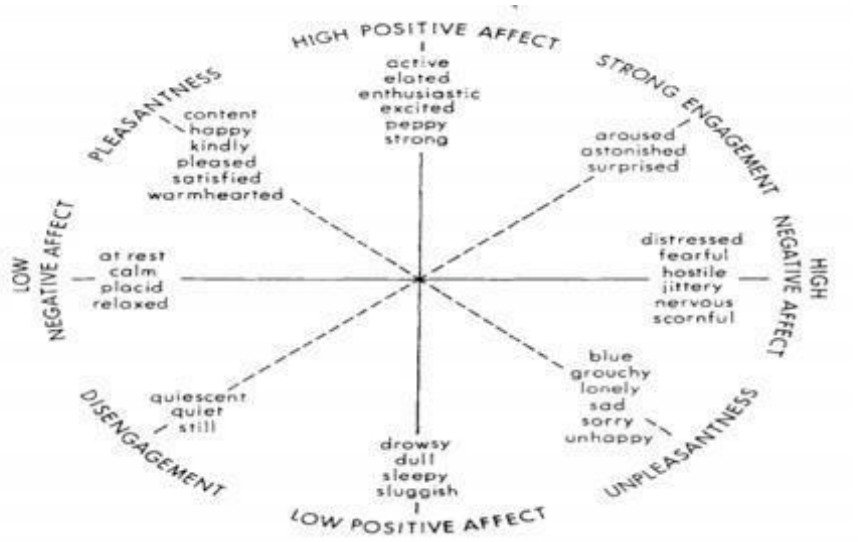


Fig. 1.1: Effect's two-dimensional organisational structure. from article "Toward a Consensual Structure of Mood" A. Watson and Tellegen. Psychological Bulletin, vol. 98, no. 221, 1985. American Psychological Association copyright, 1985

a student through various innovative teaching strategies, these strategies have been used to better understand the effects of these strategies on the students understanding in a classroom along with the amount of focus that the teacher needs to exert while taking a class forms the crux of the paper. This is made possible with the help of the attentivity model which uses two models to capture the attention and engagement of a student. The model uses eye-tracking technology and emotion detection to collect data from different students during the lecture.

The usage of the transfer learning approach with modified CNN and the eye tracking system allows the current paper's model to stand out from those utilized in other works., the experiments before have set to use the approach of using head pose which may still not produce accurate results and using eye tracker calculates the gaze, hence providing a better estimation of attention.

2. Related Work. The importance of the teacher-student relationship for teaching and learning is mentioned in a work by Ashley S. Potvin titled Students speaking to you: teachers respond to student surveys to improve classroom atmosphere. According to the information gathered from the students' experiences, teachers' conversations can be divided into two categories: routines for their classroom setting and pedagogical practice. The survey's findings impact instructors' actions by motivating them to evaluate their classrooms formally. Data collection and data analysis are used to classify the teacher's talk. This analysis of data helps teachers to improve their relationship with students and to create a caring classroom environment [5].

A paper by Zeinab Abulhul titled Teaching Strategies for Enhancing Student's Learning mentions Learning is enhanced only when students actively do something to learn rather than passively listening to the teacher [6]. Teachers have to be compelled to return up with effective teaching methods associated implement innovative solutions to satisfy each student's individual wants within the class. a number of the strategies that lecturers may use in their teaching methods for encouraging students to participate actively in learning method are giving a short lecture as an introduction to the category topic, conducting group action activities, cluster reports that facilitate students summarize the class meeting, on-line resources, and review lesson key points.

Another study called the Transfer learning approach tries to conserve time and resources by avoiding the requirement to train several machine learning models from scratch to complete related tasks. An increase in online learning has been observed as a result of the COVID-19 pandemic, according to a study by Vedant Bahel and Karan K. V[7]. This efficiency gain can be attributed to the use of natural language processing or image

categorization, two machine learning techniques that require large amounts of resources. Smart online learning systems must start by measuring student engagement. Analyzing student behavior is necessary to gauge pupils' levels of participation. This can be achieved by taking a screenshot of the pupils' video feed and sending the faces that are spotted.

They suggested seven potential web-based visualizations in this study that could help them scale the complexity of the depth data they had gathered to track changes in head posture and emotion[8]. While a lecture is taking place, a smart application for teachers is being created to comprehend students' emotions and gauge their degrees of involvement. Computer vision-based approaches are used to evaluate both online and offline lecture films to extract the emotions of the students, as emotions are crucial to the learning process. Using a pre-trained Convolutional Neural Network, the six types of basic emotions—angry, disgusted, fearful, happy, sad, surprised, and neutral—are extracted[9]. Natural user interaction in e-learning is the most ongoing task for researchers. E-learning will be more effective when following the activities of traditional classroom teaching. This work investigated existing methods and presented a survey on head movements and emotion recognition[10].This study examines several face detection, feature extraction, and expression classification methods and techniques and comes to the conclusion that various algorithms and feature extraction approaches may effectively recognise facial expressions. The eye-ball tracking technique is also described in detail because it is crucial for nonverbal communication [11].

This study offers a methodology for automatically recognising facial expressions and focuses on the connection between facial expressions and learning outcomes with an eye toward the unique characteristics of online teaching mode [12]. The automated computer instructor known as AutoTutor converses with students in everyday language while simulating a human tutor. We discovered dominant frequent item sets that predict the following set of replies using a priori data mining techniques[13]. This study investigates the relationship between interactions and epistemic emotions in large-scale MOOC (massive online open courses) data. By assembling the large amount of data generated, this study combined deep learning and social network analysis (SNA) to identify patterns of epistemic emotions concerning interactions on a MOOC platform. [14].

In this study, they demonstrated that the combination of parameters such as MFCC, Root mean square, Spectral contrast, Tonnetz, Zero-crossing rate, Mel spectrogram frequency, and Chroma is effective in properly identifying emotion from the audio frequencies. RAVDESS and TESS, two open-source audio emotion datasets, were used to train the model [15].

The present focus of this study is on using frequency analysis to predict four emotions: angry, sad, neutral, and pleased. To do this, they first extract seven attributes, namely from a single audio file, and 195 sub-features. The qualities includeMFCC, Root Mean Square, Tonnetz, Spectral ContrastZero-crossing rate, frequency of the Mel spectrogram, and Combining chroma helps identify emotion from the audio frequency range precisely [16].

3. Methodology. As mentioned in the content above we will know the engagement of the student in the classroom, but to do so we need to have a model which can aid us in achieving the goal. We use the attentivity model which is an integration of the emotion recognition model made with the help of a Convolutional Neural Network (CNN) with the architecture of VGG-16 (although with a few modifications to make the model compatible with the input size of the dataset) along with the eye tracking system which was made with the help of image processing in the real-time.

The emotion recognition model uses a dataset of around 37K images with each image being of the size 48x48 as we can see in Figure 3.1, these images are in the black and white mode, and the name of the dataset is the FER dataset. The model made takes the RELU activation function to process the input data along with the batch normalization technique to make the model run for a lower amount of time in each epoch, the absence of batch normalization can make each epoch of the model run for at least 90 minutes each. The model is later flattened with the use of the same activation function and then later the output layer in the model has been taken as the SoftMax function which helps us in making a fully connecting layer. The model uses the Maxpooling function to pool the best feature onto the next layer as seen in Figure 3.2.

Though the model is can be perfectly made with the help of a current number of images in the dataset, it achieved an accuracy of 88% with the use of the CNN model using the VGG-16 architecture as seen in the layers in Table 3.1 and Figure 3.3 below, the need to add more images to the dataset was understood due to two schools of thought, one being the improvement of the model by increasing the number of inputs for better

Fig. 3.1: Images from FER dataset showing expressions of anger, fear, happiness and sadness

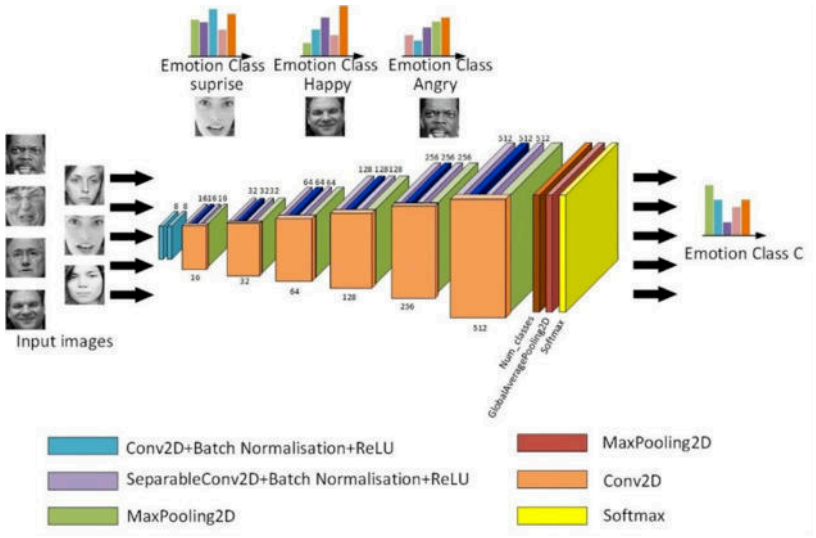


Fig. 3.2: CNN Model for Emotion Recognition

predictability and other being the inclusion of more local emotions as per saying the emotions of self and the local people with the help of transfer learning process.

3.1. Mathematical representation of the CNN model. The initial 4 convolution layers with ELU activation function is as follows

$$hl_{nj}^{1} = conv_{n+1,j}^{1} = Act\left(u_{2n,j}^{1}\right)$$

$$Act = \begin{cases} x, \ x \geq 0 \\ \alpha\left(e^{x} - 1\right), \ \mathrm{x} \ll 0 \end{cases} \tag{3.1}$$

After the 4 convolution layers, there are two flatten layers and they are mathematically represented as given in eqn. (3.2) to (3.4)

$$h_{Z}^{1} = \left\{a_1 h_1^1, a_2 h_2^1, a_3 h_3^1, a_4 h_4^1\right\} \tag{3.2}$$

$$out_{f1}^{1} = Flatten\left(F_n \ Act\left(h_z\right)\right) \tag{3.3}$$

Table 3.1: Layers in CNN

	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7
Type	Convolution	Convolution	Convolution	Convolution	Flatten	Flatten	Dense
Activation	Elu	Elu	Elu	Elu	Elu	Elu	Softmax
Input layer	48X48 X32	24X24X64	12X12X128	6X6X256	256	512	512
Output layer	24X24X64	12X12X128	6X6X256	256	512	512	7
Batch norm	Yes	Yes	Yes	Yes	Yes	Yes	-
padding	Same	Same	Same	Same	-	-	-
Dropout	0.2	0.2	0.2	0.2	0.25	0.25	-
Kernel Site	2	2	2	2	2	2	-
Maxpool size	2X2	2X2	2X2	2X2	-	-	-

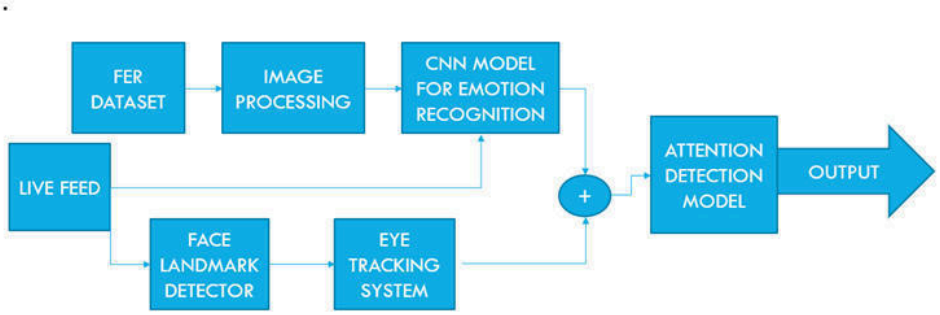


Fig. 3.3: Proposed model for attention estimation system

$$out_{f2}^{1} = Flatten\left(F_{n+1}\ Act\left(out_{f1}^{1}\right)\right) \tag{3.4}$$

Finally, the dense layer is calculated as given in eqn. (3.5)

$$out_{F}^{1} = Dense\left(Den_{c}, Act_{soft\max}\left(out_{f2}^{1}\right)\right) \tag{3.5}$$

The above mathematical represents the layers in the CNN model from eqn. (3.1) to (3.5). It shows the types of layers, number of layers and the activation functions used for achieving the desired output to calculate the attentiveness in students. Eqn. (3.6) is to calculate the percentage obtained the normal distribution

$$Z = (X - \mu)/\sigma \tag{3.6}$$

where X is data, μ = Mean of the data and σ = Standard Deviation.

The model was later able to achieve an accuracy of around 96%. This is paired with the eye tracking system, which uses image processing by taking a photo in real-time from the video being taken and extracting the eyes in particular from the given image with the help of harr_cascade_classifier [17]. Now, taking this cropped part we try to distinguish the white part of the eye from the colored part and approximately calculate the centroid of the eye after extracting the pupil of the eye by taking the average of the x – coordinates and the y- coordinates of the eye. The ratio of the eye position and the relative center position of the pupil is calculated and the vertical and horizontal ratio of the eye is calculated which is bound to be between the values 0 and 1. Now, these values state the extreme position of the eye. This method works better than calculating the head pose of a particular individual, which helps in accurately calculating the attention of the person as the person

Fig. 3.4: Result from main runner function integrating both the emotion recognition model and eye tracking model

can be keeping the perfect head pose and still not be concentrating which can be perfectly identified with the help of an eye-tracking system.

The proposed model as in Figure 3.3 shows the integration of the models formed by combining the emotion recognition model and the eye tracking model to get the attention detection model, which takes input as a live feed which is then converted to data accordingly by both the models and the calculation of attentiveness is performed in the final stage.

Taking both of these models, we make a report with the help of the effective circumplex model as mentioned above to state the various emotional engagement of the student. This experiment is carried out for the various teaching strategies and the report generated through these models will help us in identifying the various advantages and disadvantages of various teaching strategies as seen in Figure 3.4.

The following experiment being done is done based on the computational model built with the help of CNN whereas the earlier models used the normal methods of taking the emotions based on what can be seen or taking one from the student. But as discussed the experiment here is focused on the computational model. The CNN model made uses the relu activation function the usage of relu helps to prevent the exponential growth in the computation required to operate the neural network. If the CNN scales in size, the computational cost of adding extra ReLUs increases linearly. The experiment conducted uses around 57,600 images in the data in total from all the methods conducted. The experiment is made possible with the help of an attention estimation system, which uses emotion recognition made with the help of CNN to track the emotions and the eye tracking system to know the attentivity of the student. Every student was observed taking the emotions of the students into account and their emotions and attentivity throughout the experiment were noted. The students were mainly checked upon the three methods: role play, predict-observe-explain, and lastly a normal ppt lecture. As mentioned, the experiment focused on the emotion and attention of the student and also the lecturer in the session. The sessions were designed in a way to differentiate between the normal teaching method and the innovative teaching methods to prove the involvement of students in the lectures of innovative teaching methods. The experiment also helps us to draw out the amount of attention that the teacher had to use in various teaching methods. The involvement of engagement emotions as per the circumplex model describes the correlation between the emotions and the engagement of the student. The model takes the images of the videos of classes and stores them in a folder which later are used to calculate the attentivity the frame rate has been set to 2s to get as much varied data as possible and also the model works faster with a lesser frame rate.

3.2. Algorithm for the proposed model. The steps for the calculation of attentiveness of students in a classroom using CNN model is as follows:

Step1. Collect live feed from a classroom with many students. Feed is collected for various teaching strategies to identify the advantages and disadvantages of them.

Step2. The live feed is sent to the two models - Emotion recognition model and eye tracking system.

Step3. The feed is sent through the efficient deep learning CNN model for emotion recognition. The model

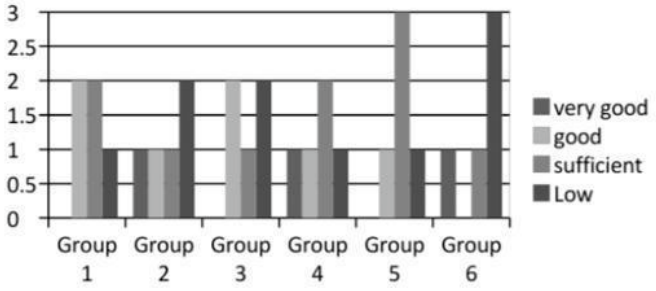


Fig. 4.1: Students' performance analysis by comparing against the role play technique of study

has 4 convolution layers, 2 flatten layers and 1 dense layer through which the images pass. The ELU and SoftMax activation functions are used.

Step4. The results from both the models are generated and used to create the attention detection model.

Step5. The result of attention detection model calculates the attentiveness of students. If the results are not satisfied again the data will be checked from the step 1 till the optimal solution is obtained.

3.3. Dataset used. To train the CNN model for emotion recognition, the FER data set was used. The data consists of 48 x 48 pixel grayscale images of faces. The faces are placed such that the face is centred and occupies almost the same amount of space in each image. This data set has around 37000 such images and are labelled as angry, disgust, happy, sad, neutral, surprised and fear. Hence, there are 7 classes in total.

4. Experimental Results. The first model to be experimented upon is the role-play model. The model consisted of around 19,200 images obtained through the video lectures. Role-play is a technique that allows students to explore realistic situations by interacting with other people in a managed way to develop experience and trial different strategies in a supported environment.

There have been various experiments done on this technique. The implementation of role play in the education of pre-service is where they try to understand the way a student is behaving to role play and without role play. They divided the students into 6 groups to determine its effects as seen in Figure 4.1.

The result of the student's reaction to role-playing confirmed that gaining knowledge of models: made it college students less difficult in knowledge gaining knowledge through role play than gaining knowledge via way of means themselves (100%); being capable of a domesticate mindset and educating the student's responsibility (90%); turned into clean to put into effect in college-level (90%); cultivated the mindset and educated student's awareness (100%); being capable of domesticating the mind-set and to educate the student's independence (75%); educated the scholars as a potential teacher (100%).

Individual results of the experiment conducted show the change in students' engagement to differ by a great margin of 81% when the experiment has been done from our end. This acts as a clear indication of the superiority of a teaching method applied in the enhancement of students understating of concepts. The use of concepts not only allowed us to identify the role of the teacher in making learning more prudent but also the use of groups to allow students to explore the topic to point out the mistakes in it their 'own way. The way of the experiment is sure different yet both of the experiments determine that role play is a better way of teaching strategy. As seen in Figure 4.2, the percentage change in positive emotions of the students is greater in this section as amounting to about 45.4% of the total suggesting more involvement of the student in a class the next model, we have used is for predict observation explain.

In Predict-Observe –Explain (POE) strategy students are required to predict the outcome of an event or experiment. The experiment is then performed and observations made by students are probed. The use of the POE experiment was also previously conducted to conduct a study to compare it with the effect of the Predict-Observe-Explain (POE) strategy on students' conceptual mastery and critical thinking in learning oscillations and waves by Dandy Furqani1*, Selly, to compare Fernie, Nanang Winarno1 who used the method using the

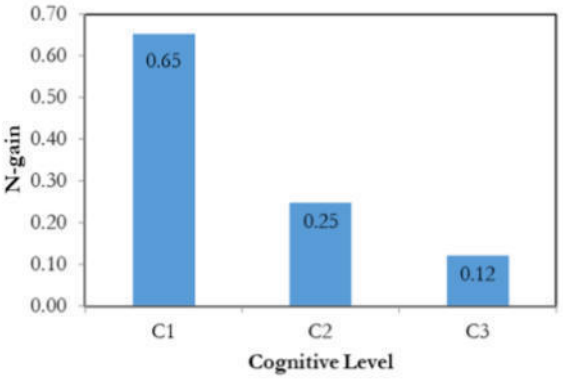


Fig. 4.2: Various cognitive levels in the stages of Predict Observe Explain (POE)

wave and vibration concepts to understand the student's understanding of the concept. [18].

The POE method shows an improvement in the conceptual mastery of university students, indicated by the normalized N-gain score of 0.29 about vital question ability. The POE method seems to be suitable for improving the ability to ask vital questions [19]. Using the POE method, the final result confirmed that university students use the increase in vital questioning from level 1,30 (challenged thinker) to 2,07 (beginner thinker). POE is suitable for implementing the know-how of the students. The results being compared shows the effectiveness of the technique while the experiment from the perspective of the other researcher may not show a better result when compared to our experiment, it can also mean the involvement of other factors which could affect the understanding of a student in this method, yet there is no denying the fact of its effectiveness to be better than another method of teaching. As shown in Figure 4.2 the cognitive levels according to the various methods in predict (c3), observe (c2), and explain (c1).

The results obtained with the help of the 3 methodologies take attention and engagement as the core details to identify the attention of a particular student during the entire session, as mentioned in the experiment section a snap of every 2 sec is recorded in the system and an aggregate is generated by the combination of attention and engagement, now taking a ratio of 3:7 we can calculate the normal distribution of the given sample as given in table 4.1.

The results are obtained after experimenting with the model on the students and the data obtained from them is taken as the percentage calculated as proportions obtained after the normal distribution of data to a group of 8 recorded as from p1...p8 is done for all the 3 categories as shown below in table 4.2.

When predictions and observations are inconsistent with each other the students' explanations are explored and as the experimental results suggested the change in positive emotions of the students is around 43.8% lastly the normal teaching method using the ppt had attention estimating a 10.7% of the total. Now we can conclude that innovative teaching methods have shown more positive results than normal teaching methods as seen in Figure 4.3.

Though the innovative teaching method sure makes a difference there is yet to observe the attention of a lecturer during the process and doing so we have obtained the following results. The role plays though producing promising results also invoke a lecturer to pay more attention towards the whole session which was calculated to be around 2.5% of the total. The ppt follows the lead with 25.4% of the total engagement, not producing results on par with role-playing or predict-observe-explain. Lastly, the predict-observe-explain uses the least engagement by the lecturer and yet produces more results as seen in Figure 4.4.

These two results help us to understand the difference between innovative teaching strategies and also normal methods and also their correlation with the amount of attentiveness paid by the student and also the lecturer in these sessions.

Table 4.1: Sample data obtained from the emotion recognition and eye tracking.

S.NO	Horizontal Ratio Average	Vertical Ratio Average	Attentiveness	Engagement
1	-1	-1	Not Attentive	0
2	-1	-1	Not Attentive	0
3	0.60622645	0.734680954 Attentive	Attentive	0.710312457
4	0.59317995	0.736744499 Attentive	Attentive	0.707882594
5	0.57516877	0.736418482 Attentive	Attentive	0.706953285
6	0.57913352	0.757323251 Attentive	Attentive	0.706096863
7	0.58773354	0.748190621 Attentive	Attentive	0.705938095
8	0.58765688	0.737449167 Attentive	Attentive	0.705532512

Table 4.2: Percentage of normally distributed engagement of students across different methods of teaching.

S.NO	Normal Activity (ppts etc)	Role Play Activity	PEO Activity
P1	10.53168	44.6858	43.1109
P2	11.22503	47.6277	45.9492
P3	12.53242	53.1749	51.3009
P4	5.314356	22.5488	21.7541
P5	16.14461	68.5014	66.0873
P6	9.917002	42.0778	40.5948
P7	9.139172	38.7774	37.4108
P8	10.79573	45.8062	44.1919

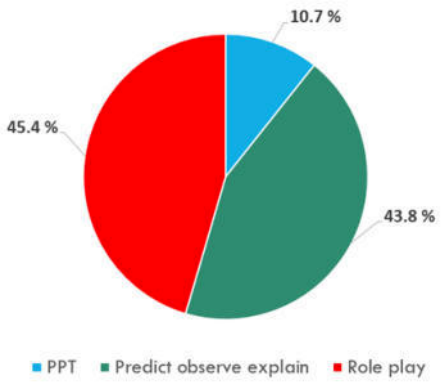


Fig. 4.3: Percentage of positive emotion change in various teaching strategies for students

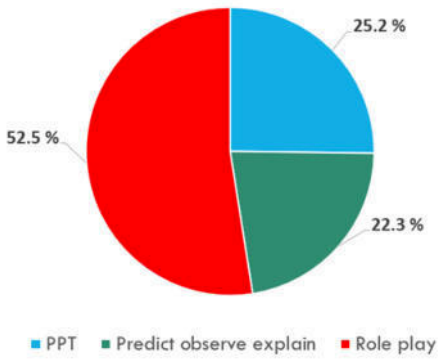


Fig. 4.4: Percentage of positive emotion change in various teaching strategies for lecturers

Though the role play seems to act better in gaining attention, yet has to also be provided with much attention by the lecturer, whereas the involvement of the lecturer is least in the predict observe explain was still able to draw out nearly as much attentiveness as the role-play and the normal ppt method has been proved to be least efficient of all methods.

5. Conclusion. The model successfully was able to prove the superiority of innovative teaching strategies in increasing the engagement of a student in a class. The model can be made better with the inclusion of various other types of inputs such as audio, video, etc which would allow the model to work in diverse situations, and also the model can be made more appropriate. The use of emotion detection can also be done for texts where a sentence can be analyzed to be displaying a certain kind of emotion possibly using methods like using a significant number of words must be entered into clustering techniques like K-Means, Cosine Similarity, Latent Discelet Allocation, and others[20]. The innovative teaching strategies like role play and predict observe explain proved to be more engaging in a class and thus produced better results whereas on other hand the conventional teaching strategy shows less promising results. Though this stands true, there is also another angle for the things to be seen, where we analyze the engagement of the teacher, which would help us understand the amount of involvement and focus that the teacher needs to exert in the whole session. Even though where we were able to see practically less difference in the role play and ppt though we can see the difference in the amount of output being given by both of them when compared side to side. Thus taking everything into account, we can finally conclude without any doubt that the innovative teaching strategies far exceed engaging a student in class rather than conventional teaching methods.

REFERENCES

[1] Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and memory. Frontiers in psychology, 8, 1454.
[2] Carroll, J. M. (1997). Human–computer interaction: Psychology as a science of design. International journal of human-computer studies, 46(4), 501-522.
[3] Carroll, J. B. (1963). A model of school learning. Teachers college record, 64(8), 1-9.
[4] Dharani, T., & Aroquiaraj, I. L. (2014). Content based image retrieval system with modified knn algorithm. International Journal of Innovations in Scientific and Engineering Research (IJISER), 1(1).
[5] Potvin, A. S. (2021). "Students speaking to you": teachers listen to student surveys to improve classroom environment. Learning Environments Research, 24(2), 239-252.
[6] Abulhul, Z. (2021). Teaching Strategies for Enhancing Student's Learning. Journal of Practical Studies in Education, 2(3), 1-4.
[7] Karan, K. V., Bahel, V., Ranjana, R., & Subha, T. (2022). Transfer learning approach for analyzing attentiveness of students in an online classroom environment with emotion detection. In Innovations in Computational Intelligence and Computer Vision (pp. 253-261). Springer, Singapore.

[8] Kalliatakis, G., Stergiou, A., & Vidakis, N. (2017). Conceiving human interaction by visualising depth data of head pose changes and emotion recognition via facial expressions. Computers, 6(3), 25.
[9] Hasnine, M. N., Bui, H. T., Tran, T. T. T., Nguyen, H. T., Akçapınar, G., & Ueda, H. (2021). Students' emotion extraction and visualization for engagement detection in online learning. Procedia Computer Science, 192, 3423-3431.
[10] Krithika, L. B., & GG, L. P. (2016). Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. Procedia Computer Science, 85, 767-776.
[11] Shubham Kadam, Vaidehi Dhawale, Sonali Patil (2019), Facial Gesture Detection and Eye Tracking during the virtual interview, Pramana Research Journal, 9(6), 170.
[12] Zhan, Z., Shen, T., Jin, L., Huang, F., & Xu, H. (2021, March). Research on Evaluation of Online Teaching Effect based on Deep Learning Technology. In 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (Vol. 5, pp. 249-253). IEEE.
[13] Myers, M. H. (2021). Automatic detection of a student's affective states for intelligent teaching systems. Brain Sciences, 11(3), 331.
[14] Han, Z. M., Huang, C. Q., Yu, J. H., & Tsai, C. C. (2021). Identifying patterns of epistemic emotions with respect to interactions in massive online open courses using deep learning and social network analysis. Computers in Human Behavior, 122, 106843.
[15] Suhail MSK, Guna Veerendra Kumar J, Mahesh Varma U, Hari Kiran Vege, Swarna Kuchibhotla. (2020). MLP Model for Emotion Recognition using Acoustic Features. SK et al., International Journal of Emerging Trends in Engineering Research, 8(5), 1702.
[16] Mannem, V., Kuchibhotla, S., "Deep learning methodology for recognition of emotions using acoustic features", International Journal of Pharmaceutical Research, 12(4), 2020
[17] Veni, S. (2016). Image Processing Edge Detection Improvements and Its Applications. International Journal of Innovations in Scientific and Engineering Research (IJISER), 3(6), 51-54.
[18] Furqani, D., Feranie, S., & Winarno, N. (2018). The Effect of Predict-Observe-Explain (POE) Strategy on Students' Conceptual Mastery and Critical Thinking in Learning Vibration and Wave. Journal of science learning, 2(1), 1-8.
[19] Bonwell, C. C., & Eison, J. A. (1991). Active learning: Creating excitement in the classroom. 1991 ASHE-ERIC higher education reports. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183.
[20] Dewan, M., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: a review. Smart Learning Environments, 6(1), 1-20.

Edited by: Vinoth Kumar
Received: Jun 6, 2022
Accepted: Nov 30, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 159–170, DOI 10.12694/scpe.v23i4.2012

GAUGING STRESS, ANXIETY, DEPRESSION IN STUDENT DURING COVID-19 PANDEMIC

ASTHA SINGH *AND DIVYA KUMAR †

Abstract. During the beginning of COVID-19 pandemic, studies came across the world concerning with health issues. Researches began to find the repercussions of the virus. The virus was found to be versatile as it changes its nature and targets the lungs of a person. Later, it was seen an astonishing massacre around the world due to the virus. Many people have lost their life but many more people are still suffering with bad psychological state. Researchers began to research on the nature virus but very few researches were made on the other side-effects of this pandemic. One such crucial subject to attend in contemporary world is the effect of COVID-19 on psychological state in general population. This side-effect may lead to raise an alarming situation in future that could result in more death cases. The proposed paper presents a study on the detection of stress and depression in people caused by the pandemic. The proposed methodology is based on perceived questionnaire method through which people's responses are recorded in the form of text. COVID victims have been interrogated against a set of questions and their responses are recorded. The methodology performs text mining of their responses that also include the people's reaction from social networking sites. The text processing of people's responses is done by natural language processing (NLP). NLP is used to interpret textural facts into meaningful segments that must be understandable to machine. The refined data has been transformed into PSS (perceived stress scale) scaling factor that ranges from 0 to 4 showing various level of stress. The proposed system utilized artificial intelligence in which na'ive Bayes classifier, K-nearest neighbor (KNN), Decision tree and Random forest algorithms are applied to predict the emotional state of a person. The proposed system also uses data from social networking site for testing purpose. The model successfully shows a comparative study of such three classifiers for the classification of stress level into stress, anxiety and depression.

Key words: Text processing, perceived questionnaire method, Natural language processing, naïve Bayes, K-nearest neighbor, principle component analysis

AMS subject classifications. 68T05

1. Introduction.

Stress, depression, and anxiety (SAD) are the most common psychological adversity that affects human life indefinitely. The bad mental state may excite many other diseases in the human body. In India, the reasons for stress or depression are mostly related to personal or professional life. But the current situation in which COVID-19's second wave overwhelmed India's healthcare system resulted in a sudden rise of infected people, which goes about 4 lakh per day [1]. The number of deaths reported was around 4000 per day after March 2021. The rate of spread has seen confounding effects due to the heavier population density in India. India till now reported about 331,909 death cased since Jan 2020. With the huge death reports, India became the second country that suffered from the most infected cases and deaths. With the huge number of death reports due to COVID-19, India is also suffering from a psychological imbalance that may exist for a long duration. According to WHO [2] (world health organization), psychological adversity has also been affected by quarantine for several months, during which people became less interactive with their society. COVID-19 also affects India's economy badly, due to which financial crises have been seen. Common people in India suffered financial losses due to unemployment also. All these unpleasant events contribute to mental disruption in Indian society. The WHO shared their concern on the consequences of obstruction in daily routines and livelihoods of people in Indian society. Several deaths in the second wave [3] caused rapid inclination in loneliness, anxiety, depression, insomnia, suicidal cases, etc. The lockdown also ignited domestic violence, as revealed by one of the studies by Abramson et al. According to Loiwal et al. [4], the second wave of the coronavirus has caused a 20% increase in mental disease, according to the Indian Psychiatric Society. According to a report by Jayakumar et

*Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India (asthasisodia995@gmail.com).
†Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India (divyak@mnnit.ac.in)

al. in Kerala, India, banning alcohol sales during the lockdown has also revealed mental disease among alcohol addicts. The strength of the psychological imbalance among persons after the second wave has not yet been carefully examined because of the lockdown and quarantine regulations. On social networking platforms where users share their experiences, stress and sadness are easily tracked.

The study of stress detection using facial images or video datasets may not seem feasible for the people affected by the pandemic. Such methods require a lab setup and volunteers directly or indirectly affected by COVID. The people spend most of their time at home due to protocols of lockdown and quarantine. They mostly shared their experiences on social media. Suicidal cases have also seen an increase in India. People show reluctance in making gatherings or attending any experimental subject. The experimental setup based on facial images or video datasets is complex and expensive.

The proposed method applied the perceived questionnaire method to determine the psychological state level. This method is based on the interrogation of COVID victims. The method records people's responses in the form of text data. Other textural data from a user who may not appear in an interrogation session has been collected from social networking sites. The responses are compiled and analyzed using the natural processing language (NLP). The NLP breaks down each patient response and correlates them in perceived scaling factor ranges from 0 to 4. The perceived stress scale records people's perceptions and converts them in to score points. The COVID victims are questioned following standard protocol, and their answers are recorded. To estimate the severity of the stress level, the proposed method used the naive Bayes classifier, K-nearest neighbor (KNN), Decision tree, and Random forest algorithms. As a result, the technique divides the reactions into stress, anxiety, and sadness. The classification results for stress, anxiety, and depression have been calculated in a comparative classification study that has been evaluated. The pre-processing techniques are utilized and consider best as it generates effective filtered response relevant for the detection of stress, anxiety and depression.

The following sections comprise the remaining portion of the paper:- Natural language processing, the perceived stress scale, and naive Bayes algorithms are discussed in Section 2 of the proposed methodology. The results section in Section 3 illustrates how well the algorithm worked. We'll talk about the work's conclusion in Section 4. References are included in the final section.

2. Proposed methodology. In the proposed methodology, first the dataset of questionnaire has been taken from standard Microdata Library[9]. This dataset specifically prepared for COVID victims in six different states of India including Jharkhand, Rajasthan, Uttar Pradesh, Andhra Pradesh, Bihar, and Madhya Pradesh. It covers indications related to agriculture, migration, rural, labour markets, consumption patterns, access to relief and healthcare etc. The datasets has been recorded in the may-2020 after the outbreak of second wave of COVID-19 in India [16]. The dataset is sponsored by worldbank [17]. The dataset contain questionnaire for stress detection targets questionnaire related to daily-life activities of COVID victims. The responses from the peoples are collected through personal interrogation, online survey and social media where questionnaire set has been posted. The proposed method also takes textural responses from general post from social networking sites. The correlations of such public posts have been checked with the questionnaire dataset. These responses must correlates with the questionnaire set otherwise the responses are discarded. The responses are based on perception of peoples. Then, the proposed model applies natural processing language [14] (NLP) that performs pre-processing and analysis of textural responses. Then the refined data will be mapped with the PSS scale that ranges from 0 to 4 defining various levels of stress. Then naïve Bayes classifier is used that perform classification of PSS responses based probabilistic approach. The workflow of the proposed system has been described in figure 2.1.

As depicted in figure 2.1, the responses are captured from several peoples over a set of 500 questionnaire data. Each person replied based on his/her perception.

2.1. Natural language processing (NLP). Natural language processing has been applied to the various textural responses recorded from the public domain. The NLP applies syntax analysis and semantic analysis of the data. In syntax analysis, operations like lemmatization, morphological segmentation, word segmentation, sentence breaking, parsing, and stemming are involved. Lemmatization is converting a complex sentence or word into a simple form. Morphological segmentation is used to segregate the sentence into distinct units. Parsing performs grammar checks for the textural responses. Sentence breaking is the breaking of a large sentence into words. Stemming involves the conversion of words into their root form. The semantic analysis

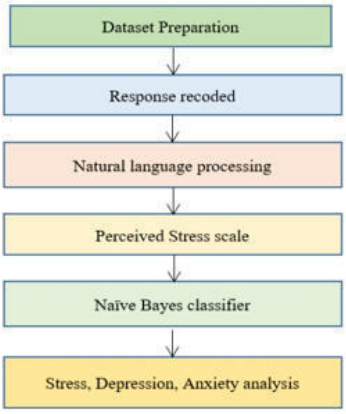


Fig. 2.1: Basic working flow of the proposed model

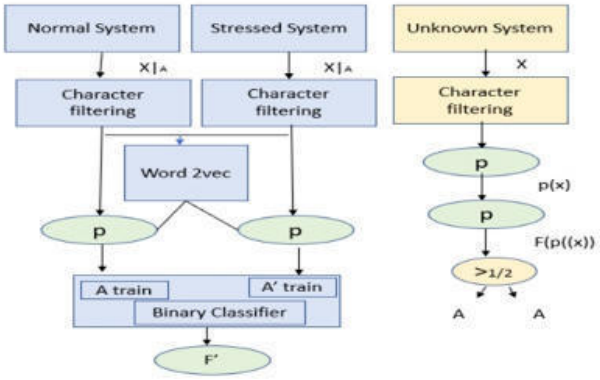


Fig. 2.2: The core working of NLP

NLP performs a meaning check on the textural data. It analyzes the sense of the words in the responses. NLP applies from social networking sites to obtain textural data and correlates it with the existing questionnaire. NLP eradicates all the ambiguity, extra spaces, duplicity, etc., from the dataset and prepares it to correlate with the PSS scaling factor. Figure 2.2 shows the basic working of NLP. The NLP process is used to interpret the meaning of each word and letter of the response taken against questionnaire dataset.

Figure 2.2 shows the character filtration of responses in which several modules are used. The processed language is divided into training and testing set for the validation. The modules used in the character filtration are described in figure 2.3.

3. NLP gives a refined output that is further mapped with PSS scaling factor.

3.1. PSS (Perceived stress scale). An effective psychological tool for assessing stress perception is the Perceived Stress Scale (PSS) [19]. It is a gauge of how stressful certain circumstances are in a person's life. Things are made to change how the unruly, unpredictable, and overburdened respondents feel about their lives. Numerous direct questions about the current state of stress are also included on the scale. The PSS is intended to be applied to community samples that have completed at least a high school degree. Things are simple to comprehend, and some responses are also simple to comprehend. The questions are also generic,thus no substance is directed at any inferior group. PSS questions inquire about emotions and thoughts from the previous month. Respondents were questioned about how frequently they experienced each feeling. The PSS

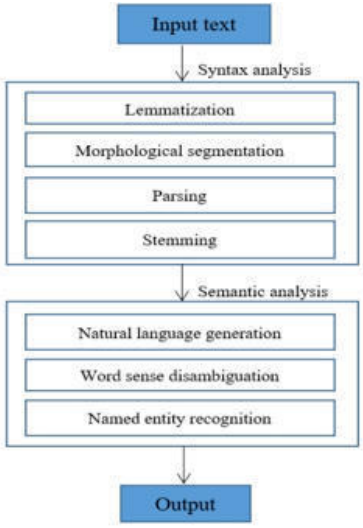


Fig. 2.3: The core working of NLP

might be between 0 and 40. Low stress is estimated to be between 0 to 30. Scores between 14 and 26 are regarded as moderate stress, whereas scores between 27 and 40 are regarded as high stress. The scores are converted into 5 PSS scaling factors by the suggested methodology. The PSS response scores are calculated using the terms 0, 1, 2, 3, and 4. Each score value identifies a particular stress level in an individual.

0 *: - Stress never occurs.*
1 *: - Almost never, yet there is still a risk for tension.*
2 *: - Stress occurs sometimes and forms anxiety.*
3 *: - Stress occurs fairly often considered into moderate anxiety*
4 *: -Stress occurs very often is considered into depression.*

PSS has 14 total scores, which includes the 5 answer points. The total PSS score is created by reversing the PSS scores 4,5,6,7,9,10, and 13 such that 0=4, 1=3, 2=2, 3=1, and 4=0, then adding together all 14 values. A high PSS score indicates a high level of stress.

Evidence for Validity. Higher PSS scores cause the following association:

1. *failure to quit addiction*
2. *failure in curing diabetics*
3. *Health issues, depressive symptoms*
4. *Continuous headache and body pain.*

3.2. Classification algorithms. A comparative result of these three classifiers is calculated and shown in the result section.

3.3. Naïve Bayes. The Bayes theorem has been used in the nave Bayes classifier to determine the likelihood that a feature vector belongs to a category that includes sadness, stress, and anxiety. The classifier is used to analyze the data it receives from the PSS scaling factors.The classifier is used to map the chances that various feature spaces belong to a class. The feature that has the highest probability is categorized into a particular class. The dataset has been split into training and testing portions that are each 70:30.

The equation of Bayes theorem is defined as:

$$P(Y/K) = P(K/Y)P(Y)/P(K)$$

Here,

1. *P(Y|K) is the probability of Y such that event K is already true.*

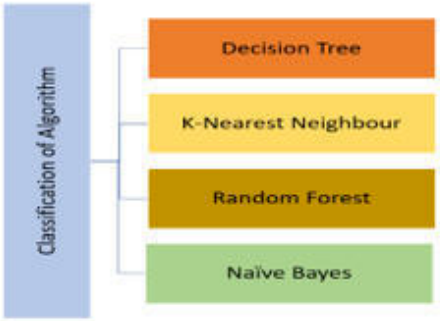


Fig. 3.1: Graphical Representation of Classification Algorithm

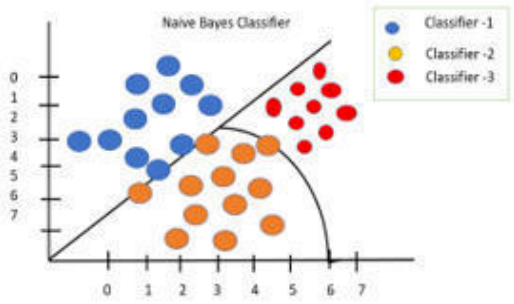


Fig. 3.2: Representation of Naïve Bayes Classifier

2. *P(Y) prior probability of classes.*
3. *Y and K are the two events.*
4. *probality of prediator given class*

The naive Bayes classifier compares the probabilities of belongingness of various feature space in a class. Feature that has maximum probability is classified into a specific class. The dataset has been divided into 70:30 for training and testing respectively.

3.4. K-nearest neighbor (KNN). The KNN algorithm is based on a distance metric assessed by the degree of linguistic similarity. In PSS records, the distance algorithm has been used, and the textural data has been divided into categories for stress, depression, and anxiety.

Step 1:- Input PSS records

Step 2:- The similarity between two selected textural data has been measured by using the following equations:

$$I(A) = \sum_{m=0}^{m} \sum_{n=0}^{n} M_x M_y^{'}$$

$$D_1(I_x, I_y) = \sum_{m=1}^{N} |M_m - M_y^{'}|$$

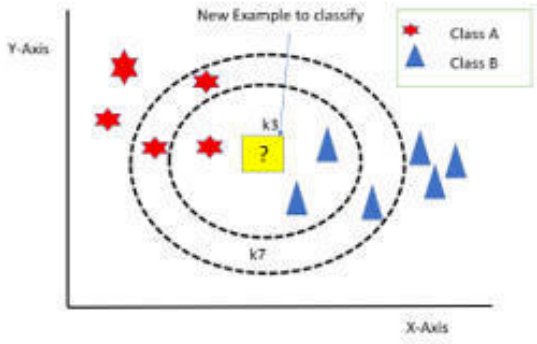


Fig. 3.3: Representation K-Nearest Neighbors

$$D_2\left(I_x, I_y\right) = \sqrt{\sum_{m=1}^{N} |M_m - M_y'|}$$

$$D_3\left(I_x, I_y\right) = \sqrt{\sum_{m=1}^{N} |M_m - M_y'|}$$

$$D_{cos}\left(I_x, I_y\right) = \frac{\vec{M_m}\vec{M_m'}}{|M_m|.|M_m'|}$$

Step 3:- Return
Step 4:- Accuracy measure

$$I\left(A\right) = \sum_{m=0}^{m}\sum_{n=0}^{n} I_x I_y$$

3.5. Decision Tree. Decision tree algorithm takes the input PSS record and transforms it into a tree having root node, parent node and the child not. Each node shows some information. Information of parent node may depend on the information of child nodes. The classification of entire information is represented by leaf node. The decision of stress, anxiety and depression is hierarchically decided by information [20] of non-leaf nodes.

The emotion recognition model uses a dataset of around 37K images with each image being of the size 48x48 as we can see in the figure 2.2, these images are in the black and white mode , the name of the dataset is FER dataset. The model made takes the ELU activation function to process the input data along with the batch normalization technique to make the model run for lower amount of time in each epoch, absence of batch normalization can make each epoch of the model to run for at least 90 minutes each. The model is later is flattened with the use of same activation function and then later the output layer in the model has been taken as the SoftMax function which helps us in making a fully connecting layer. The model uses the Maxpooling function to pool the best feature onto the next layer as seen in the figure 2.3.

3.6. Random Forest. Random forest is the combination of multiple decision trees. It is used to represent data into various decision trees and combined their information to produce single output. The leaf node of each tree shows the target decision i.e. stress, depression and anxiety. The decision of all the trees is combines to

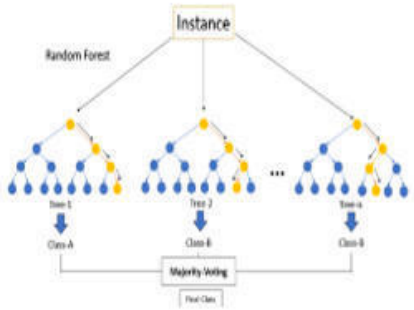


Fig. 3.4: Random Forest

obtain an equivalent classification of data. Each single tree in the model separately predicts for the classification of data. These predictions are combined together to obtain the final output of the model. The decision trees operate as an ensemble.

4. Result discussion. The results are carried out on the standard dataset taken from the Microdata library [9] in which a questionnaire has been prepared and asked from 6 different states of India. The questionnaire was taken in the form of text data. Then, NLP is applied to conduct pre-processing. The pre-processed data is further converted into PSS scaling factors. There are other noise removal techniques such as filtration methods depending on datasets. The utilized pre-processing techniques in the proposed model is effective as the questionnaire dataset is concerned. The questionnaire responses contained in terms of textual data that needs to be normalized. It has been done by the proposed pre-processing techniques in low cost and high effectivity shown by the robustness. Table 4.1 shows the sample of the questionnaire and the respective PSS records scaling from 0 to 4.

Table 4.1 shows the samples of questions taken from the standard dataset. The responses are transformed into PSS scaling factors. The median and mode has also been calculated for the PSS ratings. These PSS records are classified into stress, anxiety and depression. The classification task has been performed four types of classifier i.e. Naïve Bayes, K-nearest neighbour, Decision tree and the random forest.

Table 4.2 shows the confusion matrices obtained by all the four classifiers that is used to classify anxiety, depression and stress.

Th below equations are used for the calculation of accuracy rate, precision rate, recall and F1 score:

$$\text{Acurracy Rate} = \text{Diagonal sum (MP)} \;/\; \text{Total number of particles}$$

$$\text{Error Rate} = 1\text{- Acurracy Rate}$$

$$\text{Precision Rate} = \text{MP}/\ \text{MP}+\ \text{NF}$$

$$\text{Specificity} = 2^{*}\ (\text{Precision} * \text{Recall}\ )/\ (\text{Precision} + \text{Recall})$$

where
– MP indicate -(True positive)= Matrix diagonal
– NF indicate - (False Negative) = Consistent row for class
– FM indicate - (False Positive) = Corresponding column for class
– NF indicate (False Negative) =Sum of all column and row

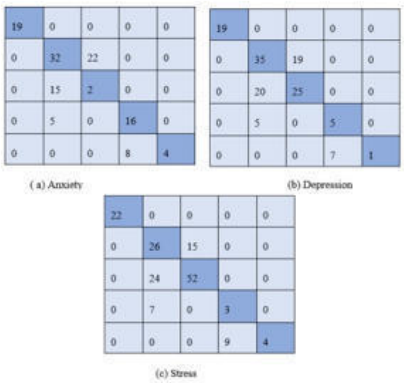


Fig. 4.1: Graph Show Confusion matrices for SAD Naïve Bayes Classifier

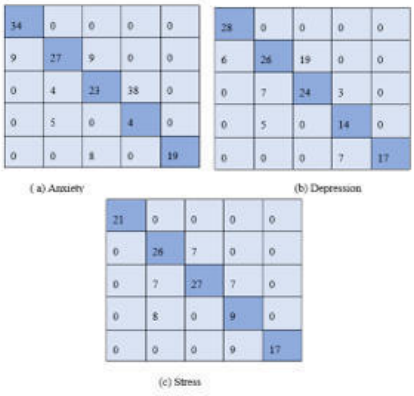


Fig. 4.2: Graph shwo Confusion matrices SAD KNN

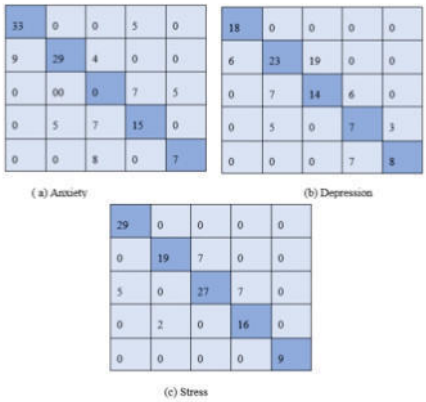


Fig. 4.3: Graph showing the Confusion matrices for SAD Decision tree

Table 4.1: Sample of questionnaire and the related PSS response (rated in a scale of 0-4) as reported from people

Source of Stress	Never (0)	Almost Never (1)	Some-times (2)	Fairly Often (3)	Very Often (4)	Media n	Mod e
Have your personal plans been changed or affected by the COVID in the last 7 days?	28	24	35	22	5	11	5
Did you do any work for pay during COVID at a job or business in the last 1 month?	14	18	25	21	14	6	4
Think about your life for past 30 days, how do you think, was it a really tough time?	20	17	21	29	27	4	6
Have you currently covered by any of the types of medical insurance or any health plan?	23	14	29	38	12	2	5
Have you had any friend or member who is very close to you die from COVID-19 or recovered from 2nd wave of COVID.	8	5	18	30	28	4	3
How did you communicate with your family, friends and relatives during the last 1 month? Did you use any phone, text, email, or Internet something to connect with you close one?	32	20	24	15	6	2	3
After 1 march 2020, what did you think about the change in your life Did you talk with any of your neighbors?	8	13	30	32	14	3	5
After the second wave of COVID, did you voluntarily appear in any organization or association as a part of helping hands?	11	24	28	18	6	4	4
In the past 1 month, have you ever felt nervous, anxious, sad, lonely etc.	14	20	24	28	17	9	4
Do you feel any slackening in health sector in the past 1 month?	17	23	31	16	8	4	8
Did lockdown seriously affect your job or financial source?	27	27	10	22	7	3	4
Have you realized any major changes in your life after 1 march 2020?	8	21	30	25	21	4	5
Have you ever visited in hospital for any reason after 1 march 2020?	16	14	32	28	9	5	4

Figures 4.1–4.4 are showing confusion matrices for anxiety, depression and stress. These confusion matrices are generated by all the four classifier used in the proposed scheme.

Table 4.1 is showing the result on the basis of confusion matrices for SAD. These confusion matrices are represented by all the four classifier used in the proposed model.

Table 4.2 depicted for the result scores including rate of accuracy, error value, rate of precision, rate of recall and F1 score. These scores are calculated for the respective anxiety, depression and stress represented as A, D and S. The various classifiers have been shown in this table based on which the results have been calculated.

Table 4.2: Measure of accuracy, error rate, precision, recall and F1 score for anxiety (A), depression (D) and stress (S)

Name of Classifiers	**Psychological status**	**Value of Accuracy**	**Value of Error Rate**	**Data of Precision**	**Data of Recall**	**F1 score**
Naïve Bayes	A	0.738	0.27	0.523	0.658	0.658
	D	0.851	0.28	0.574	0.798	0.798
	S	0.708	0.29	0.596	0.798	0.745
KNN	A	0.745	0.28	0.795	0.715	0.748
	D	0.796	0.25	0.853	0.896	0.821
	S	0.854	0.28	0.841	0.889	0.856
Decision tree	A	0.874	0.248	0.896	0.854	0.895
	D	0.824	0.241	0.813	0.896	0.865
	S	0.887	0.267	0.796	0.874	0.879
Random forest	A	0.752	0.224	0.745	0.897	0.785
	D	0.789	0.296	0.758	0.874	0.796
	S	0.796	0.274	0.746	0.752	0.798

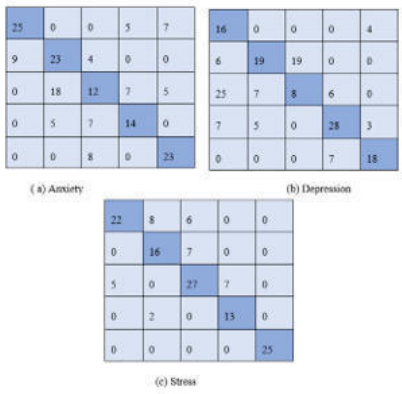


Fig. 4.4: Confusion matrices for anxiety, depression and stress for Random Forest

Figure 4.5 shows the graphical comparison of accuracies obtained by all four classifiers for stress, depression, and anxiety. Figure 12 shows that the highest accuracy has been observed in the decision tree classifier compared to other studied classifiers. The relevance of the study is to prevent suicidal cases and improve the treatment policy of the patients having mental disorder. The questionnaire dataset is effective as it contains the exact patient's responses and easy to interpret using tools and techniques. The limitation of the model is its static nature as the responses against the questionnaire dataset are static and it may not be effective as other dynamic dataset such as video-based interrogation, audio responses etc. The future of the model is in the medical science where it helps to reduce cost for the detection of mental disorder. Various other language filtration processing tools can be utilized in the place of NLP but the NLP method is found to be more effective as per the textual questionnaire responses is concerned of the proposed model. The frequency analysis is very crucial as it give details about intensity information based on statistical scores. The frequential information shows the exact

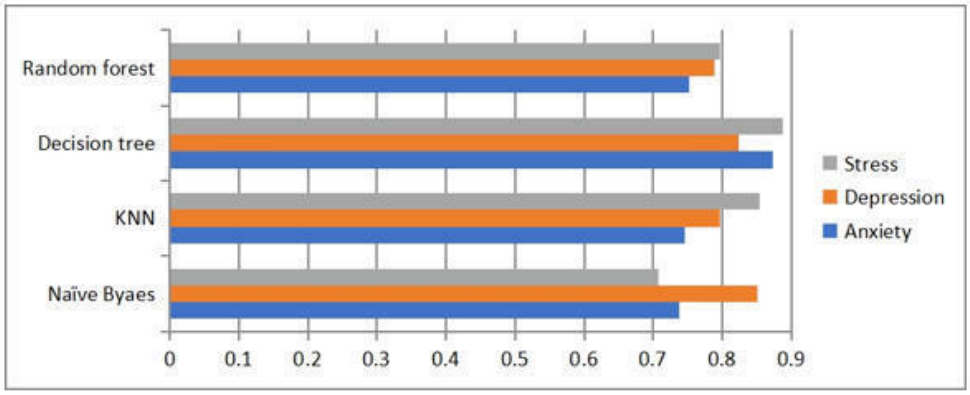


Fig. 4.5: Graphical representation of accuracy analysis for stress, depression and anxiety

variation of psychological activities that helps to reach to the accurate detection of the stress, depression and depression.

5. Conclusion. The proposed work conducted an experiment to identify stress, depression, and anxiety using four different types of classifiers. The model successfully identifies how people's psychological states relate to the COVID-19 pandemic. The dataset for the model is a collection of questionnaires that were created for the COVID second wave. The suggested system uses offline and online questioning to gather survey responses from various people. Utilizing natural processing language, the responses are analyzed and processed. PSS scaling factors have been created from the processed data. The data was then categorized into stress, depression, and anxiety using classification algorithms. The model is effective for the detection of stress, depression, and anxiety, according to the results section. The intensity level of the adverse psychological disorder is used to distinguish among stress, anxiety and depression. The proposed model concludes that the decision tree classifier is found to generate more accuracy as compared to other classifiers.

REFERENCES

[1] A. Afzali, A. Delavar, A. Borjali, and M. MIRZAMANI, "Psychometric properties of DASS-42 as assessed in a sample of Kermanshah High School students," 2007.
[2] N. Bayram and N. Bilgel, "The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students," *Social psychiatry and psychiatric epidemiology*, vol. 43, no. 8, pp. 667–672, 2008.
[3] A. T. Beck, R. A. Steer, G. K. Brown, and others, "Manual for the beck depression inventory-II," *San Antonio, TX: Psychological Corporation*, vol. 1, p. 82, 1996.
[4] M. Hamilton, "A rating scale for depression," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 23, no. 1, pp. 56–62, 1960.
[5] A. Ghaderi and M. Salehi, "A study of the level of self-efficacy, depression and anxiety between accounting and management students: Iranian evidence," *World Applied Science*, vol. 12, no. 9, pp. 1299–1306, 2011.
[6] M. Gururaj, "Kuppuswamy's Socio-Economic Status Scale," *A Revision of Income Parameter For*, pp. 1–2, 2014.
[7] M. Hamilton, N. Schutte, and J. Malouff, "Hamilton anxiety scale (HAMA)," *Sourcebook of Adult Assessment: Applied Clinical Psychology*, pp. 154–157, 1976.
[8] J. D. Henry and J. R. Crawford, "The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample," *British journal of clinical psychology*, vol. 44, no. 2, pp. 227–239, 2005.
[9] COVID-19 Related Shocks Survey in Rural India 2020.
[10] B. S. Kee, "A preliminary study for the standardization of geriatric depression scale short form-Korea version," *J Korean Neuropsychiatr Assoc*, vol. 35, no. 2, pp. 298–307, 1996.
[11] B. Kuppuswamy, "Manual of socio-economic status scale," *Delhi: Manasayan Publication*, 1962.
[12] L. Manea, S. Gilbody, and D. McMillan, "Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis," *Canadian Medical Association Journal*, vol. 184, no. 3, pp. E191–E196, 2012.

[13] S. Naveen, M. Swapna, and K. Jayanthkumar, "Stress, anxiety and depression among students of selected medical amd engineering colleges, Bangalore - A comparative study," 2015.
[14] A. Raskin, J. Schulterbrandt, N. Reatig, and J. J. McKEON, "Replication of factors of psychopathology in interview, ward behavior and self-report ratings of hospitalized depressives.," *The Journal of nervous and mental disease*, vol. 148, no. 1, pp. 87–98, 1969.
[15] M. Shah, S. Hasan, S. Malik, and C. T. Sreeramareddy, "Perceived stress, sources and severity of stress among medical undergraduates in a Pakistani medical school," *BMC medical education*, vol. 10, no. 1, p. 2, 2010.
[16] P. Svanborg and M. Åsberg, "A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Åsberg Depression Rating Scale (MADRS)," *Journal of affective disorders*, vol. 64, no. 2, pp. 203–216, 2001.
[17] P. Vitasari, M. N. A. Wahab, A. Othman, T. Herawan, and S. K. Sinnadurai, "The relationship between study anxiety and academic performance among engineering students," *Procedia-Social and Behavioral Sciences*, vol. 8, pp. 490–497, 2010.
[18] W. W. Zung, "A self-rating depression scale," *Archives of general psychiatry*, vol. 12, no. 1, pp. 63–70, 1965.
[19] Sher, K.J., Wood, P.K. & Gotham, H.J. (1996). The course of psychological distress in college: A prospective high-risk study. Journal of College Student Development, 37(1), 42–51.
[20] Jogaratnam, G. & Buchanan, P. (2004). Balancing the demands of school and work: Stress and employed hospitality students. International Journal of Contemporary Hospitality Management, 16(4), 237–245.

Edited by: Vinoth Kumar
Received: Jun 4, 2022
Accepted: Dec 5, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 171–179, DOI 10.12694/scpe.v23i4.2015

INTEGRATING COLLABORATIVE FILTERING TECHNIQUE USING RATING APPROACH TO ASCERTAIN SIMILARITY BETWEEN THE USERS

PAVITHRA C* AND SARADHA M†

Abstract. The recommender system handles the plethora of data by filtering the most crucial information based on the dataset provided by a user and other criterion that are taken into account.(i.e., user's choice and interest). It determines whether a user and an item are compatible and then assumes that they are similar in order to make recommendations. Recommendation system uses Singular value decomposition method as collaborative filtering technique. The objective of this research paper is to propose the recommendation system that has an ability to recommend products to users based on ratings. We collect essential information like ratings given by the users from e-commerce that are required for recommendation, Initially the dataset that are gathered are sparse dataset, cosine similarity is used to find the similarity between the users. Subsequently, we collect non-sparse data and use Euclidian distance and Manhattan distance method to measure the distance between users and the graph is plotted, this ensures the similar liking and preferences between them. This method of making recommendations are more reliable and attainable.

Key words: Collaborative filtering; Euclidian distance method; Manhattan distance method; SVD.

AMS subject classifications. 68T35

1. Introduction. Singular value decomposition is a significant technique used for recommendation system. SVD is widely used in developing various models. In precise SVD is used in e-commerce recommendations, it contributes to decrease the range of the data sets values by using matrix factorizing method, (cuts the space dimensions from N-Dimensions to K-Dimensions, where strictly K is lesser than N). Since singular value decomposition is a method of linear algebra, it uses the matrix structure to solve the system of problems where each column represents an item and each row user represents the user. Each element present in the matrix are the ratings that are given by the users to the respective items.

1.1. Singular Value Decomposition. Matrix Factorizing is done using singular value decomposition method and given by

$$A = PSV^T$$

The dataset value (user-item-ratings) of the matrix are decomposed using singular value decomposition method. This helps the matrix decompose into three types as given below, thus the factors of the matrix are obtained,

$$A_{m,n} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{pmatrix} \tag{1.1}$$

*Research Scholar, Department of Mathematics, Reva University, Rukmini knowledge park, Bengaluru 560064, India (pavithracshekar5@gmail.com)

†Associate Professor, Department of Mathematics, Reva University, Rukmini knowledge park, Bengaluru 560064, India (ishuharisri@gmail.com)

$$A_{m,n} = \begin{pmatrix} p_{1,1} & \cdots & p_{1,r} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,r} \end{pmatrix} \quad \begin{pmatrix} s_{1,1} & \cdots & 0 \\ 0 & \vdots & \vdots \\ \vdots & \vdots & s_{r,r} \end{pmatrix} \quad \begin{pmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & \vdots & \vdots \\ v_{r,1} & \vdots & v_{r,n} \end{pmatrix} \tag{1.2}$$

where A is a $m \times n$ utility matrix, P is a $m \times r$ left singular orthogonal matrix, this represents the affinity between users and latent factors. Latent factors are defined as describing the nature of the items. S is a $r \times r$ diagonal matrix, recounts the strength of each latent factor and V is a $r \times n$ diagonal right singular matrix, which indicates the relationship between items and latent factors. From (1.1) and (1.2) we see that, the utility matrix is decomposed into orthogonal matrix and diagonal matrix.

Latent factor contributes to reduce the dimension of the matrix in singular value decomposition method. The below equation is the mapping that accelerates a clear representation of relationships between users and items.

$$r_{u,i} = x^t \cdot y_u$$

Let xi represent each item, and yu represent each user. The expected rating by a user on an item $r_{u,i}$ can be given as Here, $r_{u,i}$ is a form of factorization in singular value decomposition.

1.2. Recommendation system. The role of recommendation system is to give suitable and relevant suggestions to the users, this determines the relationship between the users and the products, and also helps the organization to provide an appropriate product suggestion. Recommendation system is incorporated with collaborative filtering method, which uses the concepts of singular value decomposition. Most of the e-commerce uses recommendation system for their customer support interface.

Recommendation system has few beneficial aspects, to both user and organization. It contributes to avoid the local transaction caused by choosing products or items. Therefore, It helps in getting the suitable suggestion for the relevant search depending on the previous search history.

Systems that make recommendations employ a variety of technologies. These systems can be divided into two main categories called content based filtering and Collaborative filtering.

1.3. Content-based Filtering. Making decisions using similar features is a machine learning technique and are called content based filtering. This method is frequently applied in recommender systems, that are algorithms created to promote or suggest products to users based on data gathered about the user.

1.4. Collaborative Filternig. Collaborative Filtering claims an assumption, this approach describes the user who liked a product now will tend to have the similar likings in the future. This is done by analyzing the behavior of the user in the previous searches, likes, add carts, ratings etc. by using this approach the model finds the relationship and inter-dependence of two variables (i.e., users and items). Thus, Collaborative filtering is performed by using the technique of singular value decomposition.

1.5. Rating-based Recommendation. The familiarity of a product or Item can depend on high user ratings. In recommender system the customers or users give their perspective in the form of ratings and comments also users tend to give "explicit feedback", Also further, clicks, shop for, and search history are called "implicit feedback".

Depending on the ratings given to the products or the items the product suggestions are done to the users, in advancement to the same depending on the previous searches and choices them recommendation is done. This helps in filtering the information and preferences achieved by the item given by the user.

2. Literature Review. In [1] by Utkarsh Gupta et al. Hierarchical clustering technique was used for product recommendation system, this approach gives the results of low error and better clustering of similar items. In this research paper voting technique was used to determine the rating values of the product.

In [2] by Xiaoyuan Su and Taghi M. Khoshgoftaar. discuss about the challenges and the main charges of collaborative filtering, such as privacy protection, data sparsity, scalability etc., this paper also describes three different types of collaborative filtering technique, like model based CF which finds the relationship between

the user and the items. Memory based and hybrid CF, it is the combination of both the methods model based and memory based collaborative filtering technique. This was the survey based research paper.

In [3] by Harpreet Kaur et al. presents the research work in the hybrid system, which involves the combination of conceptual contents and the CF algorithms. This means that the recommendations are done based on the stuffs and concepts involved in any closure work. It basically progresses on user with user and also user with items for the recommendation.

In [4] by Geetha G, Safa M, Fancy C, Saranya D. formulated for movie recommendation system, which helps in recommending the movies to the users. This revises the existing movie data base set for few relevant information like ratings of the movie, genre of the movie and popularity. This paper is the combination of content based collaborative filtering and hybrid filtering method. It helps in achieving more precise results for recommendations.

In [5] by Sarwar B, Karypis G, Konstan J, Riedl J. used CF recommender system that collects ratings from users for products in a particular domain. Association inferences, which are exceedingly time- and scalability-intensive and have a very likely temporal complexity, were also a reliance on CF algorithms used for recommendation systems. The more efficient and scalable matrix operations are used in modern methods.

In [6] by C. S. M. Wu, D. Garg, and U. Bhandari. have used collaborative filtering technique for suggesting in e-commerce. Here the suggestions are done based on the user ratings. This paper has included Apache Mahout Framework, and the efficiency are checked between user based and item based recommendations.

In [7] by Pennock and Horvitz. contributed to the clustering method of solving the problem, it works on different data set size of the user feature vectors. A new concept of LGS was introduced to check the perform ability of the collaborative filtering methods. LGS stands for Latent Genre Space.

In [8] by Debadrita Roy, Arnab Kundu. have contributed to the recommendations with respect to movies. This is achieved with the help of collaborative filtering technique. Expectation Maximization Algorithm are used for movies recommendations. Clustering the data was also used for recommendation system desings.

In [9] by J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. developed a research paper which is the open ended research paper, where the author ends the paper with the question and the challenged faced by the recommendations like privacy policies and terms conditions. Design decisions are developed for rich interaction interface. The primary use was to privacy structuring regarding rating systems.

In [10] by Desrosiers C, Karypis G. Implemented the simple logic, the product used or liked by the user in the past will also be liked in the future. Here collaborative filtering techniques are used and developed to meet the expectations of the users.

3. Background of the work.

3.1. Collaborative Filtering. The most common method for creating recommendation systems is collaborative filtering, which has been effectively used in numerous applications. The CF recommender system gathers user feedback for products in a certain domain in the form of ratings. Also CF algorithms which was for recommendation systems relied on association inferences, which are extremely time- and scalability-intensive and have a very probably about time complexity. Modern techniques use matrix operations, which are more effective and scalable [5].

3.2. Memory based collaborative filtering. It is based on both the item's description and the user's preference profile. In memory-based collaborative filtering, we use key-terms in addition to the user's profile to propose items by indicating the user's preferred likes and dislikes. In other words, products that were previously favored are recommended via a memory-based CF algorithm. It looks at previously rated things and suggests the best option [3][5].

3.3. User-based prediction system. When compared, User-based filtering method it is expected to be better to be working with huge amounts of data, while item based collaborative filtering methods are used when the data set are small in size.

We say that user based collaborative filtering technique is used for the large amount of data. As discussed above CF is used in determining the likings and disliking of the product or the items. Here Figure 3.1 represents the flow chart for user-based prediction system. This is done by monitoring the previous search history of

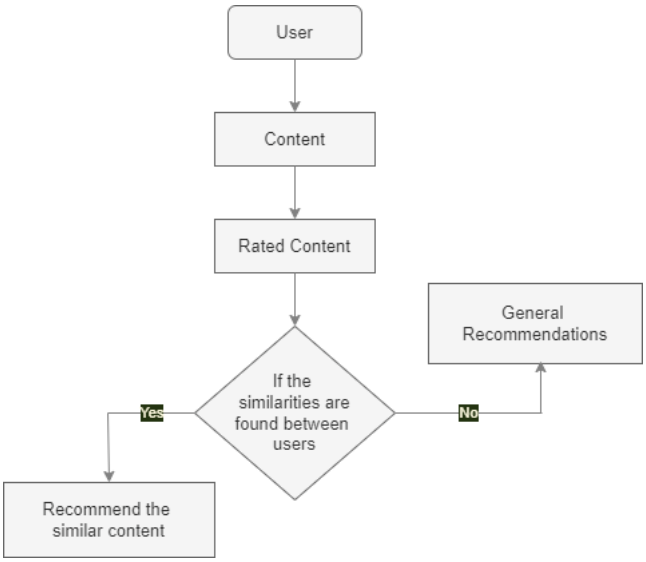


Fig. 3.1: Flow Chart

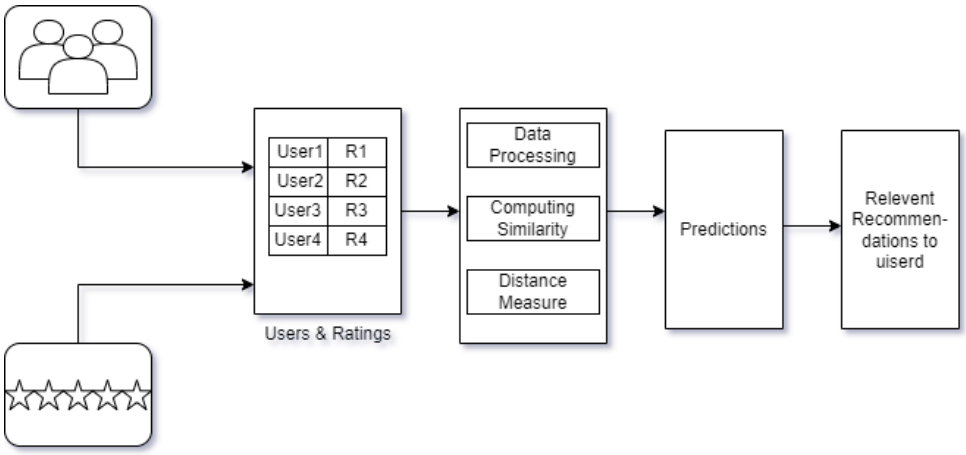


Fig. 4.1: Proposed Structure

any users. Today most of the e-commerce are using the Collaborative filtering technique for improving the recommendation suggestion to their customers and users.

4. Proposed System and Methodology. To find the similarity between the users, the following claim is raised,tabular column(users vs items)the User1 who have rated 4.5/5 for “Item1” and “Item3” and if User2 has rated 4/5 for product “Item1” and “Item2”,then we conclude that user1 and user2 have similar liking. Also, “Item1” will be recommended to User1 and “Item3” will be recommended to User 2.

The structure of the proposed system is given in Figure 4.1. The set of data(ratings) given by the users are collected from e-commerce and tabulated. The obtained data are pre-processed and the missing values are computed. The similar users are identified using cosine similarity and centered similarity. An illustration was performed by collected twenty-five data samples from e-commerce, graph was plotted across users and their ratings for respective products. The distance between the sample points was computed and classified into

Table 4.1: Hyper-parameter values

Hyper-parameter	Values
Input values	28
Missing values	17
Sparse data	17
User values	11
Product samples	7
Cosine Similarity (U1, U2)	0.093
Cosine Similarity (U1, U3)	-0.587
Non-sparse dataset	25
Classified Clusters	5

clusters, and they represent the users with similar likings and preferences.

In Specific we are using cosine similarity method and distance measure i.e., Euclidean distance method and Manhattan distance method to find the similarity between the users. Here Cosine similarity finds the comparability between two vectors in inner product space establishes whether two vectors are roughly pointing in the same direction by calculating the cosine of the angle between them. In text analysis, it is frequently used to gauge document similarity. The method of finding the distance between any two points is known as Euclidean distance method. It helps in determining the length of the line segment between any two points. Also Manhattan method is used as a metric for measuring distance between any two points in an N-dimensional vector space.

This new approach of identifying the similarity and common likings between the users was performed because the gathered data was sparse data, making it difficult to make the right suggestions to the users. As a result, we start by turning all of the sparse data into easily recognizable unique values, then we continue to compute how similar they are to one another, using cosine similarity method. Subsequently, we collect a non-sparse dataset from e-commerce to compute similarity between the users, we utilize the distance measure approach to calculate similar users with particular items.

5. Implementation. The role of the proposed system is to determine the missing rating of the products or items. The tabular column of the rating data sets is computed by initially assigning nil values and further the values are neutralized. Cosine similarity method is implemented between the users to find the similarity between them and it is given by

$$Cosinesimilarity(u,v) = Similarity(u,v) = \frac{\sum(r_{u,i}-r)(r_{v,i}-r)}{\sqrt{\sum(r_{u,i}-r)^2}\sqrt{\sum(r_{v,i}-r)^2}} \tag{5.1}$$

Table 5.1 represents the score spread of ratings given by four different users for seven different items or products. Since not all the users rate all the seven products, there exists few missing values in the datasets. For further analysis, we require to complete the table by marking the missing values with zero and converting the missing dataset values to sparse data. A dataset with sparse data is one in which a sizable portion of the cells do not really contain any data, but occupies the storage space in the file.

By using (5.1), similarity betweeen user1 and user2 is given by Sim(U1, U2) = 0.2 and Sim(U1, U3) = 0.5. This is contradicting the theory of similarity for common liking recommendation, and thus, the following table

Table 5.1: User Ratings

	user1	**user2**	**user3**	**user4**	**user5**	**user6**	**user7**
Item1	4			5	1		
Item2	5	5	4				
Item3				2	4	5	
Item4		3					3

Table 5.2: Generalizing missing values

	user1	**user2**	**user3**	**user4**	**user5**	**user6**	**user7**
Item1	4	0	0	5	1	0	0
Item2	5	5	4	0	0	0	0
Item3	0	0	0	2	4	5	0
Item4	0	3	0	0	0	0	3

Table 5.3: Cosine similarity

	user1	**user2**	**user3**	**user4**	**user5**	**user6**	**user7**
Item1	4	0	0	5	1	0	0
Item2	5	5	4	0	0	0	0
Item3	0	0	0	2	4	5	0
Item4	0	3	0	0	0	0	3

Table 5.4: Centered Values

	user1	**user2**	**user3**	**user4**	**user5**	**user6**	**user7**
Item1	2/3			5/3	-7/3		
Item2	1/3	1/3	-2/3				
Item3				-5/3	1/3	4/3	
Item4		0					0

was evolved, sim(A,B)= cos(rA,rB). In this method rating values are Ignored. This is contradicting the theory of similarity for common liking recommendation, and thus, the table 5.2 was evolved.

Similarity(U1, U2) = 0.38 and Similarity(U1, U3) = 0.322. Therefore, Sim(U1, U2) is greater than Sim(U1, U3). The difference between them are not significant. The similarity is not precise and accurate.

By Neutralizing the Table 5.3 we get Table 5.4.

By using (5.1) we compute Similarity(U1, U2)= 0.093 and Similarity(U1, U3)= -0.587. Table 5.4 represents the centered values of ratings. Therefore, Sim(U1, U2)is greater than Sim(U1, U3). This shows that there exists a similarity between User1 and User2 and non-similarity between User1 and User2 (Since, Negative value).

Our next step, is to use distance measure methods which makes conclusion about an item, this method will calculate the "distance" between the target movie and every other item in its database, then it ranks its distances as the most similar recommendations.

5.1. Distance Measure. Here we are using methods for calculating distance between points are Euclidian and Manhattan distance are the two methods that are used for computing the relationship between any two variables and objects. Euclidean distance method is the familiar or called basic method of finding the distance between the points.

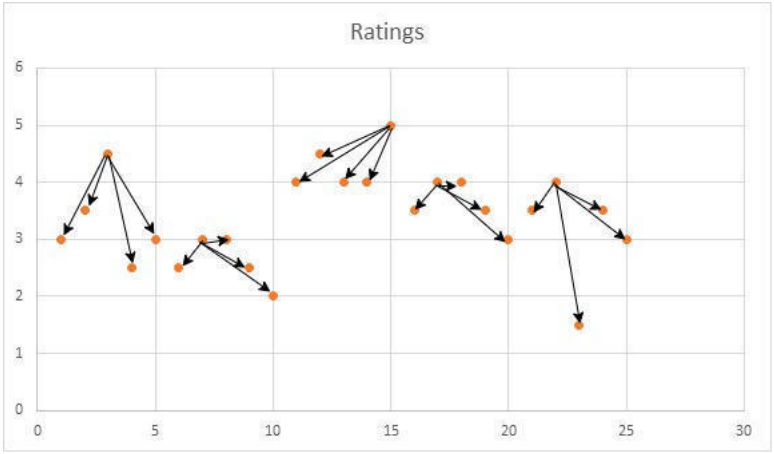


Fig. 6.1: Scatter Plots

Euclidean Distance: Euclidean distance is defined as, calculating the square root of the sum of the squared differences between a new point (x) and an existing point (y).

$$d_(i,j) = \sqrt{(x_{i,1} - x_{j,1})^2 + (x_{i,2} - x_{j,2})^2 + ... + (x_{i,3} - x_{j,3})^2} \tag{5.2}$$

or

$$d_(i,j) = \sqrt{\sum (x_{i,k} - x_{j,k})^2} \tag{5.3}$$

Manhattan Distance: Manhattan distance is defined as finding the absolute values of the difference of real vectors, and it is given by

$$d_(i,j) = |(x_{i,1} - x_{j,1})| + |(x_{i,2} - x_{j,2})| + ... + |(x_{i,n} - x_{j,n})| \tag{5.4}$$

or

$$d_(i,j) = \sum |(x_{i,k} - x_{j,k})| \tag{5.5}$$

6. Experimental Setup. We have considered 25 users and the corresponding rating values from the E - Commerce, the chosen values are plotted in scattered graph shown in the below Figure 6.1 and they are classified using clustering technique based on the nearest neighbor concept. Clustering is defined as making groups with the existing data, to obtain a new form of clusters or groups.

This results in achieving the similarity of the data in same group and dissimilarities when compared to the data values from one grouped values to another. Often, distance measures are used, an appropriate arbitrary point was chosen from cluster and two methods of calculating the distance between every other point from an arbitrary point (i.e., Euclidean distance method and Manhattan distance method) was performed, this is achieved by using multiple iterations, and the multiple bar graph are plotted. In this experiment we use (4) Euclidean distance method and (6) Manhattan distance method to find the distance between them.

The proposed system is contributing in classifying the similar types of the users based on the rating values. Table 6.1 represents the resulting values of distance measure. The proposed system is developed using collaborative filtering approach which helps e-commerce to find the similar likings of the users based on their reviews and reactions, and the appropriate suggestion are filtered. Subsequently, we find the subset of the group of the users from the bigger data sets that are collected from the e-commerce.The above Figure 6.2 represents the similarity between clusters and ratings.

Table 6.1: Comparison values

	Euclidean Distance K-Means	Manhattan Distance K-Means
c1	3.08	5
c2	1.22	2
c3	1.87	3.5
c4	1.22	2
c5	2.78	4.5

Fig. 6.2: Cluster Vs Rating

Here the datasets which was gathered from e-commerce is the set of sparse data; thus this does not support in achieving the appropriate recommendations to the users. Therefore, as the first step we convert all the sparse data into a distinguishable unique values and continue to calculate the similarity between them, and to measure the strength of the similar users we use distance measure method to determine the correlation between the users and the products.

This approach of finding the similarity between the users is a reliable method if the datasets that are obtained are sparse data. The methods like cosine similarity and distance measure methods which are implemented on the sparse data produced desired results. This method of making recommendations are more reliable and attainable for sparse data samples. This proposed system works well for sparse data and produces the reliable results.

7. Limitations of the work proposed and their future improvement.

7.1. Limitations. For this proposed system, the data which is gathered from the e-commerce is more often the sparse dataset. Thus, determine the missing value and averaging the datasets become the mandatory steps to be carried out to find the similarity between the users. In this research paper the recommendations are done only based on the ratings given by the users and neglecting other parameters.

7.2. Future Work.

a. For the aim of making recommendations, we'll strive to combine the genre dimension with other product dimensions (reviews, reactions, etc.). It will improve how effective our suggestion is.
b. Expand the user and product databases respective data sizes.
c. Use additional collaborative filtering methods to compare recommendations.

8. Conclusion. In this research paper, we have developed a recommendation system for e-commerce based on rating system. We used singular value decomposition method as collaborative filtering technique.

This approach helps us in identifying the similarities between the user and make relevant suggestions of the product. We gather vital data, such as user evaluations provided by e-commerce, which are necessary for recommendations. Cosine similarity is utilized to determine how similar the users are to one another for sparse data. For the non-sparse dataset, we assess the distance between users and depict the graph using the Euclidian and Manhattan distance methods, which guarantees that their tastes and preferences are similar. This approach to giving recommendations is more trustworthy and practical.

REFERENCES

[1] Utkarsh Gupta, Dr Nagamma Patil, "Recommender System Based on Hierarchical Clustering Algorithm Chameleon" IEEE International Advance Computing Conference (IACC) 2015.
[2] Xiaoyuan Su and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques Advances in Artificial Intelligence Volume 2009, Article ID 421425, 19 pages, October 2009.
[3] Harpreet Kaur Virk, Er. Maninder Singh," Analysis and Design of Hybrid Online Movie Recommender System "International Journal of Innovations in Engineering and Technology (IJIET)Volume 5 Issue 2,April 2015.
[4] Geetha G, Safa M , Fancy C, Saranya D4, "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System" National Conference on Mathematical Techniques and its Applications (NCMTA 18) 2018.
[5] Sarwar B., Karypis G., Konstan J., Riedl J., "Item-based Collaborative Filtering Recommendation Algorithms," Published in the Proceedings of the 10th international conference on World Wide Web, Hong Kong, May 15, 2001.
[6] C. S. M. Wu, D. Garg, and U. Bhandary, "Movie Recommendation System Using Collaborative Filtering," In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), pp. 11-15, IEEE, 2018 November.
[7] D. M. Pennock and E. Horvitz, Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach, In IJCAI Workshop on Machine Learning for Information Filtering, Stockholm, Sweden, August 1999.
[8] Debadrita Roy, Arnab Kundu, "Design of Movie Recommendation System by Means of Collaborative Filtering" International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013.
[9] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative Filtering Recommender Systems" The Adaptive Web, LNCS 4321, pp. 291 – 324, Springer-Verlag Berlin Heidelberg 2007.
[10] Desrosiers C, Karypis G, "A comprehensive survey of neighborhood-based recommendation methods". In: Ricci F, Rokach L, Shapira B, Kantor P, Recommender systems. Springer, Boston, pp 107–144, 2011.

Edited by: Vinoth Kumar
Received: Jun 11, 2022
Accepted: Oct 10, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 181–191, DOI 10.12694/scpe.v23i4.2018

HYBRID HYPER CHAOTIC MAP WITH LSB FOR IMAGE ENCRYPTION AND DECRYPTION

S. JAHNAVI *AND C. NANDINI †

Abstract. There are number of images that transmitted through the web for various usages like medical imaging, satellite images, military database, broadcasting, confidential enterprise, banking, etc. Thus, it is important to protect the images confidentially by securing sensitive information from an intruder. The present research work proposes a Hybrid Hyper Chaotic Mapping that considers a3D face Mesh model for hiding the secret image. The model has a larger range of chaotic parameters which are helpful in the chaotification approaches. The proposed system provides excellent security for the secret image through the process of encryption and decryption. The encryption of the secret image is performed by using chaos encryption with hyper hybrid mapping. The hyper hybrid mapping includes enhanced logistic and henon mapping to improve the computation efficiency for security to enhance embedding capacity. In the experiment Fingerprint and satellite image is used as secret image. The secret image is encrypted using a Least Significant Bit (LSB) for embedding an image. The results obtained by the proposed method showed better enhancements in terms of SNR for the 3D Mesh model dataset as 77.85 dB better compared to the existing models that achieved Reversible data hiding in the encrypted domain (RDH-ED) of 33.89 dB and Multiple Most Significant Bit (Multi-MSB) 40 dB. Also, the results obtained by the proposed Hybrid Hyper chaotic mapping showed PSNR of 65.73 dB better when compared to the existing Permutation Substitution and Boolean Operation that obtained 21.19 dB and 21.27 dB for the Deoxyribonucleic Acid (DNA) level permutation-based logistic map.

Key words: Decryption, Encryption, Hybrid Hyper Chaotic map, Least Significant Bit, 3D Mesh model.

AMS subject classifications. 68P25

1. Introduction.

Information security is a part of today's world and images are the common form of multi-media on the internet [1]. Security is one of the significant issues for information transmission through the network [2]. The main aim of the watermarking technique is to perform image information conversion to keep it confidential for the process of encryption by authorized persons [3]. The encryption has the ability for recovering the original data without losing important things in it. Encrypted Secret information is sent through the internet or the wireless networks through the multi-media for a better secure data transmission over distinct communications channels [4, 5]. In the existing research, chaos-based image encryption is utilized that has robust properties with sensitivity, and unpredictability towards the initial dependent conditions [6]. The computational systems are dependent on the internet which utilized watermarking techniques [7]. The Chaos based image encryption has robust properties like unpredictability, and sensitivity towards the conditions [8]. The deterministic conditions in the chaos signify the random behavior with a deterministic system [9].

The existing models have used encryption techniques that includedigital image encryption based on digital image encryption with respect to the random sequences. Digital image encryption is based on the process of image compression coding and the image key. Chaos technology is difficult in cracking the randomness which has made the digital image encryption reliable. Chaos encryption technology refers to the higher dimensional space when proposed by researchers. The researchers faced the problem of encryption and faced difficulty to process the low efficiency of the decryption and encryption process.

. The contributions of the research work are given as follows:

- To develop a hybrid hyper Chaotic Mapping for encrypting the images that consisted of chaos encryption with Enhanced logistic map and henon map.

*Visvesvaraya Technological University, Computer Science and Engineering Department, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka 560082, India (jahnavishankar.s@gmail.com).

†Professor and Head, Computer Science and Engineering Department, Dayananda Sagar Academy of Technology and Management, India (hodcse@dsatm.edu.in)

- To embed the secret image in a cover image using Least Significant Bit (LSB)for hiding the confidential data on images.

The organization of the research paper is shown as follows. Section 2 explains about literature review of the existing methodologies. Section 3 presents about proposed Hybrid Hyper Chaotic map with LSB for image encryption and decryption. Section 4 illustrates the results and discussion. The conclusion of this research work is given in Section 5.

2. Literature Survey. Ting Luo et al. [11] developed the novel Reversible Data Hiding Method for the 3D Model based on the process of homomorphic encryption. The homomorphic Paillier cryptosystem was used to perform the 3D model encryption. The greedy algorithm was used for data hiding to classify the 3D model vertices that were referred and were embedded to increase the capacity. The embedded vertex has computed based on the reference vertex to predict the module length to generate the prediction error for embedding the data. However, visual quality reduced significantly as the embedding capacity was high.

Wei Zhang et al. [12] developed a CNN-CapsNet for the process of image scene classification for remote sensing. The Capsule network (CapsNet) was used as the proposed model which mainly performs the grouping of neurons as vectors or capsules that replaces with the neuron. The traditional neural network encodes the properties and spatial information of image features achieved equivariance showed an active area for classification. The CapsNet utilizes the capsules or a neuron group or vector that was used for replacing the neuron using a traditional neural network. The spatial properties encode the information of features in an image achieved equivariance showed improvement in the classification. The CNN- CapsNet improved scene classification by using two models such as CapsNet and CNN. The feature maps were used from only one CNN model that was different from retrained CNN

Ferhat Ozgur Catak et al. [13] utilized fully homomorphic encryption and parallel computation for the process of privacy preservation based on biometric data matching. The main aim was to use fully homomorphic encryption based on biometric matching to control the borders. The authentication for the biometric system was performed based on the hash expansion to encrypt the homomorphic features. The homomorphic encryption method showed significant drawbacks with respect to the execution of time. The matching system's deficiency suggested that the model consumed time during fingerprint matching in the encrypted domain. However, the identification of large fingerprint images required computing resources, processing capabilities, and storage.

Wanli Lv et al. [14] developed a Reversible Data Hiding for performing Encryption for the 3D mesh models. The multiple Most Significant Bit (Multi-MSB) was used for reversing the space that was used adaptively for embedding the secret message. The auxiliary information was compressed using arithmetic coding to further free up the redundant space on the 3D mesh models. The developed majority Voting system was the main principle for restoring the mesh model with higher quality. The developed model failed to make data extraction and mesh recovery as they are independent of each other. However, the decrypted cipher text obtained with plain text contained a secret message which could not separate in the clear domain as it failed to perform reversible recovery during data extraction.

Nashwan Alsalam Ali [15] developed a 3D Polygon Mesh Encryption model that maps on3D Lorenz Chaotic Map. The 3D polygon mesh model protects the encryption process using the 3D Lorenz Chaotic map. The developed model provided diffusion better by using an excellent property based on Hausdorff Distance (HD). The histogram metrics were evaluated to adopt them for calculating the matching degree between the original and the extracted model. However, the results were required to be analyzed based on histogram and Hausdorff which encrypted 3D model from different original models.

Xiaojun Lu [16] developed an Adaptive Weight Method to perform the image retrieval using multi-feature fusion. The model used an adaptive weighting method to extract the single feature trust that extracted the unsupervised features. Then, the transfer matrix was constructed based on the trust and the transfer matrix constructed was based on the weight of the single features that are obtained with several iterations. The image decomposition was performed for the process of image classification that considered the image patches achieved results. The key idea behind image decomposition was to perform the process of classifying the image patches. The model extracted a better image description for the retrieval process that showed significance for improving the retrieval performances.

Kaimeng Din et al. [17] developed a Multi-Scale U Shaped Chained M-Shape CNN model to perform

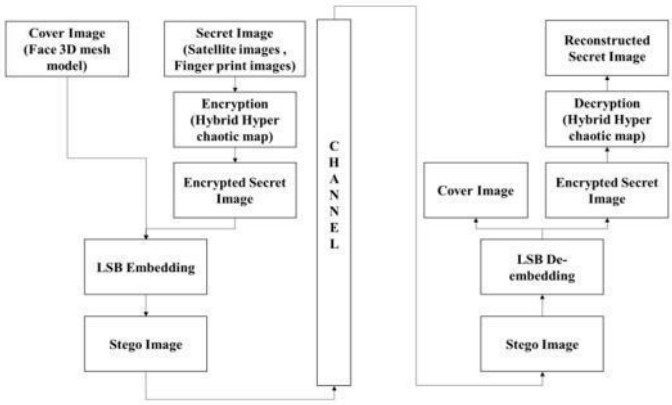


Fig. 3.1: Block diagram of the proposed method

the process of authentication for the generation of high-resolution remote sensing images. The developed model utilized a subject-sensitive perceptual hash function for generating a special case for a conventional hash function. The developed model was based on the subject-sensitive perceptual hash for achieving the CNN architecture extracted the robust features based on the high-resolution Remote Sensing (HRRS) images. Yet, the perceptual hash algorithm used with respect to different resolution images failed to locate tampered areas using deep learning approaches.

Abhimanyu Kumar Patro and Bibhudendra Acharya et al [18] developed a dual-layer cross-coupled chaotic map as an effective model to perform the process of image encryption. The developed model differed from the image encryption schemes multiple times due to two-layered cross coupled chaotic map to perform the process of permutation diffusion operations. The process of left to right flipping, block shuffling, and bit XOR performs the diffusion operation which were carried out in the 1^{st} layer based on the cross coupled chaotic map. Yet, the computational complexity was created in the developed model which was higher compared to the existing models.

Ebrahim Zarei Zefreh et al. [19] developed a hybrid model that consists of DNA computation, chaotic systems, and hash functions for performing image encryption. The developed model performed permutation for the mapping function based on logistic map that was applied for the randomly generated DNA image. The model changed the positions of the elements in the DNA image whichproved the efficiency of the developed scheme. The proposed model analyzed its security and showed improvement in its performance. However, the developed model achieved better security resulting in image encryption schemes that sufficiently showed a faster process for its applications.

Tahir Sajjad Ali and Rashid Ali et al. [20] developed Chaos-based image encryption that consisted of Boolean operation and permutation substitution. Chaotic theory was used to analyze the randomness and unpredictable behaviors of the image encryption process. The Boolean operation and permutation substitution was performed for the RGB components that showed better results for security and analyzing the performances. The model still required security and to be implemented on the real time image for the process of encryption transmission applications.

3. Proposed Methodology. The block diagram of the proposed research is shown in figure 3.1 which consists of a cover image that includes 3D mesh model images and satellite images. The finger print is used to authenticate and improve the security accessing the secret image. At the identification stage, secret image is authenticated by using a biometric identifier.

3.1. UC Merced Land Use Dataset. There are a total of 21 class land-use image datasets used for various purposes. The image has a dimension of 256 $\times$256 pixels that are extracted manually from large scaled

Fig. 3.2: Input images

images from the USGS National Map Urban Area Imagery collection. It is used in various areas from the urban side of the country.

3.2. Sokoto Coventry Fingerprint Dataset (SOCOFing). Sokoto Coventry Fingerprint Dataset (SOCOFing) is a biometric fingerprint database for research purposes. The SOCOFing used 6000 fingerprint images from the 600 African subjects that contained a distinct set of attributes that includes hand, gender, and finger name that are altered synthetically. The versions were provided with 3 different levels of alteration for central obliteration, z-cut, and central rotation, which is shown in Figure 3.2.

3.3. Chaos Encryption with Hybrid Hyper Chaotic Mapping. Once the digital images were collected, the process of encryption of the secret image is performed using the logistic map and henon mapping for performing Chaos encryption for the process of henon map hybridization. Chaos encryption is used extensively for reputation among researchers as it considers the inherent features of the chaos system. The chaos image cryptosystem has two main phases: diffusion and permutation. At the diffusion phase, each of the pixel values are altered and applied with the chaos sequences. The present research performs hyper chaotic mapping for enhancing the digital chaotic maps that are evaluated in terms of chaoticity. The statistical properties have contributed to the chaos based cryptography improvement. The bit reversal approach is used for addressing the issues and the proposed chaotic model modifies the values that are represented with fixed numbers for reversing these fractional bits' order.
With the permutation phase, the permutation for the image pixel positions are performed which overcome the scrambled time over the image without the image pixel value distribution. The diffusion phase and permutation phases used the keys K_i that generated the henon maps and logistic values. The Chaos encryption utilizes the simplest chaotic maps that are called henon map which is mathematically provided in the equation (3.1). The polynomial mapping of two degree uses logistic map that performs chaotic behaviour. The expression for the logistic map is provided in the Eq. (3.1)

$$l_{map}{=}\mu x_n\,(1{-}x_{n-1}) \tag{3.1}$$

where x_n is known as the Chaos sequence that ranges between [0,1] where the term μ is called as the control parameter which is ranging between $(3.57, 4)$.

The Modified Logistic Map's chaotic behavior evaluates the model which highlighted it effectively. The proposed method was analyzed included Bifurcation Diagram (BD), Lyapunov Exponent (LE), Shannon Entropy (SE), Correlation Dimension (CD), Correlation Coefficient (CC), Phase Diagram, and Approximate Entropy (ApEn).

Bifurcation diagram The present research uses Bifurcation diagrams for distinguishing between chaotic and non-chaotic regions. The logistic map showed limited values for parameters that suffered from the periodicity values indicated in unshaded regions.

Lyapunov Exponent (LE) Similarly, the LE is having the desired characteristic for the chaotic map which showed a desirable application under cryptography. The model showed slight changes in the control parameter conditions. The positive LE value is represented the chaotic behavior whereas, the larger LE values show better characteristics using chaotic mapping.

Shannon entropy The metrics that have been used for measuring the randomness in time series are due to chaotic trajectories. Therefore, the chaotic map is evaluated based on the SE that lies in the range of (0,1) that are divided into 210 partitions. The SE calculates the trajectories with the length of 215 that has been used for controlling the parameters. The ideal SE value is equal to 10, which shows higher randomness for a uniformly distributed chaotic trajectory that had visited equal partitions and showed better ergodicity.

Correlation coefficient (CC) The CC is used for measuring the relationship among the chaotic trajectories based on two experimental settings. The CC is used to calculate 2 trajectories that is having the initial conditions at an extreme. These are closer to each other as it has the same control parameters. Thus, the control parameters that are closer to one another starting with trajectories from the initial conditions are considered.

Correlation dimension (CD) The CD is used to observe the geometric complexity of the chaotic attractor that estimates the dimension level for a fractal. The CD measures the attractor with strangeness for the chaotic map. The CD showed a high value which implied a chaotic trajectory that has a phase space moved with the high fractal dimension. The trajectory has lies with the strange attractor that has high irregularity, and unpredictable behaviors that are suitable for cryptography.

Phase diagram The phase diagram reveals more information related to chaotic behaviors which are based on the one map iterations to investigate. The degree of complexity is based on the chaotic attractor that is dependent on the confusion capability with respect to each of the iterations. The entire phase is visited at the space and all these maps are successful in achieving ergodicity. The predictable curve can map the susceptible attacks like a return map to perform the process of signal optimization. The parabolic curve has leaked the information which is used for controlling the parameters for finding the critical points based on the phase diagram.

Approximate entropy The ApEn is used to measure the complexity among the orbits that are generated with distinct chaotic maps. The probability of chaotic orbits has demonstrated a new pattern thatincreases with embedding dimensions. The ApEn values evaluate the same chaotic maps which are set using the distinct parameter has the same bit precision values. The chaotic system showed better values of sensitivity for the initial conditions when the parameter reached 4. The henon map has two dimensional reversible non-linear chaotic maps which were iteration with the point as (x_n, y_n) mathematically expressed as shown in the Eq. (3.2).

$$h_{map}= 1-ax_n^2+y_n, \quad y_{n+1}=bx_n \tag{3.2}$$

From the above Eq. (3.2), $a\epsilon\,(0, 1.4)$, $b\epsilon(0.2,\ 0.314)$ which are known as the control parameters that are working using henon map which is dependent on the parameter values. The logistic and henon map are hybridized five times for accomplishing the transient effect based on the parameter value of the fingerprints as keys represented as K_i. This type of chaotic orbit is obtained from the previous step that is permutated with the diffusion phase on the plane image using Eq. (3.3) and (3.4).

$$x_{n+1}=l_{map}+h_{map} \tag{3.3}$$

$$mim\,(i) = permut\bigoplus K_i(x_{n+1}p\,(i)\,),\ i= 1, 2, 3, \ldots .p\times q \tag{3.4}$$

From the above Equation, p is known as the width and q is known as the height of the plane image, $p(i)$represents the pixel value of an original image, $mim(i)$called as the pixel value of an image. At last,

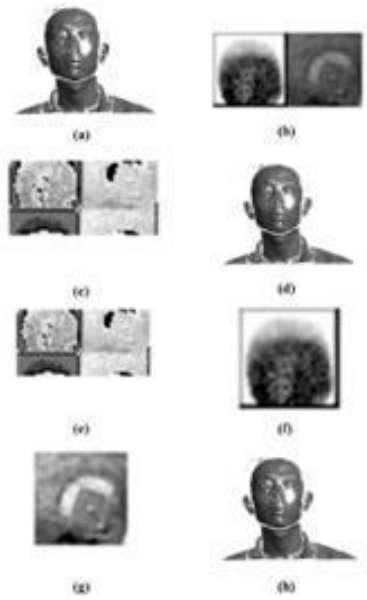


Fig. 3.3: 3D images (a) Cover image, (b) secret image, (c) Encrypted secret image, (d) Stego image, (e) Retrieved image, (f) Reconstructed fingerprint, (g) Reconstructed satellite image, (h) Reconstructed 3D image

the fingerprint image operates with a permutation process which uses an image generated by performing the process of diffusion. It is performed on the permuted image by using the hybrid chaotic model. At the diffusion stage, the output generated is cipher-image that is represented as $c(i)$ mathematically expressed as shown in the Eq. (3.5).

$$c(i) = mim(i) \ , \ i = 1, 2, 3, \ldots p \times q \tag{3.5}$$

3.4. Embedding process and Extraction Process using Least Significant Bit and Decryption Phase. Once the secret image is obtained image in the encryption process, with the cover image transformation, the embedding process is carried out to hide the secret image. The secret image performs the encryption process that considers the cover image transformation to embed the process as it is carried out. The LSB is called the bit operator for the integer positions that embed the unit value. The minimum weighting value is replaced with the sampled image pixels with the binary bits to provide information from the secret data which is hidden inside the pixels. The main purpose is to extract secret information from the locations. Thus, the increase in detection accuracy faces difficulty for the secret data. The pseudo random sequence is used for controlling the location of the secret binary information which is embedded. The LSB model is simple and easy to implement. The model is embedding and extracts the information with a higher rate of hiding capacity faster. The size of the secret image is $m \times n/8$ where the size of the cover image is represented as m×n. The 8 bit LSB has the cover image bit value which is exchanged with the secret image bit value. Thus, the secret image embeds the cover image and the values obtained are binary are converted to the decimal numbers. At the receiver side, the LSB operations are performed on the stego image with the hybrid hyper chaotic mapping. The Chaos decryption with hybrid mapping is performed based on the generated cipher text. The extraction of the secret image was limited with no loss in information and thus the secret image is the same as the original secret image, which is shown in figure 3.3 and 3.4.

4. Results and Discussion. The experimental simulation is performed using a MATLAB (2018a) environment that works with 8 GB RAM, i9 3.0 GHz processor, and 3 TB memory. Mathematical equations of

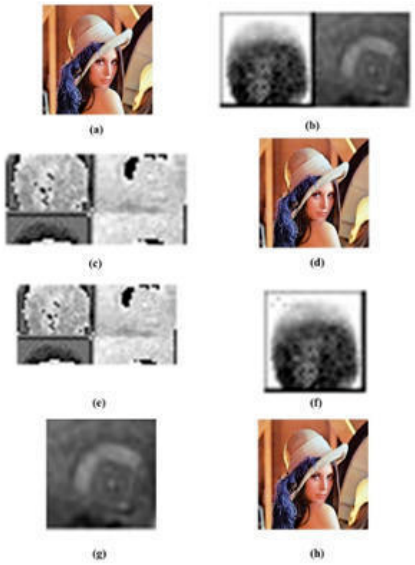


Fig. 3.4: Lena images (a) Cover image, (b) secret image, (c) Encrypted secret image, (d) Stego image, (e) Retrieved image, (f) Reconstructed fingerprint, (g) Reconstructed satellite image, (h) Reconstructed Lena image

entropy value, PSNR, UACI, SSIM, and NCC were indicated in the Eq. (4.1–4.6).

$$E\left(m\right)=\sum_{x=0}^{m-1}p(m_x)log_2\frac{1}{p(m_x)} \tag{4.1}$$

where $p(m_x)$ represents the probability of symbol occurrence and m is known as the total number of symbols represented as $m_x \epsilon m$.

$$Peak\ signal\ to\ noise\ Ratio\ (PSNR)=10\ log_{10}\left(\frac{255^2}{MSE}\right) \tag{4.2}$$

$$Mean\ Square\ Error\ (MSE)=1/pq\sum_{x=0}^{p-1}\sum_{y=0}^{q-1}\left[I\left(x,y\right)-k\left(x,y\right)\right]^2 \tag{4.3}$$

where p and q are known as the row and the column of an image. $k(x,y)$ is known as the decrypted image, and $I(x,y)$ is known as the original input image.

$$Unified\ Average\ Changing\ Intensity\ (UACI)=\frac{1}{pq}\sum_{x=1}^{p}\sum_{y=1}^{q}\frac{\left|E_1\left(x,y\right)-E_2\left(x,y\right)\right|}{255}\times 100 \tag{4.4}$$

$$Structural\ SIMilarity\ SSIM\left(x,y\right)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{4.5}$$

$$Normalized\ Cross\ Correlation\ (NCC)=\left[\frac{\sum_{x=1}^{p}\sum_{y=1}^{q}\left[I\left(x,y\right)k(x,y)\right]}{\sum_{x=1}^{p}\sum_{y=1}^{q}\left[I(x,y)\right]^2}\right] \tag{4.6}$$

Table 4.1: Results obtained by the proposed method for 3D mesh model used as cover image

Images	PSNR	MSE	NCC	AD	SSIM	NAE
Reconstructed 3D	49.008	0.818	0.999	0.171	0.997	0.008
Stego Image	47.635	1.131	0.993	0.324	0.999	0.009

Table 4.2: Results obtained for the existing approaches for performing encryption

Images	PSNR	MSE	NCC	AD	SSIM	NAE
Proposed (Finger print)	99.007	0.002	0.998	0.004	0.995	0.003
Existing (Finger print)	17.617	1127.8	0.577	1.751	0.888	0.242
Proposed (Satellite)	99.008	0.007	0.992	0.003	0.997	0.009
Existing (Satellite)	10.219	6193.2	0.633	4.980	1.090	0.427

Table 4.3: Results obtained for the proposed method

Images	PSNR	MSE	NCC	AD	SSIM	NAE
Without double encryption i.e only logistic method (Fingerprint)	11.964	4136.6	0.743	29.94	1.550	0.355
Without double encryption i.e only logistic method (Satellite)	14.234	2456.8	0.469	1.674	0.809	0.432

Table 4.4: Results obtained for the proposed research work evaluated for different types of datasets such as Fingerprint, Satellite Reconstructed, 3DImg, and Stego 3DImg

Images	PSNR	MSE	NCC	AD	SSIM	NAE
Fingerprint	99.008	0.008	0.994	0.008	0.99	0.005
satellite	99.001	0.001	0.994	0.002	0.997	0.007
Reconstructed 3DImg	49.381	0.752	0.994	0.15	0.995	0.005
Stego 3DImg	48.027	1.026	0.998	0.29	0.992	0.002

where x *and* y is known as the windows from the filter represented as 'k', I is called the original image, σ *and* $\mu$are called as the standard deviation that is having x and$y$ as mean, c_1 *and* c_2 are represented as the constants. $E_1 and\ E_2$ are indicated as encrypted images.

Average Difference (AD) calculates the difference among two adjacent frames on a set of images.

Signal to Noise Ratio (SNR) is defined as the ratio of signal power to the noise power, often expressed in decibels.

4.1. Quantitative Analysis . The proposed method results are compared with the previous researches by comparing with the previous researches that uses reconstructed 3D and Stego image for validating the results.

In table 4.1, the results obtained by the proposed method are shown in terms of PSNR values which are obtained for the reconstructed 3D image. This has obtained 49.008 dB of PSNR values and Stego image of 47.635 dB. The MSE values for the 3D reconstructed image are obtained as 0.818 and 1.131. The value of NCC is obtained as 0.999 and 0.993 for the reconstructed and Stego image. The Average Difference for the reconstructed 3D image is obtained as 0.171 and for that of the Stego is obtained as 0.324.

In table 4.2, the proposed and the existing method logistic are evaluated in terms of PSNR, MSE, NCC, AD, SSIM, and NAE. It is observed that the Proposed system yields better result comapred to exhisting system.

Table 4.3 shows the results obtained for the proposed and the existing methods evaluated in terms of PSNR,

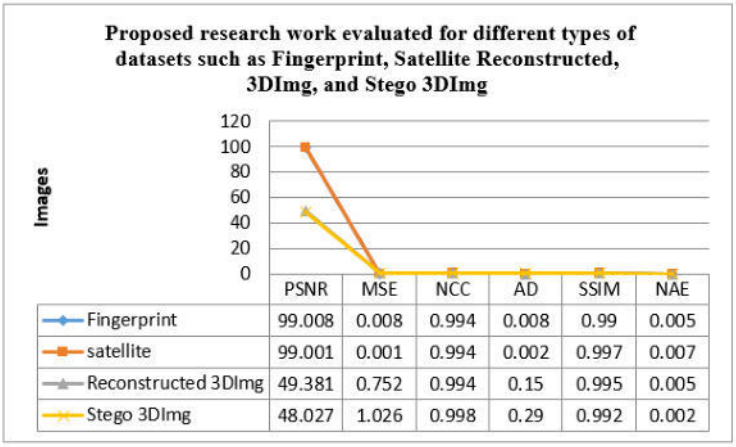


Fig. 4.1: Proposed research work evaluated for different types of datasets such as Fingerprint, Satellite Reconstructed, 3DImg, and Stego 3DImg

Table 4.5: Results obtained by the proposed research work (I)

Images	PSNR	MSE	NCC	AD	SSIM	NAE
Stego Image	65.73	0.02	0.99	0.01	0.99	0.00
Existing (logistic+henon)	29.44	74.04	0.99	0.47	1.00	0.01
Without double encryption (only logistic)	26.33	151.52	0.98	0.94	1.01	0.03

Table 4.6: Results obtained by the proposed research work (II)

Images	PSNR	MSE	NCC	AD	SSIM	NAE
Finger_print (Proposed)	55.77	7.26	0.71	0.31	1.00	0.01
Satellite(Proposed)	59.00	10.00	1.00	0.01	1.00	0.01
Satellite(Existing)	9.95	6588.95	0.66	3.33	0.99	0.45
Fingerprint(Existing)	17.21	1236.69	0.56	0.62	0.91	0.26
Without double encryption (Fingerprint)	8.61	8951.28	0.60	64.93	2.84	0.52
Without double encryption(Satellite)	13.00	3259.43	0.41	43.94	2.57	0.51

MSE, NCC, AD, SSIM, and NAE. The hybrid mapping algorithm consist of Logistc with henon map (as double encryption) and only logistic map as existing systems used for the evaluation of results for both the fingerprint and the satellite images. The Satellite images and the fingerprint images are used for the evaluation of the results of the existing and proposed approaches. The proposed method for fingerprint and satellite images obtained PSNR of 99.007 dB and 99.008dB better when compared to the existing approaches that obtained PSNR of 17.61 dB for fingerprint images and 10.21 dB for the satellite image.

Table 4.4 and figure 4.1 shows the results obtained for the proposed method without double encryption for satellite and fingerprint images. The PSNR values without double encryption uses logistic mapping for encryption, it obtains 11.96 dB and 14.23 dB for the satellite image. The value of MSE for the proposed method without double encryption for the fingerprint image and satellite image is is 4136.6.

Table 4.5 shows the results obtained for the proposed research work which validates the results for various types of datasets such as fingerprint images, satellite images, reconstructed 3D images, and Stego 3D Images

Table 4.7: Results obtained by the proposed research work (III)

Images	**PSNR**	**MSE**	**NCC**	**AD**	**SSIM**	**NAE**
Fingerprint	55.77	7.26	0.71	0.31	1.00	0.01
Satellite	59.00	10.00	1.00	0.01	1.00	0.01
Reconstructed 3DImg	64.35	0.03	1.00	0.01	0.99	0.00
Stego 3DImg	65.73	0.02	0.99	0.01	0.99	0.00

Table 4.8: Comparative Analysis for 2 D images

Authors	**Method**	**PSNR (dB)**	**MSE**
K. Abhimanyu Kumar Patro and Bibhudendra Acharya [18]	Dual-Layer Cross-Coupled Chaotic Map	-	7764.3
Ebrahim Zarei Zefreh [19]	DNA level permutation-based logistic map	21.27	-
Tahir Sajjad Ali and Rashid Ali [20]	Permutation Substitution And Boolean Operation	21.19	497.39
Proposed method	Hybrid Hyper Chaotic mapping	65.73	108.7

Table 4.9: Comparative Results for 3D mesh images

Authors	**Methods**	**SNR (dB)**
Ting Luo [11]	Reversible data hiding in the encrypted domain (RDH-ED)	33.89
Wanli Lv [14]	Multiple Most Significant Bit (Multi-MSB)	40
Proposed	Hybrid Hyper Chaotic mapping	77.85

4.2. Comparative Analysis. Table 4.8 is the comparative analysis of the proposed and the existing models. The values of the PSNR and MSE are validated for existing methods for the standard MATLAB datasets. Table 4.9 shows the results obtained for the 3D mesh model images that are evaluated in terms of SNR (dB).

The existing model Dual-Layer Cross-Coupled Chaotic Map consumed more resources that resulted in 7764.3 MSE values. DNA level permutation based logistic map obtained 21.27 dB of PSNR values due to compatibility of techniques. Also, Permutation Substitution and Boolean Operation obtained PSNR of 21.19 dB and MSE of 497.39 as they used unrealistic requirements. Whereas, the proposed hybrid Hyper Chaotic mapping obtained PSNR of 65.73 dB and MSE of 108.7 of MSE.

5. Conclusion. The present research proposed a highly secured transmission network for real-time world applications. The hybrid map was implemented for ensuring the integrity of the data and the privacy of the secret image. The proposed hybrid hyperchaotic encryption was used for performing encryption of the images based on Chaos encryption with hybrid mapping that consisted of logistic and henon maps. Yet, improvement was required for the direct and indirect recursions which were performed by combining the hybrid chaotic mapping-based encryption technique. Compared to the existing research, the proposed method has used a huge size of images that reduced the time complexity. The results obtained by the proposed method showed better enhancements in terms of SNR for the 3D Mesh model dataset as 77.85 dB better compared to the existing models that achieved Reversible data hiding in the encrypted domain (RDH-ED) of 33.89 dB and Multiple Most Significant Bit (Multi-MSB) 40 dB. Also, the results obtained by the proposed Hybrid Hyper chaotic mapping showed PSNR of 65.73 dB better when compared to the existing Permutation Substitution of

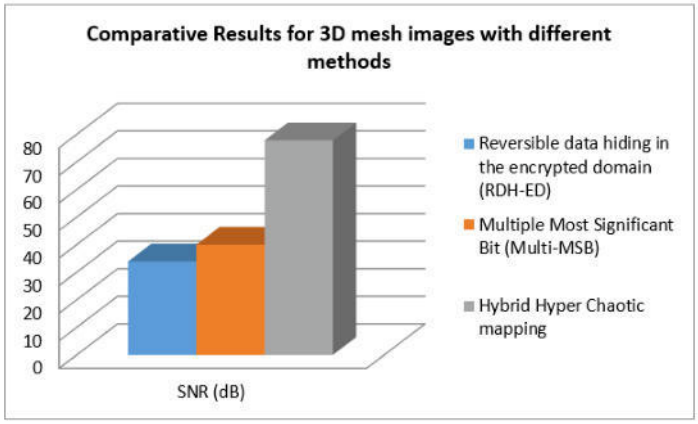


Fig. 4.2: 3D mesh model images evaluated in terms of SNR (dB) of different

21.9 dB and Boolean Operation of 21.27 dB.

REFERENCES

[1] Ali, T.S. and Ali, R., 2020. A new chaos based color image encryption algorithm using permutation substitution and Boolean operation. Multimedia Tools and Applications, 79 (27), pp.19853-19873.
[2] Gambhir, G. and Mandal, J.K., 2020. Multicore implementation and performance analysis of a chaos based LSB steganography technique. Microsystem Technologies, pp.1-11.
[3] Valandar, M.Y., Barani, M.J., Ayubi, P. and Aghazadeh, M., 2019. An integer wavelet transform image steganography method based on 3D sine chaotic map. Multimedia Tools and Applications, 78 (8), pp.9971-9989.
[4] Younus, Z.S. and Hussain, M.K., 2019. Image steganography using exploiting modification direction for compressed encrypted data. Journal of King Saud University-Computer and Information Sciences.
[5] Mahdi, M.H., Abdulrazzaq, A.A., Rahim, M.S.M., Taha, M.S., Khalid, H.N. and Lafta, S.A., 2019, May. Improvement of Image Steganography Scheme Based on LSB Value with Two Control Random Parameters and Multi-level Encryption. In IOP Conference Series: Materials Science and Engineering (Vol. 518, No. 5, p. 052002). IOP Publishing.
[6] Khan, M., Alanazi, A.S., Khan, L.S. and Hussain, I., 2021. An efficient image encryption scheme based on fractal Tromino and Chebyshev polynomial. *Complex & Intelligent Systems*, *7*(3.5), pp.2751-2764.
[7] T. Huynh-The, C.H. Hua, N.A. Tu, T. Hur, J. Bang, D. Kim, M.B. Amin, B.H. Kang, H. Seung, and S. Lee, "Selective bit embedding scheme for robust blind colour image watermarking", *Information Sciences*, Vol. 426, pp. 1-18, 2018.
[8] A.M. Abdelhakim, and M. Abdelhakim, "A time-efficient optimization for robust image watermarking using machine learning", *Expert Systems with Applications*, Vol. 100, pp. 197-210, 2018.
[9] S. Kumar, N. Jain, and S.L. Fernandes, "Rough set based effective technique of image watermarking", *Journal of Computational Science*, Vol. 19, pp. 121-137, 2017.
[10] A. Abbasi, C.S. Woo, and S. Shamshirband, "Robust image watermarking based on Riesz transformation and IT2FLS", *Measurement*, Vol. 74, pp. 116-129, 2015.
[11] Luo, T., Li, L., Zhang, S., Wang, S. and Gu, W., 2021. A Novel Reversible Data Hiding Method for 3D Model in Homomorphic Encryption Domain. *Symmetry*, *13* (6), p.1090.
[12] Zhang, W., Tang, P. and Zhao, L., 2019. Remote sensing image scene classification using CNN-CapsNet. *Remote Sensing*, *11* (5), p.494.
[13] Catak, F.O., Yayilgan, S.Y. and Abomhara, M., 2020. A Privacy-Preserving Fully Homomorphic Encryption and Parallel Computation Based Biometric Data Matching.
[14] Lv, W., Cheng, L. and Yin, Z., 2021. High Capacity Reversible Data Hiding in Encrypted 3D mesh models Based on multi-MSB Prediction. *arXiv preprint arXiv:2110.01010*.
[15] Ali, N.A., Rahma, A.M.S. and Shaker, S.H., 2021. 3D Polygon Mesh Encryption Based on 3D Lorenz Chaotic Map. *iJIM*, *15*, p.103.
[16] Lu, X., Wang, J., Li, X., Yang, M. and Zhang, X., 2018. An adaptive weight method for image retrieval based multi-feature fusion. *Entropy*, *20*, p.577.
[17] Ding, K., Liu, Y., Xu, Q. and Lu, F., 2020. A subject-sensitive perceptual hash based on MUM-Net for the integrity authentication of high resolution remote sensing images. *ISPRS International Journal of Geo-Information*, *9*, p.485.
[18] Patro, K. and Acharya, B., 2021. An efficient dual-layer cross-coupled chaotic map security-based multi-image encryption

system. *Nonlinear Dynamics*, *104*, pp.2759-2805.
[19] Zefreh, E.Z., 2020. An image encryption scheme based on a hybrid model of DNA computing, chaotic systems and hash functions. *Multimedia Tools and Applications*, *79*, pp.24993-25022.
[20] Ali, T.S. and Ali, R., 2020. A new chaos based color image encryption algorithm using permutation substitution and Boolean operation. *Multimedia Tools and Applications*, *79*, pp.19853-19873.

Edited by: Vinoth Kumar
Received: Jun 16, 2022
Accepted: Oct 16, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 193–210, DOI 10.12694/scpe.v23i4.2019

AN EFFICIENT NOVEL APPROACH WITH MULTI CLASS LABEL CLASSIFICATION THROUGH MACHINE LEARNING MODELS FOR PANCREATIC CANCER

P. SANTOSH REDDY *AND M. CHANDRA SEKHAR †

Abstract. Pancreatic cancer is right now the fourth largest cause of cancer-related deaths. Early diagnosis is one good solution for pancreatic cancer patients and reduces the mortality rate. Accurate and earlier diagnosis of the pancreatic tumor is a demanding task due to several factors such as delayed diagnosis and absence of early warning symptoms. The conventional distributed machine learning techniques such as SVM and logistic regression were not efficient to minimize the error rate and improve the classification of pancreatic cancer with higher accuracy. Therefore, a novel technique called Distributed Hybrid Elitism gene Quadratic discriminant Reinforced Learning Classifier System (DHEGQDRLCS) is developed in this paper. First, the number of data samples is collected from the repository dataset. This repository contains all the necessary files for the identification of prognostic biomarkers for pancreatic cancer. After the data collection, the separation of training and testing samples is performed for the accurate classification of pancreatic cancer samples. Then the training samples are considered and applied to Distributed Hybrid Elitism gene Quadratic discriminant Reinforced Learning Classifier System. The proposed hybrid classifier system uses the Kernel Quadratic Discriminant Function to analyze the training samples. After that, the Elitism gradient gene optimization is applied for classifying the samples into multiple classes such as non-cancerous pancreas, benign hepatobiliary disease i.e., pancreatic cancer, and Pancreatic ductal adenocarcinoma. Then the Reinforced Learning technique is applied to minimize the loss function based on target classification results and predicted classification results. Finally, the hybridized approach improves pancreatic cancer diagnosing accuracy. Experimental evaluation is carried out with pancreatic cancer dataset with Hadoop distributed system and different quantitative metrics such as Accuracy, balanced accuracy, F1-score, precision, recall, specificity, $TN, TP, FN, FP, ROC_{AUC}, PRC_{AUC}$, and PRC_{APS}. The performance analysis results indicate that the DHEGQDRLCS provides better diagnosing accuracy when compared to existing methods.

Key words: Pancreatic cancer diagnosis, Kernel Quadratic Discriminant analysis, Elitism gradient gene optimization, Reinforced Learning technique, Hadoop distributed system.

AMS subject classifications. 68T05

1. Introduction.

PDAC (pancreatic ductal adenocarcinoma) is still an incurable cancer that kills a lot of people. Despite breakthroughs in the identification and treatment of other gastrointestinal cancers, such as colorectal and gastric cancers, PDAC mortality rates only slightly exceed the number of newly diagnosed cases, with 5-year survival rates as low as 3-15 percent 3,4. These poor results are due to late and incurable stage diagnosis, as well as significant chemo-resistance in tumours[7][8]. Most treatment approaches are rendered useless due to the latter. The importance of early diagnosis is well recognised, and various healthcare groups have advocated for a move toward early detection. A lot of findings indicate the advantages of early PDAC detection.

Patients diagnosed with stage I illness have a higher chance of surviving than those diagnosed with later stages. Similarly, when compared to PDAC [2] detected when individuals have symptoms, incidentally diagnosed PDAC is related with longer median survival. Patients diagnosed at an operable stage have a significantly higher chance of survival than those detected at an inoperable stage. Unfortunately, up to 85% of instances are not surgically resectable when they are discovered, and in the United Kingdom, more than half of cases are detected after a non-specific illness course that results in an emergency hospital admission 10-13.

Biomarkers may play an important role in the early detection of PDAC by enriching for persons in high-risk groups who are more likely to develop cancer, allowing doctors to prioritize individuals for screening. Despite

*Department of Computer Science and Engineering, Presidency University, Bengaluru, India (santosh.reddyp@presidencyuniversity.in)

†Department of Computer Science and Engineering, B. N. M. Institute of Technology, Bengaluru, India (mchandrasekhar@presidencyuniversity.in)

thousands of articles, no one candidate biomarker for the early identification of PDAC has been translated into clinical application. Indeed, developing biomarkers for this disease has distinct hurdles. Due to the low prevalence of PDAC, obtaining the sufficient number of samples for biomarker development is difficult, necessitating extensive national and international collaborations. Pancreatic tumors are extremely diverse, both within and between people.

As a result, single biomarkers are unlikely to be sensitive enough to detect PDAC, and compre-hensive panels of biomarkers will be necessary. PDAC biomarker research requires an understanding of and accounting for any confounding factors68. The majority of important studies now include illness controls, such as chronic pancreatitis. 68 Furthermore, people are becoming more aware that obstruc-tive jaundice can contribute to false-positive biomarker results69-71. The fact that a large percentage of PDAC patients have diabetes mellitus (DM), 50 is not adequately accounted for in biomarker re-search, and developing biomarkers risk being associated with DM rather than PDAC.

Finally, a key limitation in early detection studies for PDAC has been the lack of specialized pre-diagnostic groups. When weighing the cost vs. value of a biomarker-assisted screening programme, take into account the costs of both the first screening and the subsequent tests required to confirm the diag-nosis. Since both genuine positive and false positive tests necessitate additional investigation, high spec-ificity biomarkers are required.

Pancreatic cancer is the fourth most prevalent cause of cancer death and the second most common cause of death from neoplasms that disrupt digestion. Regular pancreas segmentation, on the other hand, remains a point of contention for the following reasons: 1) CT scans with low soft tissue contrast. 2) Significant anatomical differences. In terms of size and placement in the abdominal cavity of patients, the pancreas is very unpredictable anatomically [6][9]. The pancreas is a flexible tissue that yields. As a result, the shape and appearance of the pancreas fluctuate significantly between individu-als.

1.1. Contributions of the paper. In order to solve the existing issues, a novel DHEGQDRLCS is introduced. The contribution of the proposed DHEGQDRLCS is listed below:

- To improve the pancreatic cancer diagnosis accuracy, a novel DHEGQDRLCS technique is introduced as a reliable diagnostic tool to improve the clinical practicality for diagnosing pancreatic cancer early based on hybridization of Kernel Quadratic Discriminant Function, Elitism gradient gene optimization, and Reinforced Learning technique.
- A Kernel Quadratic Discriminant Function is applied to the DHEGQDRLCS technique for analyzing the correlation between the data samples and the class means value by using the kernel function. Then the Elitism gradient gene optimization technique finds the maximum correlated results for classifying the samples into a particular class. After that, reinforcement learning is applied to find the minimum loss function. This process increases the accuracy and minimizes incorrect pancreatic classification.
- Extensive experimentation is conducted with Hadoop distributed system to measure the performance of the DHEGQDRLCS technique and other related works. The obtained result shows that our proposed, DHEGQDRLCS technique provides better performance than the existing Distributed eSVM (DeSVM) and Distributed eLR (DeLR) Models.

2. Literature Survey. In a recent analysis of 3.9 million cancer patients from seven countries (Australia, Canada, Denmark, Ireland, New Zealand, Norway, and the United Kingdom) assessing seven cancer sites (esophagus, stomach, colon, rectal, pancreas, lung, and ovary), PDAC was found to have the lowest 5-year survival rates (ranging from 7.9 percent in the UK to 14.6 percent in Australia). 4 Early discovery will undoubtedly aid in the improvement of these statistics, and as our review demonstrates, progress has already been made. The development and validation of biological and epidemiological markers will benefit from the construction of new customized cohorts (of people with NOD or symptoms).

Careful and ethical use of existing data, whether via social media or electronic health records, can help prediction models, while AI applied to imaging can help discover lesions early. Identifying the small number of people with MCLs who are at the highest risk of developing PDAC is still a major knowledge gap in the field of mucinous cysts. Much work, including the ongoing trials discussed here, has to be done to enhance the early detection of PDAC. The ongoing cohort studies are critical in many ways, not least because they help to raise awareness of PDAC symptoms and their relation to NOD among healthcare practitioners and patients.

Early detection advances will be combined with therapy advances to help people with PDAC live longer. According to the concept of regular automation algorithms, Support Vector Machine is an urgent classification algorithm for categorising data connected to the calculation of Wisconsin Breast Cancer data in a short amount of time [10]. Proportion of relative outcomes for four alternative algo-rithms for data retrieval and automatic automation in terms of efficacy and effectiveness. Initiates a new function for the benefit of the medical health system, with the purpose of predicting an average patient's outcome in the analysis of electronic medical processes and the recognized parameters of parameters developed for optimal operation [11].

The application coordination with the estimation of data for variable effects, categories of ef-fects, and the threshold parameter to identify the disease diagnosis generally provides efficient prognos-tic data. They use and cover medical proceedings for blood cancer, heart failure, and diabetes [12]. Using data from combination radiation (EBRT) and brachytherapy (BT), construct a function based on the red convolution neural representation to analyse rectal prescribed amount sharing and forecast rectal toxicity in patients with uterine cancer [13]. To have an impact on patient data, they established a place to live and transfer approach.

To increase the dates for video data losses and loss factors, the adaptive synthetic model technique is applied. To construct RSDM discriminate regions with the calculation model, create Gradi-ent Activation Weight Map (Grad-CAM) classes[19][20]. A combination of experimental results is used to examine the CNN-based representation for predicting rectal dose by means of transfer therapy for uterine cancer radiation.

A novel machine learning technique with the twin support vector machine (TWSVM) was developed in [21] (i.e. distributed extended SVM) for identifying pancreatic cancer early. However, the higher pancreatic cancer detection accuracy was not achieved. Unconditional and conditional logistic regression models were developed in [22] (i.e. Distributed extended LR) for pancreatic cancer diagnosis by measuring the relationship between fasting glucose levels and pancreatic cancer risk. But, the higher precision and recall analysis was not achieved.

A Fully Automatic Deep Learning Framework was developed in [23] for Pancreatic Ductal Adenocarcinoma Detection. But the Framework was not efficient to perform the multiclass classification. Multi-Omics DEep Learning for Prognosis-correlated subtyping (MODEL-P) was developed in [24] to identify Pancreatic Ductal Adenocarcinoma Pancreatic But it failed to improve the robustness of the deep learning model while handling more samples for pancreatic cancer detection. Three classification algorithms namely linear discriminant, analysis (LDA), support vector machine (SVM), and k-nearest neighbor (KNN) were developed in [25] for the classification of benign and malignant pancreatic tumors. However, the designed algorithms failed to establish a more reliable classification model for the accurate detection of pancreatic cancer.

Different Blood biomarkers classes were analyzed in [26] for differential diagnosis and early detection of pancreatic cancer. But the analysis was not considered with a large number of samples. A grouped" neural network (GrpNN) architecture was designed in [27] to generate a dimensionally reduced vector for early detection of pancreatic cancer. But it failed to apply to multi-modal clinical data sets. A combinatorial approach consisting of Particle Swarm Optimization (PSO), Artificial Neural Network (ANN), and Neighborhood Component Analysis (NCA) iterations was developed in [28] for pancreatic cancer diagnoses. But, the performance of pancreatic cancer detection was not improved.

3. System architecture. The system architecture is the conceptual model that describes the structural views. A system architecture also consists of the number of processes involved and works together to implement the overall system.

Figure 3.1 illustrates the system design architecture. First, the input is collected from the corresponding dataset. Then the input data gets preprocessed. Pre-processing is an essential step in the data mining process for transforming the raw data into a structured format that helps to improve accuracy and minimize time complexity. Next, data separation is performed to split the preprocessed dataset into the testing and training data. In machine learning, data separation is used to avoid overfitting. It is a modeling error that occurs when a function is too strongly fit to a limited set of data points. Then the testing and training samples are given to the machine learning techniques such as Distributed Extended Support Vector Machines (i.e.: Distributed eSVM), Distributed Extend Logistic Regression Model, and Distributed Hybrid Elitism gene Quadratic discriminant Reinforced Learning Classifier System. These techniques train a model on known input data and provide future outputs. The processes of three machine learning techniques are described as given below.

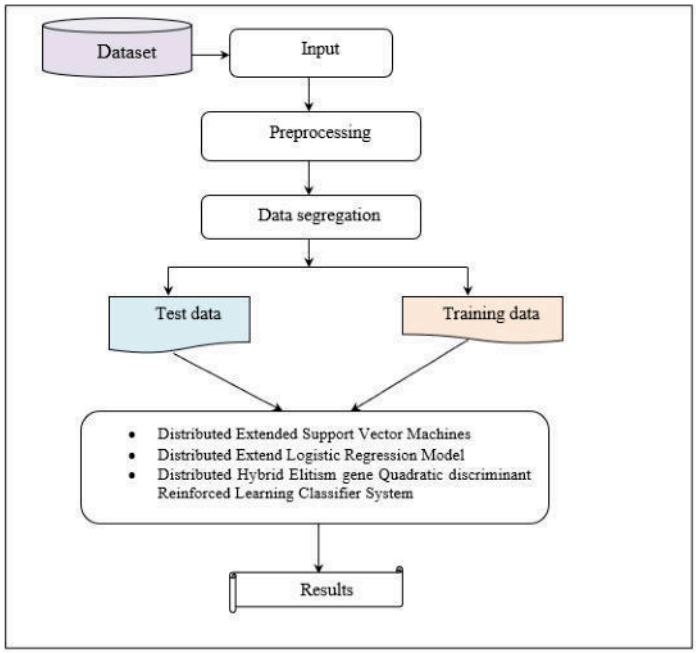


Fig. 3.1: System Design

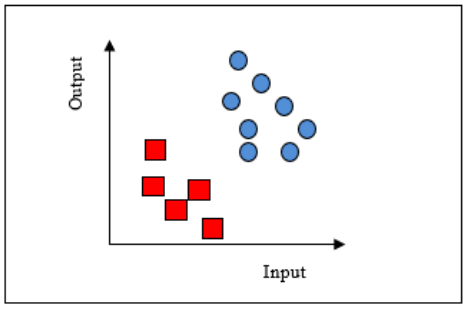


Fig. 3.2: Distributed Extended Support Vector Machines

3.1. Distributed eSVM. Distributed Extended Support Vector Machines (DeSVM) [21] is a data classification approach that uses hyperplanes to categorize the data into different classes. In DeSVM, Distributions are considered any population of data that has scattered in the two-dimensional space. The DeSVM technique is often useful for data with non-regularity, or data with an unknown distribution. Let us consider the DeSVM has two types of values, which are shown in figure 3.2.

Figure 3.2 illustrates the DeSVM principle used to build numerous separating hyperplanes from labeled data, dividing the data space into different classes where only one type of data is distributed in each segment. The DeSVM technique is often useful for data with non-regularity, or data with an unknown distribution [18]. To solve the classifier, then draw the straight line $y = ax + b$ that separates the red and blue items to classify the data above and below.

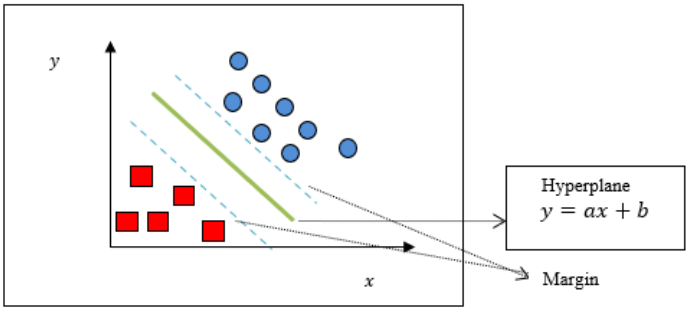


Fig. 3.3: Classification results

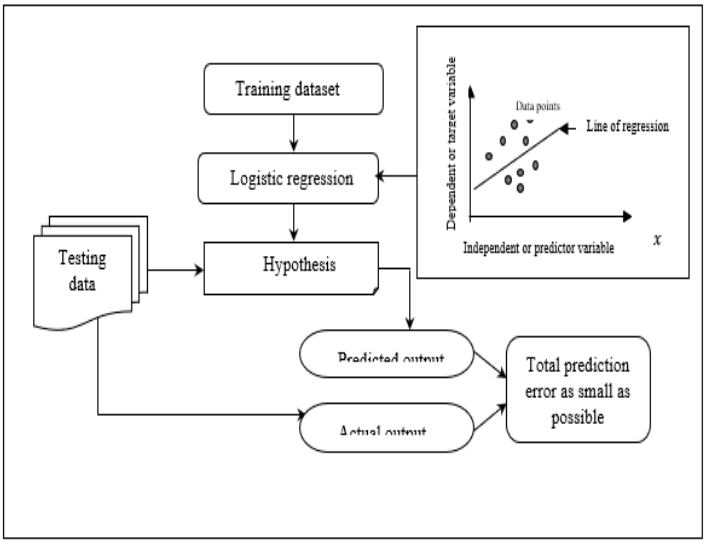


Fig. 3.4: Logistic Regression model

Figure 3.3 illustrates the classification results of DeSVM. DeSVM generates the multiple separating hyperplanes such that the data space is divided into multiple classes and each segment contains only one type of data. The DeSVM works based on finding a decision boundary (i.e hyperplane) between binary classes and maximizing the margin i.e. the distance between the decision boundary and the closest data points.

3.2. Distributed Extend Logistic Regression Model (Modified Linear Regression). The Distributed Extend Logistic Regression Model (DeLR) [22] is a machine learning technique employed to analyze the relationship between one or more independent variables (x) on the training data set and a dependent variable (y). By applying a Linear Regression, that employs a linear function called a hypothesis. To solve the parameters of this hypothesis equation, the cost function also known as the squared error function is minimized for regression problems. These parameters are determined using the gradient descent approach.

Figure 3.4 illustrates the process of Logistic Regression where the modeling technique provides the rela-

tionship or correlation between the dependent and independent variables. Logistic Regression is a type of prescriptive modeling technique that measures the relationship or correlation between the dependent and independent variables. In the two-dimensional space, 'x' axis represents the independent variable and 'y' represents the dependent variable.

Linear regression with only one variable is represented as

$$y \ = \ AX \ + \ B \tag{3.1}$$

Multiple linear regression is expressed as

$$y \ = A_0 + A_1 * X_1 \ + \ A_2 * X_2 + A_3 * X_3 + \ldots . A_n * X_n \tag{3.2}$$

where A denotes a coefficient. The goal of machine learning is to establish a mapping between data

$$f : X \rightarrow Y \tag{3.3}$$

The mapping (f) between input variables 'X' and output variables 'Y' is represented by a regression model. The Logistic Regression specifies a line that fits the relationship between the input variable (X) and the output variable (Y) by determining the specific weight of the input variable coefficient. The hypothesis analysis the training and testing results and provides the predicted output. Finally, the classification result with minimum error is obtained.

3.3. Distributed Hybrid Elitism gene Quadratic discriminant Reinforced Learning Classifier System. A Distributed Hybrid Elitism gene Quadratic discriminant Reinforced Learning Classifier System (DHEGQDRLCS) is an adaptive system that integrates machine learning, evolutionary computing, and other heuristics to deal with a Pancreatic Cancer diagnosis. The conventional DeSVM and DeLR single optimal models not having the ability to perform accurate multiclass classification while considering a large volume of input samples. Contrary to the conventional method, the purpose of DHEGQDRLCS is to accomplish the issue of a single optimal model by constructing a hybridized model. In other words, the DHEGQDRLCS have the ability to solve the multi-class classification while handling the large volume of input samples by integrating three different techniques namely Elitism gradient gene optimization, Reinforcement algorithm, and kernel Quadratic discriminant analysis.

Figure 3.5 depicts the architecture diagram of the proposed DHEGQDRLCS to perform accurate Pancreatic Diagnosis. Let us consider the Pancreatic dataset D'and collect the number of 590 data samples $S_i \in S_1, S_2, S_3, ..S_n$. The 590 urine samples are collected for the four biomarker panels, including 183 control samples, 208 benign hepatobiliary disease, and 199 pancreatic ductal adenocarcinomas (PDAC) samples. The four biomarker panels are creatinine, LYVE1, REG1B, and TFF1. The dataset includes 14 columns (i.e. attributes) listed in table 3.1.

After the data collection, training and testing are identified for pancreatic cancer diagnosis. Then the proposed technique performs samples are separated into training and testing subsets. The separation of training and testing is a technique used for evaluating the performance of a machine learning algorithm. It is also used for solving classification problems for any supervised learning algorithms. The process involves taking a dataset and dividing it into two subsets. The first subset is called as training dataset set used to fit the machine learning model. The second subset called the testing dataset is used to train the model rather than the input provided to the model. In the proposed technique, each subset has half of the samples. After the separation, the pancreatic cancer diagnosis is performed using Distributed Hybrid Elitism gene Quadratic discriminant Reinforced Learning Classifier System (DHEGQDRLCS). The proposed DHEGQDRLCS technique integrates three methods as Elitism gradient gene optimization, Reinforcement algorithm, and kernel Quadratic discriminant analysis to diagnose pancreatic disease.

The proposed DHEGQDRLCS first uses the kernel Quadratic discriminant analysis for analyzing the testing and training samples. Kernel Quadratic discriminant analysis is a supervised machine learning classifier used to classify the samples into two or more classes based on likelihood estimation with help of Gaussian kernel functions. A likelihood method is a measure of the relationship between the training and testing data samples.

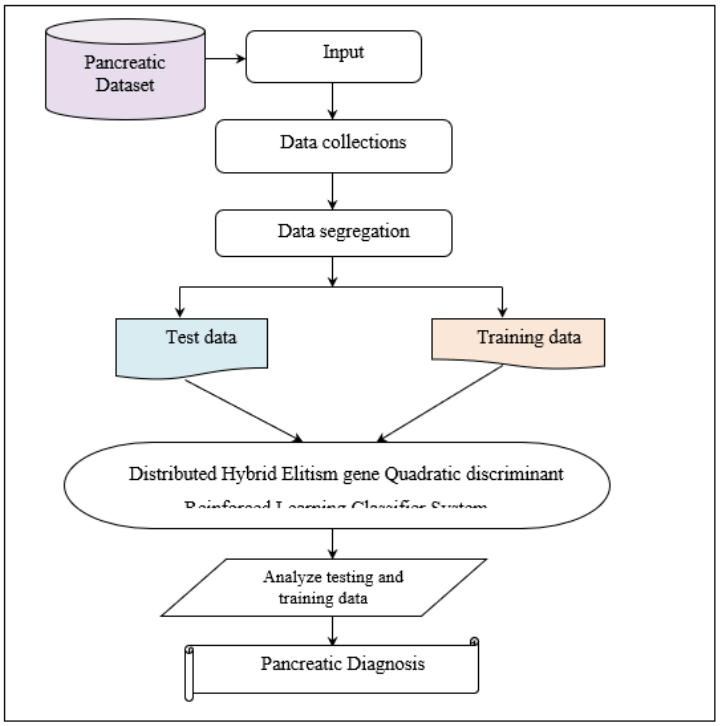


Fig. 3.5: Architecture diagram of proposed DHEGQDRLCS

Let us consider the number of training samples in the dataset. First, the number of classes 'C_j = 'C_1,'C_2,..'C_n are defined. The mean value of the class is calculated as given below:

$$M_c = \frac{1}{n} \sum_{i=1}^{n} S_i \tag{3.4}$$

From (3.4), M_c denotes a mean value of class and n denotes the number of training samples 'S_i'. Therefore, the correlation between the mean and samples is measured by applying a Kernel Quadratic discriminant analysis as given below:

$$R\left(S_i, M_{cj}\right) = \frac{1}{\sqrt{2\pi}\ d} \exp\left[-0.5\left[\frac{\left|S_i - M_{cj}\right|}{d}\right]^2\right] \tag{3.5}$$

where $R\ \left(S_i, M_{cj}\right)$ represents a Kernel Quadratic discriminant analysis output, d represents the deviation, M_c represents the mean of the class, S_i indicates a training sample. After that, the maximum correlation for the training samples being classified into the particular class is identified through the Elitism gene optimization technique.

Elitism gradient gene optimization is a metaheuristic evolutionary algorithm to generate high-quality solutions in the search problems (i.e. finding maximum correlated results. By applying the Elitism gradient gene optimization technique, first, the population of genes (i.e. correlations between the samples and mean of classes) are initialized in search space.

$$W = R_1\ \left(S_i, M_{cj}\right), R_1\ \left(S_i, M_{cj}\right)..R_n\ \left(S_i, M_{cj}\right) \tag{3.6}$$

Table 3.1: Attribute descriptions

S.no	Attributes	Description
1	sample_id	id of the sample
2	patient_cohort	Any group of individuals affected by common diseases,
3	sample_origin	origin of the patient samples (BPTB: Barts Pancreas Tissue Bank)
4	Age	Patient age
5	sex	M- Male F-Female
6	diagnosis	1. non-cancerous pancreas 2.benign hepatobiliary disease i.e., pancreatic cancer, 3. Pancreatic ductal adenocarcinoma
7	stage	stages of pancreatic cancer IA, IB, IIA, IIIB, III, IV
8	benign_sample_diagnosis	benign hepatobiliary disease samples
9	plasma_CA19_9	Monoclonal antibody levels in blood plasma, which are frequently elevated in pancreatic cancer patients.
10	creatinine	Creatinine is a protein that commonly utilized as a kidney function indicator
11	LYVE1	Lymphatic vessel endothelial hyaluronan receptor 1 is a protein discovered in the urine that may have a role in tumour spread
12	REG1B	Urinary levels of a protein linked to pancreatic regeneration,
13	TFF1	Urinary Trefoil Factor 1 levels evaluated, which could be linked to urinary tract regeneration and repair
14	REG1A	REG1A: Urinary levels of a protein connected to pancreatic regeneration,

After the initialization, the fitness is measured based on the gradient ascent function also often called steepest descent for finding a local maximum of a function (i.e. arg max).

$$F\ (W) = arg\ \max\ R\ (S_i, M_{cj}) \tag{3.7}$$

where $F\ (W)$ indicates the fitness function, $arg\ max$ denotes an argument of the maximum function. After that, the Elitism selection operation is applied to a gene optimization for finding the fittest individual among the populations by setting the threshold 't'. Therefore, the optimal selection procedure is obtained as follows,

$$Z_p = \begin{cases} F\ (W) > t\ ; & \text{select individuals} \\ \text{Otherwise;} & \text{Reject the individuals} \end{cases} \tag{3.8}$$

where Z_p denotes an elitism selection outcome (i.e. predicted classification results), t denotes a threshold, $f\ (X)$ indicates fitness. In this way, optimum correlated results are identified and classified the sample into a particular class. As a result, the optimum correlated results are used to provide the final predicted output classification results.

Then applying the model-free reinforcement algorithm considers the state-action pair at each time step to predict the accurate classification results by minimizing the loss function. The state-action pair in the reinforcement algorithm finds the total amount of expected rewards for taking an action from the state. The state describes the current situation of the classifier whereas the action describes the set of possible decisions made after observing the results space (i.e. loss function) from the states.

During the learning process, loss of the classification is estimated based on the target class and the predicted classification results as given below,

$$l = \ (Z_t - Z_p)^2 \tag{3.9}$$

$$Z_t = r + \beta \; max \; Z'(s,a) \tag{3.10}$$

where l denotes a loss function, Z_t indicates a target classification result, 'Z_p' predicted classification results, r denotes a reward, β discount factor that denotes between 0 and 1, $max \; Z'(s,a)$ denotes a maximum predicted reward from all possible actions. In order to minimize the loss function, the output results get updated based on the learning rate.

$$Z_{p_{(t+1)}} = Z_p + \quad \tau \; (r + \beta \; max \; Z'(s,a) - Z_p)^2 \tag{3.11}$$

where $Z_{p_{(t+1)}}$ denotes an updated value of the output functions, Z_p indicates the previous classification value, τ denotes a learning rate $(0 < \tau < 1)$, r denotes rewards,β indicates a discount factor between 0 and 1. This process is iterated until finding the minimum loss. Finally, accurate classified results are obtained. Based on classification results, non-cancerous pancreas, benign hepatobiliary disease i.e., pancreatic cancer, and Pancreatic ductal adenocarcinoma are correctly diagnosed with minimum error. The DHEGQDRLCS algorithmic process is explained as follows:

// Algorithm 1: Distributed Hybrid Elitism gene Quadratic discriminant Reinforced Learning Classifier System
Input: Dataset**,** Number of data samples $\boldsymbol{S_i = S_1, S_2, \ldots S_n}$ **Output:** Improve the pancreatic diagnosing accuracy

Begin
Collect the data samples ' $S_i = S_1, \; S_2, \ldots S_n$ from the dataset
Separate testing and training samples
For each training samples ' S_i'
Define number of classes 'C_j'
for each class 'C_j
Compute mean value 'M_c'
end for
Measure the correlation between training samples and mean value '$R\ (S_i, M_{cj})$'
Obtain multiple correlation results
Generate the population of correlation results 'X '
For each individual '$R\ (S_i, M_{cj})$' in population
Measure the fitness '$f\ (X)$'
Apply elitism selection **'Z_p'**
if $(f\ (X) > t)$ **then**
Select the individuals
Classify the data into particular class
else
Reject the individuals
End if
For each classification results 'Z_p'
Compute the loss function 'l'
Update the **results** '$Z_{p_{(t+1)}}$ '
Obtain the classification results with minimum loss
End for
End

Algorithm 1 describes the step-by-step process of the proposed DHEGQDRLCS for accurate pancreatic disease diagnosis. First, the number of data samples is collected from the dataset. The collected data are separated into testing and training. The input training data samples are given to the hybrid learning classifier system. The proposed classifier analyzes the training data samples with the mean of kernel Quadratic discriminant function. Then the maximum correlation results are identified by applying the Elitism gradient gene optimization. Initialize the population of correlation results. Then gradient ascent function is applied to

measure the fitness of the individual. Followed by, the elitism selection procedure is applied for finding the maximum correlated results. Finally, the samples that highly correlated to the mean value are classified into that particular class. In this way, classification results are obtained. Finally, model-free reinforcement learning is applied for minimizing the loss function by updating the learning rate. In this way, accurate classification results are observed. Based on classification results, pancreatic cancers are correctly identified.

4. Experimental settings. In this section, experimental evaluation of proposed DHEGQDRLCS and existing DeSVM [21] and DeLR [22] are implemented in Java language with Hadoop distributed system by using a pancreatic dataset. In this article, the Hadoop Distributed System is applied that provides high-performance data access while handling a large volume of datasets. The pancreatic dataset includes 590 data samples and 14 attributes. The 590 urine samples are collected for the four biomarker panels, including 183 control samples, 208 benign hepatobiliary disease, and 199 pancreatic ductal adenocarcinomas (PDAC) samples. The attributes are sample_id, patient_cohort, sample_origin, Age, sex, diagnosis, stage, benign_sample_diagnosis, plasma_CA19_9, creatinine, LYVE1, REG1B, TFF1, and REG1A. The original dataset is divided into three subsets such as dataset 1, dataset 2, and dataset 3 to compare the accuracy in this paper.

5. Performance evaluation under various metrics. In this section, the performance analysis of the proposed DHEGQDRLCS and existing DeSVM and DeLR are discussed with different metrics such as accuracy, precision, recall, F1_Score, Specificity, true negative (TN), true positive (TP), False negative (FN), False positive (FP), ROC_AUC, PRC_AUC, and PRC_APS.

Accuracy: It is defined as the number of data samples that are correctly diagnosed into different classes. The accuracy is calculated as given below:

$$Accuracy = \left[\frac{TP + TN}{TP + TN + FP + FN}\right] * 100 \tag{5.1}$$

where TP indicates a true positive i.e. number of data samples correctly diagnosed as the non-cancerous pancreas or benign hepatobiliary disease i.e., pancreatic cancer, or Pancreatic ductal adenocarcinoma, TN indicates a number of true negatives i.e. normal samples correctly diagnosed as normal, FP represents the false positive i.e. normal samples incorrectly identified as pancreatic cancer, FN indicates a false negative i.e. cancer samples incorrectly identified as normal.

Precision: The precision is computed as given below:

$$Precision \ = \left[\frac{TP}{TP + FP}\right] * 100 \tag{5.2}$$

where, TP denotes the true positive, FP symbolizes the false positive.

Recall: The formula for calculating the recall is given below:

$$Recall = \left[\frac{TP}{TP + FN}\right] * 100 \tag{5.3}$$

where TP represents the true positive, FN symbolizes the false negative.

F1_Score: It is estimated based on the average mean of precision as well as recall. The F1_Score is formulated as given below:

$$F1\_Score = \left[2 * \frac{Precision \ * \ Recall}{Precision \ + \ Recall}\right] * 100$$

Specificity: It is the test's ability to correctly reject healthy samples without a condition. The Specificity is measured as follows:

$$Specificity = \left[\frac{TN}{TN + FP}\right] * 100 \tag{5.4}$$

where TN represents the true negative, FP symbolizes the false positive.

Table 5.1: Comparison of Statistic

Parameters	**Statistic**		
	Existing DeSVM	**Existing DeLR**	**Proposed DHEGQ-DRLCS**
Accuracy	23.71	18.60	25.45
Balanced accuracy	15.27	17.69	18.98
F1_Score	13.27	2.14	14.47
Precision	13.24	15.70	16.21
Recall	13.12	19.54	20.58
Specificity	24.81	5.72	25.24
TN	25.57	22.21	19.78
TP	20.02	21.52	22.45
FN	23.89	20.74	19.44
FP	25.57	22.22	19.86
ROC_AUC	15.1	17.31	18.91
PRC_AUC	19.47	5.70	20.51
PRC_APS	21.18	9.07	22.22

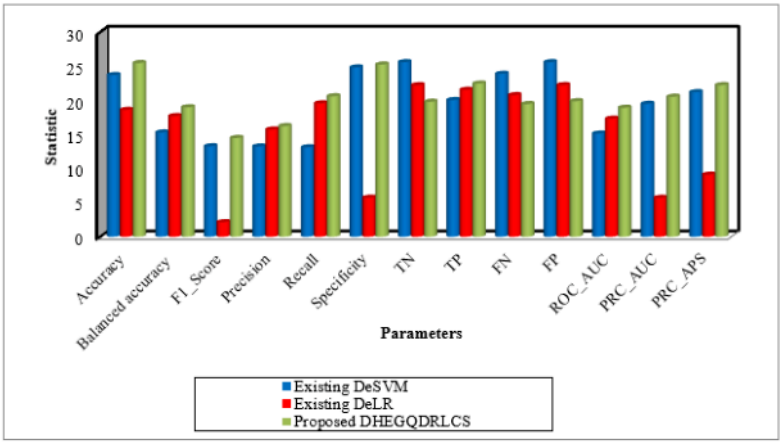


Fig. 5.1: Performance results of statistic

P- value: p-value is the particular statistical measure, such as the mean or standard deviation for the null hypothesis test.

$$P - value = \frac{X - \mu}{\sigma}$$

where X denotes input samples, μ mean and 'σ' denotes a deviation.

The performance analysis of the DHEGQDRLCS technique is conducted and results are depicted in table 5.1 and graph 5.1.

Figure 5.1 depicts the performance results of statistics with respect to a number of parameters such as Accuracy, balanced accuracy, F1-score, precision, recall, specificity, TN, TP, FN, FP, ROC_AUC, PRC_AUC, and PRC_APS. In figure 5.1, three different colors green, blue and red indicate the performance outcomes of the statistic using three methods namely Proposed DHEGQDRLCS, Existing DeSVM, and Existing DeLR respectively. Figure 5.1 clearly shows that performance results of statistics are significantly improved using

Table 5.2: Comparison of the p-value

	p-value		
	Existing DeSVM	**Existing DeLR**	**Proposed DHEGQ-DRLCS**
Accuracy	7.00E-06	0.00	0.00
Balanced accuracy	0.000483	0.00	0.00
F1_Score	0.001312	0.34	0.00010
Precision	0.001333	0.00	0.00010
Recall	0.001417	0.00	0.00
Specificity	4.00E-06	0.06	0.00
TN	3.00E-06	0.00	0.00
TP	4.50E-05	0.00	0.00
FN	6.00E-06	0.00	0.00
FP	3.00E-06	0.00	0.00
ROC_AUC	0.000526	0.00	0.00
PRC_AUC	5.90E-05	0.06	0.00001
PRC_APS	2.50E-05	0.01	0.00001

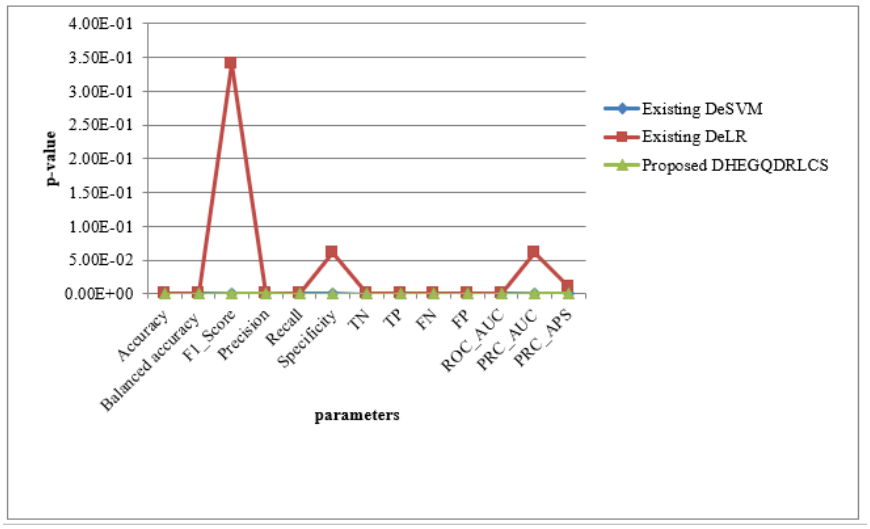


Fig. 5.2: Performance results of the p-value

proposed DHEGQDRLCS when compared to existing DeSVM and DeLR. This is because the DHEGQDRLCS uses the kernel Quadratic discriminant function to analyze the training samples and identify the maximum correlated results by applying the Elitism gradient gene optimization. Based on the optimization results, the accuracy of pancreatic cancer diagnosis is increased by improving the true positive and minimizing the true negative, true positive, and false negative.

The experimental results of p-value using three methods Proposed DHEGQDRLCS, existing DeSVM, and existing DeLR are shown in table 5.2 and Figure 5.6. The p-value is a statistical test that measures the probability of extreme results of the statistical hypothesis test. Among three different methods, the proposed DHEGQDRLCS provides an improved performance than the other two existing methods. As shown in figure 5.6, the DHEGQDRLCS obtains the p-value equal to 0. 000 are lesser than which is less than. 05. Then, the

Table 5.3: Comparison of Mean and Standard deviation using dataset 1

Parameters	Mean			Standard deviation		
	Existing DeSVM	Existing DeLR	Proposed DHEGQ-DRLCS	Existing DeSVM	Existing DeLR	Proposed DHEGQ-DRLCS
Accuracy	0.77	0.75	0.79	0.02	0.02	0.01
Balanced accuracy	0.68	0.68	0.69	0.02	0.03	0.01
F1_Score	0.42	0.42	0.43	0.03	0.04	0.02
Precision	0.35	0.33	0.37	0.03	0.04	0.02
Recall	0.54	0.57	0.58	0.04	0.05	0.03
Specificity	0.81	0.79	0.85	0.02	0.02	0.01
TN	347.5	338.60	364.20	7.6	8.78	5.63
TP	43.5	45.70	46.90	2.84	3.86	2.33
FN	36.5	34.30	33.10	2.84	3.86	2.21
FP	82.3	91.20	65.60	7.42	8.60	5.58
ROC_AUC	0.71	0.72	0.73	0.03	0.04	0.03
PRC_AUC	0.34	0.37	0.38	0.03	0.02	0.01
PRC_APS	0.34	0.38	0.39	0.03	0.02	0.01

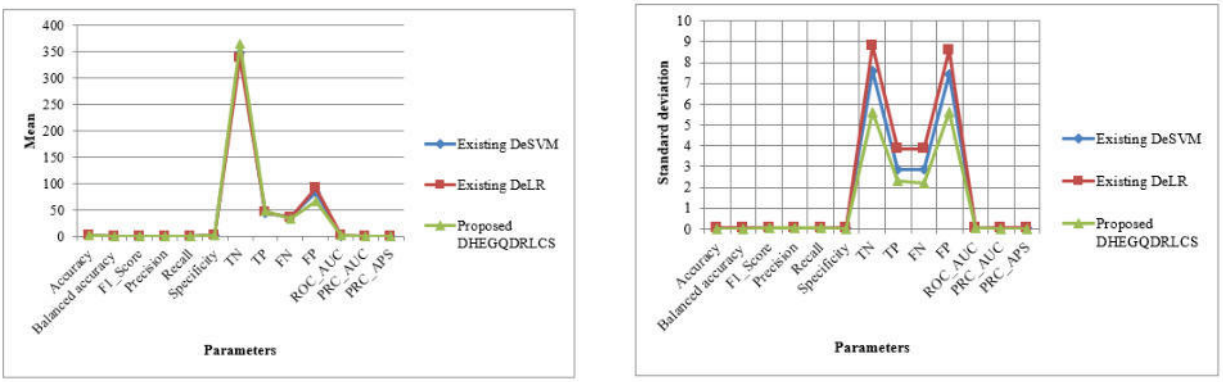


Fig. 5.3: (a) Performance results of mean deviation using dataset 1; (b) Performance results of standard deviation using dataset 1

results are statistically significant.

Table 5.3 and figure 5.3 reveal the experimental results of the mean and standard deviation using dataset 1. The performance of the mean is estimated using three methods DHEGQDRLCS, existing DeSVM, and existing DeLR depending on a number of parameters. From the observed results, it is demonstrated that the proposed DHEGQDRLCS technique provides improved mean value results and minimizes the deviation when compared to existing techniques. Let us consider the input data samples, the mean of the DHEGQDRLCS technique by computing the accuracy is 0.79, and the accuracy of existing DeSVM and DeLR are observed as 0.77 and 0.75 respectively. Similarly, the various parameters are estimated to get final results.

Figure 5.4 illustrates the experimental results of mean and standard deviation with the different numbers of parameters. As shown in the figure, the mean and standard deviation is estimated by using three methods DHEGQDRLCS, existing DeSVM, and existing DeLR. Among the three methods, the performance of mean using DHEGQDRLCS is improved and the standard deviation is minimized when compared to DeSVM and

Table 5.4: Comparison of Mean and Standard deviation using dataset 2

Parameters	Mean			Standard deviation		
	Existing DeSVM	Existing DeLR	Proposed DHEGQ-DRLCS	Existing DeSVM	Existing DeLR	Proposed DHEGQ-DRLCS
Accuracy	0.79	0.74	0.80	0.01	0.02	0.01
Balanced accuracy	0.59	0.69	0.70	0.04	0.04	0.03
F1_Score	0.3	0.43	0.44	0.07	0.04	0.03
Precision	0.31	0.33	0.36	0.04	0.03	0.02
Recall	0.3	0.62	0.63	0.11	0.06	0.05
Specificity	0.88	0.77	0.89	0.02	0.02	0.01
TN	378	330.40	319.70	9.4	6.77	5.71
TP	24.2	49.20	50.00	8.47	5.12	4.52
FN	55.8	30.80	28.00	8.47	5.12	4.52
FP	51.8	99.40	46.10	9.39	6.48	5.42
ROC_AUC	0.65	0.73	0.74	0.04	0.04	0.03
PRC_AUC	0.28	0.39	0.42	0.04	0.02	0.01
PRC_APS	0.29	0.39	0.43	0.04	0.02	0.01

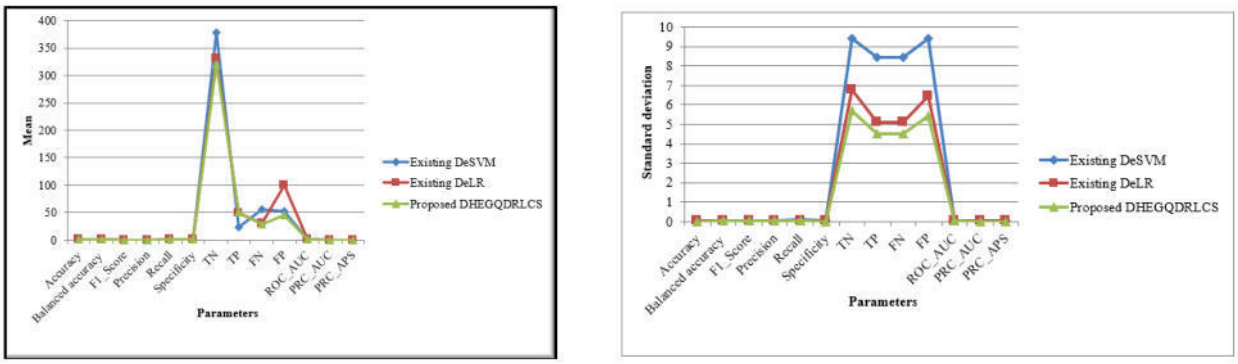


Fig. 5.4: (a) Performance results of mean using dataset 2; (b):Performance results of standard deviation using dataset 2

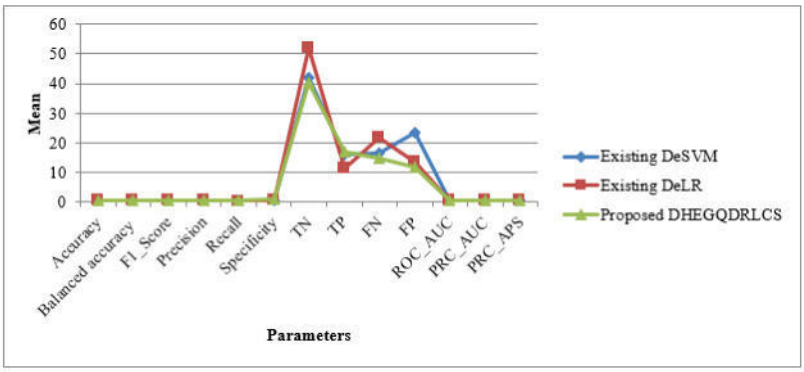
DeLR. This is due to the maximum correlated function used for early identification of pancreatic cancer. The false positive rate is minimized by applying reinforcement learning to accurately detect pancreatic cancer and minimize the loss.

Table 5.5 and figure 5.5 illustrate the performance analysis of the mean and standard deviation using dataset 3 with respect to three different methods DHEGQDRLCS, existing DeSVM, and existing DeLR. The observed results indicate that the DHEGQDRLCS provides better performance when compared to conventional methods. Let us consider the accuracy parameter in the mean estimation. The observed performance of accuracy using DHEGQDRLCS is 0.66 and the accuracy of existing methods is 0.59 and 0.64. The results indicate that the accuracy using DHEGQDRLCS is significantly improved by improving cancer detection and minimizing the loss function.

Table 5.7 indicates that the specificity cutoff with mean for four different stages stage I, stage II, stage III, and stage (IV). The machine learning-based Hadoop distributed eSVM, Distributed eLR, and DHEGQDRLCS

Table 5.5: Comparison of Mean and Standard deviation using dataset 3

Parameters	Mean			Standard deviation		
	Existing DeSVM	Existing DeLR	Proposed DHEGQ-DRLCS	Existing DeSVM	Existing DeLR	Proposed DHEGQ-DRLCS
Accuracy	0.59	0.64	0.66	0.07	0.07	0.03
Balanced accuracy	0.57	0.57	0.59	0.06	0.05	0.03
F1_Score	0.44	0.39	0.46	0.07	0.07	0.05
Precision	0.41	0.49	0.52	0.06	0.12	0.05
Recall	0.49	0.34	0.51	0.12	0.11	0.10
Specificity	0.64	0.79	0.82	0.11	0.15	0.07
TN	41.9	51.80	40.20	7.4	10.14	6.44
TP	16.2	11.30	17.00	4.42	3.83	2.67
FN	16.6	21.50	14.80	3.5	3.41	2.34
FP	23.3	13.40	12.00	7.53	9.66	3.80
ROC_AUC	0.58	0.58	0.59	0.07	0.07	0.06
PRC_AUC	0.43	0.45	0.47	0.07	0.08	0.06
PRC_APS	0.45	0.46	0.47	0.07	0.07	0.05



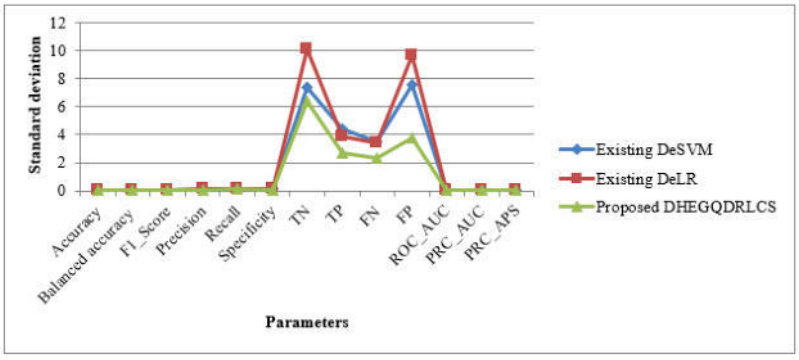


Fig. 5.5: (a) Performance results of mean using dataset 3; (b) Performance results of standard deviation using dataset 3

Table 5.6: Model accuracy with comparisons

Methods		Accuracy	F1_Score	Precision	Recall	TN	TP	FN	FP	ROC_AUC
Existing DeSVM	statistic	23.71	13.27	13.24	13.12	25.57	20.02	23.89	25.57	15.10
	p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	avg_mean	0.72	0.39	0.36	0.45	255.80	27.97	36.30	52.47	0.65
	avg_std	0.03	0.06	0.04	0.09	8.14	5.24	4.94	8.11	0.05
Existing DeLR	statistic	18.60	2.14	15.70	19.54	22.21	21.52	20.74	22.22	17.31
	p-value	0.00	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	avg_mean	0.71	0.41	0.38	0.51	240.27	35.4	28.87	68	0.68
	avg_std	0.04	0.04	0.05	0.06	8.56	4.27	4.13	8.25	0.05
Proposed DHEGQ-DRLCS	statistic	25.45	14.47	16.21	20.58	19.78	22.45	19.44	19.86	18.91
	p-value	0.00	0.11	0.03	0.00	0.00	0.00	0.00	0.00	0.00
	avg_mean	0.75	0.44	0.41	0.57	238.36	37.96	25.3	41.23	0.68
	avg_std	0.01	0.03	0.03	0.05	5.92	3.17	3.02	4.93	0.04

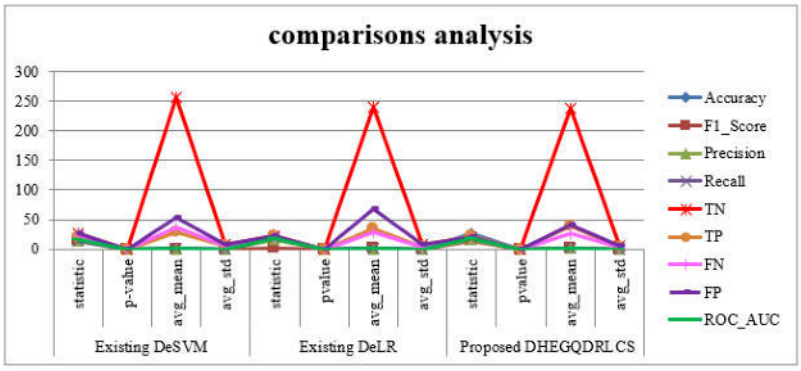


Fig. 5.6: Model accuracy with comparisons

are applied for pancreatic cancer detection. The experiments demonstrated that the proposed DHEGQDRLCS improves the performance of the classifier for early identification of pancreatic cancer with stage values ranging from 0.6 to 0.10 and a mean specificity cutoff. The algorithms in the Hadoop distributed file system of DHEGQDRLCS provide better accuracy than other existing approaches in this comparison research.

5.1. Advantages and Limitation of DHEGQDRLC. The architecture of DHEGQDRLCS is used to increase the accuracy of early pancreatic cancer diagnosis based on the hybridization of three techniques namely kernel quadratic discriminant function, Elitism gradient gene optimization, and reinforcement learning algorithm. The hybridization process minimizes the loss function of disease diagnosis and increases the classification results. The limitation of the proposed DHEGQDRLCS is not having the ability to analyze time consumption and space complexity in the present state when applying the large volume of the dataset.

6. Conclusion. In this section, early detection of pancreatic cancer is very significant before cancer swells to other organs in the body. However, early detection of pancreatic cancer is difficult because this cancer has non-specific symptoms. The conventional distributed eSVM, Distributed eLR approaches perform pancreatic

Table 5.7: Cutoff values with a mean of proposed models

specificity cutoff WITH MEAN	**stages**
0.6 to 0.7	I
0.7 to 0.8	II
0.8 to 0.9	II
0.9 to 0.10	IV

cancer of pancreatic cancer. But the higher accuracy and loss rate was not minimized. Therefore, a novel DHEGQDRLCS is developed to increase the classifier's performance for early pancreatic cancer diagnosis with four different stages based on p-value and statics, according to preliminary data samples. In the proposed DHEGQDRLCS, the kernel quadratic discriminant function analyzes the testing and training data samples using the kernel function. Then the Elitism gradient gene optimization is applied to provide the final disease diagnosis results. In order to minimize the loss function of disease diagnosis, the reinforcement learning algorithm is applied to find the optimal learning rate. Finally, the hybridized classifier provides precise classification results with a minimum loss function. Based on the classification results, pancreatic cancer is correctly identified. A comprehensive experimental evaluation is carried out using three datasets with different parameters such as Accuracy, balanced accuracy, F1-score, precision, recall, specificity, $TN, TP, FN, FP, ROC_{AUC}, PRC_{AUC}$, and PRC_{APS}. The quantitative performance result indicates that the presented DHEGQDRLCS achieves higher accuracy in a pancreatic cancer diagnosis than the conventional methods. The proposed future work could increase the number of applications for pancreatic disease detection and also improve accuracy.

REFERENCES

[1] Stephen P Pereira, Lucy Oldfield, Alexander Ney, Phil A Hart et al. "Early detection of pancreatic cancer" , The Lancet Gastroenterology & Hepatology, 2020

[2] Santosh Reddy P, Chandrasekar M. "PAD: A Pancreatic Cancer Detection based on Extracted Medical Data through Ensemble Methods in Machine Learning" , International Journal of Advanced Computer Science and Applications, 2022

[3] Md Manjurul Ahsan, ShahanaAkter Luna, Zahed Siddique. "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review" , Healthcare, 2022

[4] Stefanus Tao Hwa Kieu, Abdullah Bade, Mohd Hanafi Ahmad Hijazi, HoshangKolivand. "A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions" , Journal of Imaging, 2020

[5] Rahaman, M.M.;Li, C.;Yao, Y.;Kulwa, F.;Rahman, M.A.;Wang, Q.;Qi, S.;Kong, F.;Zhu, X.;Zhao, X.Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transferlearningapproaches.J.X-RaySci.Technol. 2020, 28,821–839.

[6] Ma,J.;Song, Y.;Tian,X.; Hua,Y. Zhang, R.;Wu, J.Survey on deep learning for pulmonary medicalimaging.Front.Med. 2019,14,450–469.

[7] Ardila,D.;Kiraly,A.P.;Bharadwaj,S.;Choi,B.;Reicher,J.J.;Peng,L.;Tse, D.; Etemadi, M.; Ye, W. Corrado, G.; et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography.Nat.Med. 2019, 25,954–961.[CrossRef]

[8] Kieu, S.T.H.;Hijazi, M.H.A.;Bade, A.;Yaakob, R.;Jeffree, S. Ensemble deep learning for tuberculosis detection using chest X-Ray and canny edge detected images. IAES Int. J. Artif. Intell.**2019**, 8, 429–435.

[9] Ayan,E.;Ünver,H.M.Diagnosis of Pneumonia from Chest XRay Images using DeepLearning .Sci.Meet.Electr.-Electron.Biomed.Eng.Comput.Sci. 2019.

[10] Salman, F.M.; Abu-naser, S.S.; Alajrami, E.; Abu-nasser, B.S.; Ashqar, B.A.M.COVID-19 Detection using Artificial Intelligence Int.J.Acad.Eng.Res. 2020,4,18–25.

[11] Gao, X.W.; James-reynolds, C.; Currie, E.Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture .Neurocomputing 2019,392,233–244.[CrossRef]

[12] Gozes,O.;Frid,M.;Greenspan,H.;Patrick,D. Rapid AIDevelopment Cycle for the Coronavirs (COVID-19) Pandemic: Initial Results for Automated Detection & Patient Monitoring using Deep Learning CT Image Analysis Article . arXiv 2020 ,arXiv : 2003 .05037.

[13] Mithra, K.S.; Emmanuel, W.R.S. Automated identification of mycobacterium bacillus from sputum images for tuberculosis diagnosis. Signal Image Video Process. 2019.

[14] Samuel, R.D.J.;Kanna, B.R.Tuberculosis (TB) detection system using deep neural networks. Neural Comput.Appl. 2019, 31,1533–1545.

[15] O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.;Walsh, J. Deep

Learning vs .Traditional Computer Vision.Adv.Intell.Syst.Comput.**2020**, 128–144.
[16] Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop,Swinoujscie,Poland.
[17] Shorten,C.;Khoshgoftaar,T.M.A survey on Image Data Augmentation for DeepLearning . J.BigData 2019,6.
[18] Ker, J.; Wang, L. Deep Learning Applications in Medical Image Analysis. IEEE Access 2018, 6, 9375–9389.
[19] Wang, C.; Chen, D.; Hao, L.; Liu, X.; Zeng, Y.; Chen, J.; Zhang, G. Pulmonary Image Classificatin Based on Inception-v3 Transfer Learning Model .IEEE Access 2019 ,7, 146533 –146541
[20] Sadewo W, Rustam Z, Hamidah H and Chusmarsyah A.R, "Pancreatic Cancer Early Detection Using Twin Support Vector Machine Based on Kernel" Symmetry, 2020, 12, 1-8.
[21] Jacobson S, Dahlqvist P, Johansson M, Svensson J, Billing O, Sund M, Franklin O, "Hyperglycemia as a risk factor in pancreatic cancer: A nested casecontrol study using prediagnostic blood glucose levels", Pancreatology, Elsevier, 2021, 21, 1112-1118
[22] Jacobson S, Dahlqvist P, Johansson M, Svensson J, Billing O, Sund M, Franklin O, Alves N, Schuurmans M, Litjens G , S. Bosma J, Hermans .J and Huisman H., "Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography. Cancers, 2020, 14(2), 1-10.
[23] Ju J, V. Wismans L, A.M. Mustafa D, J.T. Reinders M, H.J. van Eijck C, A. P. Stubbs A, and Li Y, "Robust deep learning model for prognostic stratification of pancreatic ductal adenocarcinoma patients", iScience, Elsevier, 2021, 24 (12), 1-18
[24] Yana Z, Ma C, Mo J, Han W, Lv X, Chen C, Chen C, Nie X, "Rapid identification of benign and malignant pancreatic tumors using serum Raman spectroscopy combined with classification algorithms", Optik, Elsevier, 2020, 208, 1-5
[25] N. Al-Shaheri F, S.S. Alhamdani M, S. Bauer A, Nathalia Giese, Markus W. Büchler, Thilo Hackert, J rg D. Hoheisel, "Blood biomarkers for differential diagnosis and early detection of pancreatic cancer", Cancer Treatment Reviews, Elsevier, 2021, 96, 1-16
[26] Park J, Artin M G, Lee K E., Pumpalova Y S, Ingram M A, May B L, Park M, Hur C, Tatonetti N P., "Deep learning on time series laboratory test results from electronic health records for early detection of pancreatic cancer", Journal of Biomedical Informatics, Elsevier, 2022, 131, 1-11
[27] Savareh B A, Aghdaie H A, Behmanesh A, Bashiri A, Sadeghi A, Zali M, Shams R, "A machine learning approach identified a diagnostic model for pancreatic cancer through using circulating microRNA signatures", Pancreatology, Elsevier, 2020, 20, 1195-1204

Edited by: Vinoth Kumar
Received: Jun 16, 2022
Accepted: Oct 19, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 211–223, DOI 10.12694/scpe.v23i4.2021

PREDICTION OF NAC RESPONSE IN BREAST CANCER PATIENTS USING NEURAL NETWORK

SUSMITHA UDDARAJU *, G. P. SARADHI VARMA† AND M. R. NARASINGARAO‡

Abstract. Breast cancer is now the most prominent female cancer in both developing and developed nations, and that it is the largest risk factor for mortality worldwide. Notwithstanding the well-documented declines in breast cancer mortality during the last twenty years, occurrence rates continue to rise, and do so more rapidly in nations where rates were previously low. This has highlighted the significance of survival concerns and illness duration treatment. Patient data after first chemotherapy is collected from the hospital and this data is then analysed using neural network. Proposed architecture gives result as the patient is responding to the chemotherapy or not. Moreover, it also gives the risk factor in surgery. Early prediction of such things gives broader idea about how treatment should go. Once the Breast cancer is detected and if chemotherapy is done, then it becomes very important to check whether patient is responding to the chemotherapy or not. So, the proposed system architecture is designed in such a way that it detects if the patient is responding to the chemotherapy or not. And if patient is not responding to the chemotherapy, then patient should go to the surgery. The proposed system is also compared with the existing algorithms machine learning and neural network techniques like support vector machine (SVM) and Decision Tree(DT) algorithms. The proposed neural network architecture gives 99.19% accuracy where SVM and DT gives 89.15% and 74.82%. Bosom disease is known to have asymptomatic stages, which is distinguished simply by mammography and around 10% of patients getting mammography recovers further assessments, and among them 8 to 10% require bosom biopsy. Alert the cautious consideration of the radiologist to peruse mammograms to perceive mammograms is generally 30 to 60 seconds for every picture. In any case, the weakness and explicitness of human radiologist's mammography was controlled by 77-87% and 89-97%, individually. As of late, twofold peruses are allowed with most screening programs, yet this will additionally disintegrate the time heap of human radiologists. As of late, the headway of man-made brainpower (AI) has made it conceivable to recognize programmed infection on clinical pictures in radiology, pathology, and even gastrointestinalities. For bosom malignant growth screening, all the more profound examinations have additionally been led, 86.1 to 9.0% responsiveness and 79.0 to 90.0% exceptional elements. By and by, there are a couple of distributions for built up disease location of mammography under Asian with higher bosom thickness contrasted with white individuals. Bosom thickness can influence the malignant growth pace of mammography pictures. Hence, the motivation behind this study was to create and approve a profound learning model that consequently recognizes threatening bosom sores in Asian advanced mammograms and to inspect the exhibition of the model by bosom thickness level. We have acquainted our own pretreatment technique with expand the exhibition of the model. Furthermore, we tried to lead a meta-examination to contrast and accessible investigations on AI-based bosom malignant growth recognition. Apparently, this is probably the greatest review done on Asians.

Key words: Neural network, chemotherapy, Residual Cancer Burden (RCB), Pathologic complete response (PCR)

AMS subject classifications. 68T05

1. Introduction.

Bosom malignant growth has turned into the most far and wide female disease in both creating and created countries, and it is the biggest reason for death among ladies nations from one side of the planet to the other [1]. Therapies for this sort of disease are agonizing, tedious, and may not be guaranteed to guarantee that malignant growth some way or another doesn't spread to different pieces of the body. As an outcome, different scholastics are attempting to more readily grasp the connections that exist in this confounded issue and to track down viable answers for work on the possibilities of treatment adequacy and, as an important result, the personal satisfaction of ladies who truly are going through it. [2]. The series of investigates on the attainability of therapies for bosom disease in ladies is increasing at similar rate as the quantity of instances of contamination around the world. As a reaction, clinical analysts are being completed in a joint effort with specialists in the factual or computational fields to mirror shrewd models equipped for rearranging and

*Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. (susmithauddaraju@yahoo.com).
†Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.(gpsvarma@gmail.com).
‡Professor, Department of CSE, Gandhi Institute of Technology and Management, Visakhapatnam, India. (rmanda@gitam.edu).

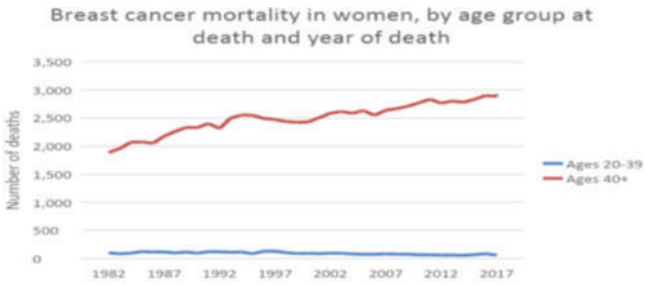


Fig. 1.1: Breast cancer incidence rates by age

demonstrating the viability of bosom disease medicines. In research, studies have been embraced to evaluate mental components of disease treatment and its relationship to treatment confidence, post-treatment impacts, connections, and effects on sleep deprivation. [3]. A few investigations take a gander at how people who have sought through treatment think [4]. One more ongoing line of examination investigates how comorbidities, age, and time since analysis impact therapy in the early discovery of disease [5] and, at long last, the effect of chemical substitution treatment on bosom disease cell development in menopause. The viability of bosom malignant growth medicines can be inspected from various perspectives, including social, monetary, and ladies' wellbeing. It is basic, be that as it may, to feature the significance of examination in view of information gathered from explicit patient gatherings. Teng et al. [6] learned at the bosom malignant growth instances of 5,279 ladies with invading pipe and lobular carcinoma. They were analyzed somewhere in the range of 2006 and 2010, and their data was assembled from the National Cancer Institute's SEER Cancer Registry. The first review intended to utilize Bayesian induction based prognostic displaying. In figure 1.1 we can see worldwide breast cancer incidence rates by age in women's. Here we can observe that the women above the age 40 are highly diagnosed with breast cancer [22].

Fluffy brain organizations (FNN) perform well when it connects with recovering attributes and relationships utilizing information bases in regards to clinical medicines. Canny half and half models have been used to expect emphaticness engine issues in youngsters [7] and to screen fetal wellbeing [8]. Classifier for the segmentation is stated by the Ajay Ladkat et al [9]. Different models were utilized to recognize mental imbalance in youngsters [10], grown-ups [11], and teenagers [12] utilizing electroencephalography (EEG).

At long last, it tackled difficulties in anticipating bosom malignant growth considering other assessment standards proposed by Silva Araujo et al. [13], for example, resistin, hyperglycemia, age, and BMI. Subsequently, utilizing Lymph Node Ratio Estimation investigation, such a methodology can work emphatically in the distinguishing proof of bosom malignant growth Survival Prognosis. Different AI approaches are tended to in paper by S. L. Bangare et al. [17], alongside numerical clarifications and instances of their utilization in genuine circumstances. S. L. Bangare et al. exhibited amazing review utilizing AI approaches in their distribution. K. Gulati et al. [18] and LMI Leo Joseph et al. [19] led research utilizing AI and profound learning draws near. Shachi Mall et al. [21] shown the utilization of Machine learning for sickness discovery and so on. The creators have introduced most ideal component extraction methods for shape portrayal and utilized the Neural Networks, for example, Capsule Network, KNN for the grouping.

The primary commitment of this paper is to configuration new methodology for the characterization of impacts after the hemotherapy. The paper comprises of the proposed framework engineering alongside the measurements. The outcomes have been checked and confirmed from the master specialists. Likewise, the proposed framework's outcomes have been contrasted and existing framework and there similar diagram is additionally plotted in outcome meeting. At last, decisions about the exercises acted in the paper is introduced in last meeting.

Patients who went through cutting edge bosom mammography and endocrine treatment were signed up for this survey. Just subjects beyond 18 years old who were not associated with the technique were enlisted. Distant clinical records, no necessary affirmation of dubious bosom injury, missing mammograms, or the idea of

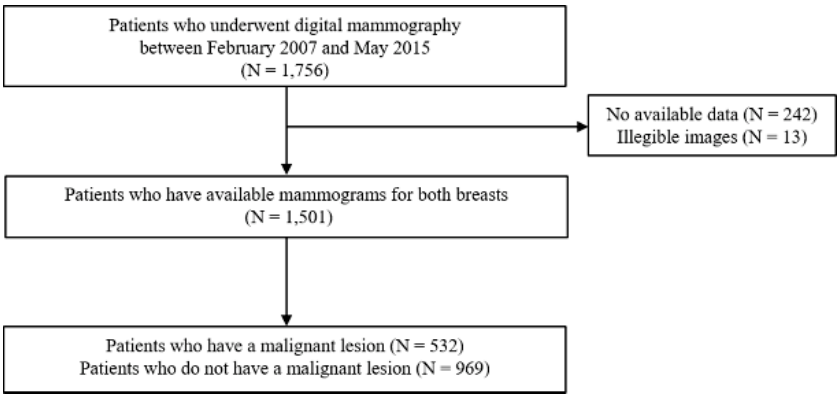


Fig. 1.2: Flow diagram

mammograms can't be translated true to form because of unsettling, defocus, or irritating spots. Low subjects were prohibited [1]. This study was supported by the Institutional Review Board (No. 201903004) and consented to the rules of the Helsinki Declaration. Board for Institutional Review has removed the prerequisite for designed informed assent on the grounds that the survey represents an immaterial gamble to the issue. Skulls and normal along the side slanted mammograms of each bosom were gotten from taking an interest subjects utilizing a photo placement and relating outline and # 40. Hallym University Sacred Heart Hospital focusing on PACS and # 41; 2560x3328 pixels. The computerized mammography convention is agreeable with the European Federation of Medical Physics Organizations. All pictures were taken utilizing a Lorad Selenia PC mammography apparatus (Hologic Incorporated, Bedford, USA). The display of the unit was guaranteed inside the predefined flexibility's brought out for quality control all through the assessment period [2]. Individual information, clarifications, or showed information that portrays the possibility of the laterality or mammography view has been taken out from the picture, Every mammography was then re-examined two radiologists, also checked for the presence of undermining wounds. Clinical records of possibly threatening bosom sores, past mammography estimations, and neurotic reports were reflectively explored and followed up for no less than 5 years after mammography assessment. Threatening sores must be affirmed by careful histopathology. The support flowchart is introduced in Figure 1.2.

2. Related works.

2.1. Data Preprocessing. All pictures were resized to 1000x1300 pixels before investigation. The pictures were then preprocessed with the CLAHE calculation (contrast-restricted versatile histogram evening out) to limit contrast contrasts between the pictures [3]. CLAHE is an adjusted form of versatile histogram balance. This is a picture handling calculation that changes the brilliance of every pixel in a picture and applies histogram adjustment locally to contiguous pixel regions to work on the nearby differentiation of the picture [4]. The CLAHE calculation tackles the clamor over amplification issue of versatile histogram evening out by restricting the incline of the combined thickness work while ascertaining histogram adjustment [5]. In this review, CLAHE was carried out utilizing the Python programming language OpenCV library rendition 4.1.2.30. We flipped every one of the mammograms of the left bosom upward and gave the picture an organization like that of the right bosom to make a bound together profound learning model for the right and left bosoms. Then, for each bosom, the crania-caudal and average sidelong slanted pictures were trimmed to eliminate just 10% of the vacant space and associated evenly. At last, the size of the connected pictures has been changed to 900x650 pixels [6]. 1.2. Record structure All combined pictures were ordered into two classes, harmful and harmless, in view of the presence or nonattendance of threatening sores in the pictures. The whole dataset was parted into preparing and test datasets, and 10% of the pictures in every classification were haphazardly relegated to the test dataset. The remainder of the preparation datasets was then additionally partitioned into the right preparation and tuning datasets in an 8: 1 proportion utilizing arbitrary tasks by classification

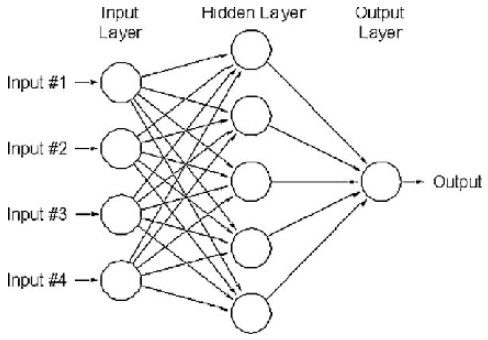


Fig. 2.1: Artificial Neural Network

[7]. The preparation, tuning, and test datasets were fundamentally unrelated and absolutely thorough. Since the dangerous growth bunch was around one-fifth the size of the non-harmful cancer bunch, the threatening growth bunch was extended in the preparation dataset to ease the class unevenness. Pictures of the threatening cancer bunch in the preparation dataset were amplified up to 5x, creating 10% and 20% amplified pictures of the dataset and 10% and 20% decreased pictures.

2.2. Training Convolutional Neural Networks (CNNs). Two CNN structures, DenseNet169 and EfficientNetB5, were utilized to foster profound learning models. Basically, DenseNet169 is a CNN structure highlighting high thickness impedes that are connected with the information include guides of each past sub block and utilized as the info highlight guide of a specific sub block [8]. This tight association tackles the disappearing angle issue and diminishes the quantity of boundaries [9]. EfficientNetB5 was planned utilizing MBconv blocks that all the while control the equilibrium of organization width, profundity, and goal through support learning [10]. This organization beats the picture arrangement organization of past ImageNet datasets with less boundaries and induction time [11]. Both of these CNN models prepared on the ImageNet dataset and optimized on the prepared dataset for this review. Using the Adam Optimizer, we used 0.9 for Beta 1 and 0.999 for Beta 2 to limit double cross entropy. The underlying learning rate was positive, and the learning rate decreased by 10 to 10 years until it reached 107. The group size was configured at 4, and the way racy factor was fine. Early stopping was applied after 30 years, with a 20-year tolerance. The dropout was not used for DenseNet169, but was used for EfficientNetB5 with a 0.4 dropout rate. The Pytorch structure was used in a handling unit designed by NVIDIA GeForce Titan RTX.

2.3. Gradient-Weighted Class Activation Mapping (Grad-CAM). Two CNN structures, DenseNet-169 and EfficientNetB5, were utilized to foster profound learning models. Basically, DenseNet169 is a CNN structure highlighting high thickness impedes that are linked with the info include guides of each past subblock and utilized as the information include guide of a specific subblock [12]. This tight association tackles the evaporating slope issue and diminish the quantity of boundaries [13]. EfficientNetB5 was arranged using MBconv blocks that meanwhile control the harmony of association width, significance, and objective through help learning [14]. This association beats the image game plan association of past ImageNet datasets with less limits and acceptance time [15]. Both of these CNN models pre-arranged on the ImageNet dataset and changed on the planning dataset for this audit. The Adam Optimizer was used to restrict equal with beta1 of 0.9 and beta2 of 0.999, compute cross entropy. The basic learning rate was OK, and the learning rate decreased by 10% every 10 ages until it showed up at 10-7. The bundle size was set to 4 and the weight decay factor was certifiable. Early stopping was used after 30 ages with diligence of 20. The dropout was not used for DenseNet169, but was used for EfficientNetB5 with a dropout speed of 0.4. The NVIDIA GeForce Titan RTX plans-handling unit utilises the Pytorch framework. A back spread mind network tends to a neuron and adjoining layers that are related by loads.

Equations (2.1) and (2.2), which describe how hidden nodes are activated, can be used to explain the

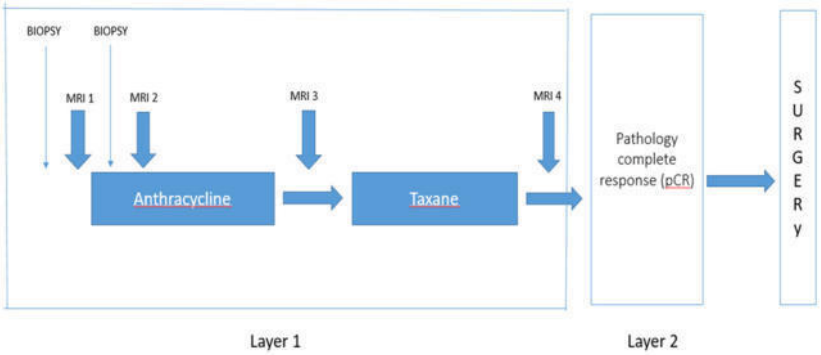


Fig. 3.1: Collection of Data

calculation.

$$I_j = \sum_i w_{ji} y_i + a_j \tag{2.1}$$

where, $f_j = activationfunction$

$$T_j = f_j < I_j \tag{2.2}$$

w_{ji} =weight attached to the nodes in input layer I and hidden layer j's connecting link
a_j =the predisposition related at every association connect between the information layer and secret layer.
y_I =contribution at hubs in input layer.
I_j =sum of the weight inputs multiplied by predisposition
T_j =result of initiation work at stowed away layer.
The guideline of result layer can be derived utilizing Eqn. (2.3) and Eqn. (2.4).

$$I_n = \sum_j w_{nj} y_j + b_n \tag{2.3}$$

$$T_n = f(I_n) \tag{2.4}$$

w_{nj} =weight related at the association interface between hubs in secret layer j and hubs in yield layer n.
b_n =the inclination related at every association interface between the secret layer and result layer.
y_j =yield at hubs in secret layer
I_n =summation of weighted yields at the result layer.
T_n =last result at the result layer.

3. Methodology. Patient data after first chemotherapy is collected from American Oncology Institute (AOI) Manglagiri, Guntur. American Oncology Institute (AOI), the best hospital for cancer in Vijayawada operates a full-fledged facility at NRI Hospital, Manglagiri. The data is collected in the form of an excel sheet. The dataset consists of 884 instances with 11 input features. The data is collected before, in-between and after Neo adjuvant chemotherapy(NAC) of the patient as shown in figure 3.1. The initial course of treatment for locally progressed breast cancer is neo adjuvant chemotherapy (NAC) (BC). Its major goal is to obtain a pathological full response in addition to making breast conserving surgery practicable (pCR) Input parameters of the dataset and its description are as stated in table 3.1.

Dataset contains patient socioeconomics and pre-treatment boundaries. Complete dataset is partitioned into the race of the patient, Estrogen Receptor Status (ERS), Progesterone Receptor Status (PRS), Hormone Receptor Status, Her2 classification Status, reciprocal bosom disease status, and record cancer laterality. Other

Table 3.1: Input parameters in database and their values

Input Parameters in database	Values of parameters
Race of the patient	Caucasian, African American, Asian, Native Hawaiian/Pacific Islander, American Indian/ Alaskan Native
Estrogen Receptor Status (ERS)	Positive or Negative
Progesterone Receptor Status (PRS)	Positive or Negative
Hormone Receptor Status	Negative if both ERS and PRS are Positive Positive if both ERS and PRS are Negative
Her2 category Status	Positive or Negative
Bilateral breast cancer status	Yes or No
Index tumour laterality	Left or Right

Table 3.2: Output parameters in database and their values

Output parameters in database	Values of parameters
pathological complete response (pCR)	Yes, No, or No surgery
Residual Cancer Burden (RCB) class	Class 0 to 3, No surgery

than this four more MRI features are used as input in this research. Totally 11 features are used as input features in this work. Race of the patient incorporates Caucasian, African American, Asian, Native Hawaiian/Pacific Islander, American Indian/Alaskan Native. Trama centers, PRS and Her2 status are either certain or negative. Chemical receptor status is negative in the event that the two ERS and PRS are negative, positive assuming ERS or PRS is positive. Patients either have reciprocal bosom disease before neoadjuvant treatment or not. File cancer laterality is left or right.

We cannot feed collected database directly to the neural without pre-processing. Some fields in the collected dataset are of the string type. So, by using linear encoding method the string is converted into the logical numbers. Proposed framework accepts these highlights as information. Then this information is taken care of in lined up with the thick layer with 16 and 64 neurons into it, etc as show in figure 3.2. By utilizing thick layer, the dimensionality of the vectors gets changed. Each neuron in the past layer is associated with each neuron in the following layer. Here for every layer except last layer activation function used is Relu. And in last layer activation function is softmax. The neural network is consisting of two parallel running convolution neural networks. It is the novelty of the proposed neural network that two neural network accepts the data simultaneously and gives one result.

Proposed system takes input features as input. Then the input is fed parallel to dense layer with 16 neurons and 64 neurons into it. 16 and 64 neurons are fed parallel to 64 and 16 neurons. 64 and 16 neurons are fed parallel to 128 and 32 neurons and so on as shown in figure 3.2. In the last layer the dimension changes from 2 to 1, which means the data is combined and only one result is taken out from the system. The outcome result is about chemotherapy response along with the criticality due to cancer tissue. By using the dense layer the dimensionality of vectors get changed. Every neuron in previous layer is connected to every neuron in the next layer. The activation function we used for this network is RELU except for the output layer. The activation function we used for output layer is SOFTMAX.

In the last layer the dimension changes from 2 to 1, which mean the data is combined and only one result is taken out from the system. The outcome result is about chemotherapy along with the criticality due to cancer tissues.

It gives an idea about how a patient responds to the Chemotherapy. If the patient does not respond to the chemotherapy, then it suggests the need for surgery. Residual Cancer Burden (RCB) class suggests the criticality of the patient after first Chemotherapy.

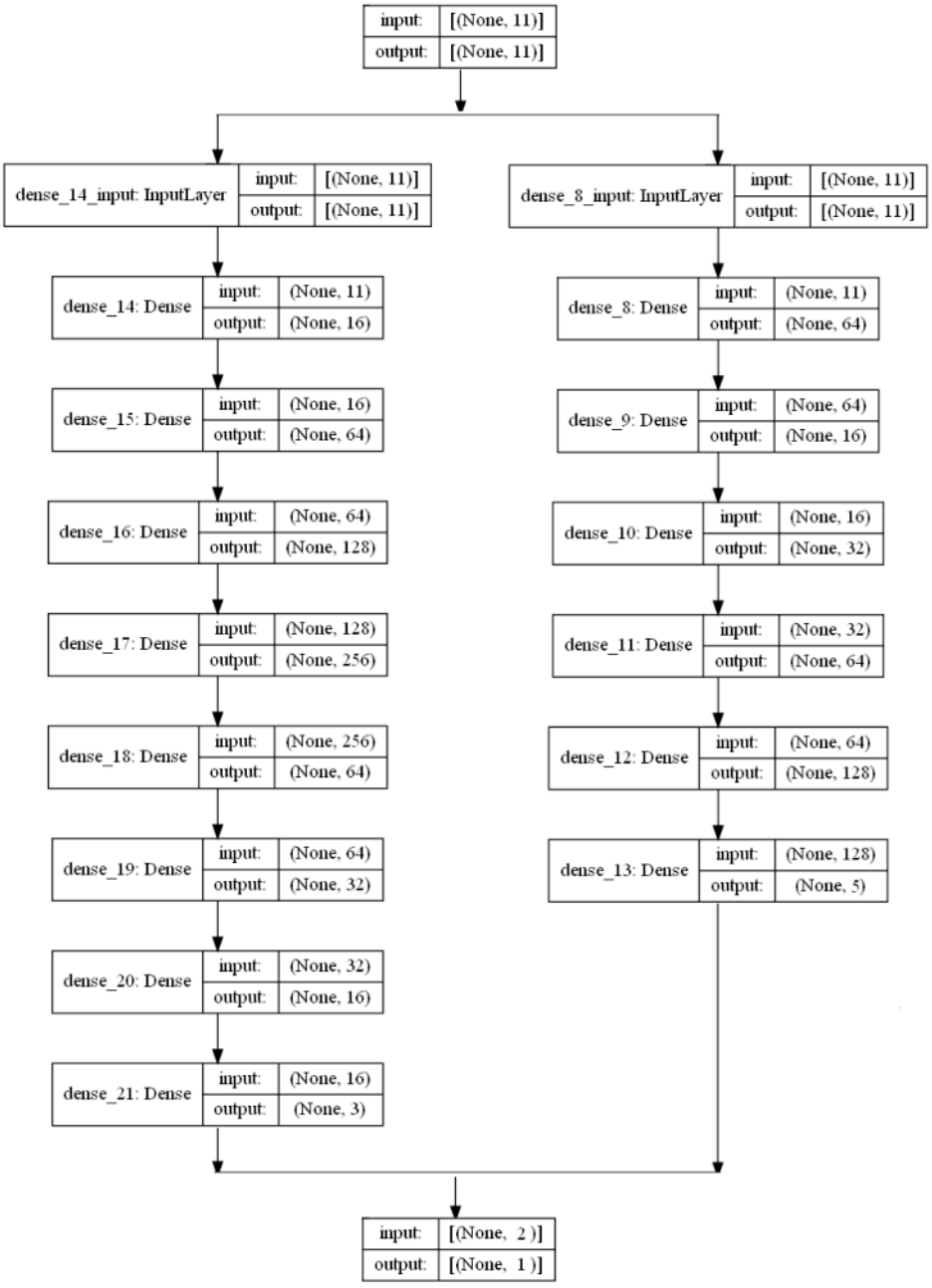


Fig. 3.2: Proposed system architecture

The Naive Bayes Classifier (NBC) is a very valuable Bayesian learning strategy. This classifier uses Bayes' hypothesis to accept freedom between indicators. Basically, NBC accepts that the existence of a property in one class is freed from the existence of another class. Bayesian hypothesis finds accidental probabilities. Numerically, this is in condition (1). (19). The factors of conditions (19), (20), and (21) are expressed as follows. I.) $P(Y|X)$ is the inverse likelihood of class Y given the indicator (X). II.) $P(Y)$ is the prior probability of the class. III.)

Table 3.3: Mammogram data composition

	Whole Dataset		Training Set		Test Set	
	Breast n	Patient n	Breast n	Patient n	Breast n	Patient n
Overall	3002	1501	2701	1484	301	284
Non-Malignant	2465	1496	2218	1427	247	235
Malignant	537	532	483	478	54	54
A	152	76	132	74	20	18
Breast B	594	297	532	292	62	57
density C	1560	780	1405	774	155	149
D	696	348	632	344	64	60

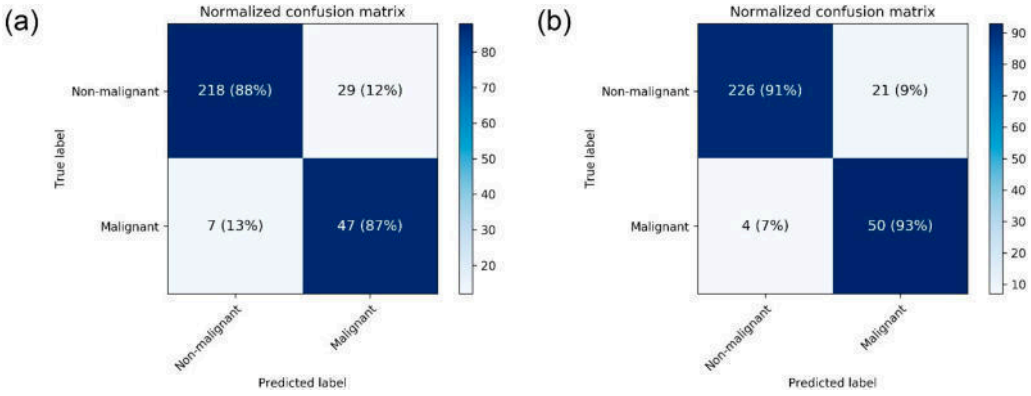


Fig. 3.3: Breast cancer detection heat map of confusion matrix

$P(X|Y)$ is the probability of the indicator, also known as the probability. IV.) $P(X)$ is the past probability of the indicator.

3.1. Clinical Subject Demographics. Finally, 3002 in total combination 1501 patients' mammograms were examined. recalled for review. The normal age of the members was 48.9 ± 11.1 years. The entire dataset contained 537 malicious images and 2465 non-malicious images [15]. The dangerous cancer group was more experienced than the non-dangerous growth (52.7 11.2 vs. 48.1 10.9 years; p 0.001) group. A number of many images of big breasts of thickness a C or a D (2256, 75.1 percent). Table 3.3 shows the information structure of the prepared and test datasets. The test dataset contained 54 images named Malicious Images and 301 images, including 247 non-vengeful images. The normal time for members to take mammograms on the 49.9 10.9 years made up the test dataset.

3.2. Detection of Breast Cancer by CNN Models' Performance. Table 3.3 shows the measurements for a typical CNN design exhibition. The recommended AUC for the most malignant breast growth location on the mammogram was changed to 0.952 ± zero half for the DenseNet 169 method and 0.954 ± 0.020 for the EfficientNet B5 method. The proposed accuracy was changed to 88.1 ± 0.2% for the DenseNet 169 method and 87.9 ± 4.7 for the EfficientNet B5 method. For the DenseNet169 model, the recommended responsiveness and specificity scores were 87.0 ± 0.0 and 88.4 ± 0.2, respectively. Figure 3.2 shows a standardized confusion framework for CNN structures that separates harmful images from harmless images.

4. Results and experimentation.

4.1. Grad-CAM. The results are tested on the image dataset also to verify the results collected in the form of excel sheet. The result of the classification for the image as input is shown in figure 3.3. Figure 4.1 shows an illustration of a GradCAM picture. The CNN model effectively distinguished harmful injuries. In

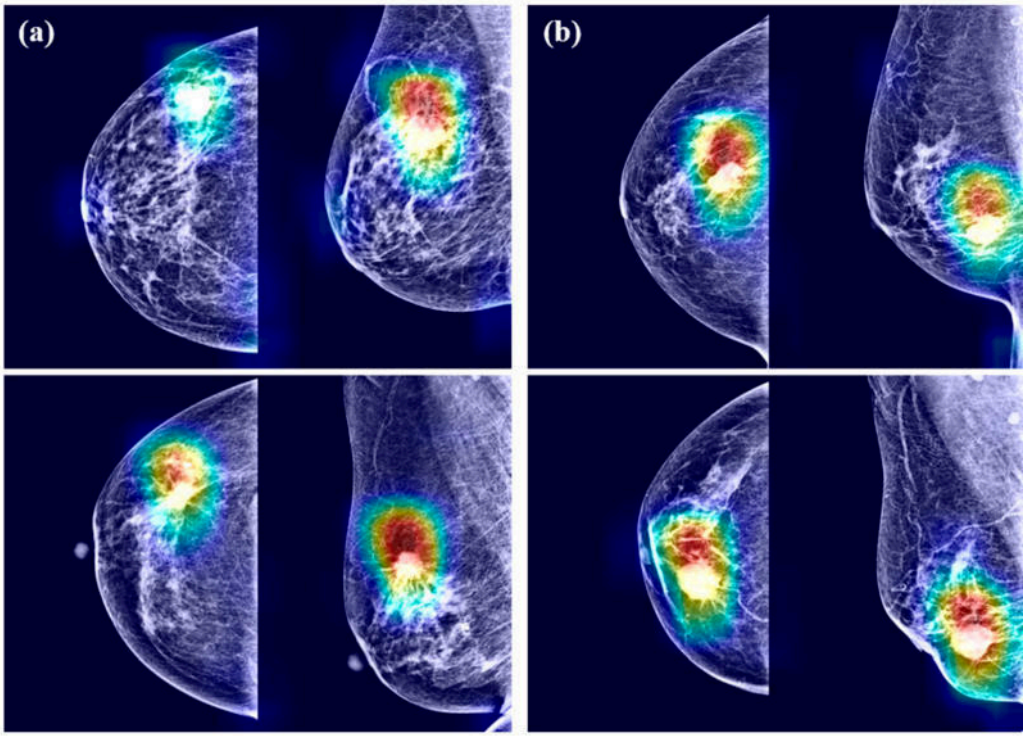


Fig. 4.1: Gradient-weighed class for mammograms by DenseNet-169 and (b) EfficientNet-B5

the combined mammogram, harmful sores were generally appropriately recognized in both caudal and average sidelong slanted sees. The CNN model zeroed in on the point of interaction between bosom disease and the encompassing parenchyma [16, 17, 18, 19]. Radiodensity regions offered more to disease forecast than radiation porous regions. Specifically, CNN models would in general recognize irregularities as mass and calcification instead of compositional mutilations and imbalances. Nonetheless, the radiation murky construction of the typical design was not thought of as being available in many images is logica GradCAM [20, 18, 19].

Look at the consequences of different grouping models and analyze them. It ought to be noticed that the outcomes showed in Table 4.1 have a mix of SVMLDA, NLDA, NNRFE, NNPCA, NLDA and NNRDA and NN alone, with the best exhibition of 99.07% in regards to explicitness. In any case, the blend of SVMLDA and the mix of NNLDA (98.41%), exactness (98.41%), update (98.41%), accuracy (98.41%), callback (98,41%), callback (98.41%) I got the best presentation as far as And precision (98.82%). The most unfortunate fitting ML model for responsiveness and review was a mix of Nb-CFS, and precision, particularity, exactness, Nb and Nb were scant of RFE pretreatment information. Likewise, dimensionalization influences the exhibition of the ML calculation. Reproduction results show that the CFS and RFE techniques work on the precision and particularity of the outspread piece SVM, and the LDA works on the two its exactness and awareness (discovery of threatening cases). For ANN, CFS increments awareness, and RFE, PCA, and LDA increment responsiveness and arrangement precision. NB exactness is CFS and LDA have improved, and awareness is upgraded by LDA. AI grouping calculations are additionally analyzed utilizing their particular ROC plots and the region underneath the kappa esteem. SVMLDA, NNLDA, and NBLDA had the best locales under the ROC bend. The worth is 0.9994 and its superior execution is shown. Furthermore, the best kappa worth of 0.9748 was gotten from SVMLDA and NLDA. Aspect decrease is vital in the arrangement cycle. Highlight extraction assumes a significant part in arrangement models. The primary motivation behind performing highlight extraction ought to work on prescient execution and guarantee quicker expectation. In this manner, the advantages of component extraction can't be overstated. Highlight extraction makes it simpler to envision

Table 4.1: Summary of machine learning models' results

Machine Learning Models	Accuracy	Area under ROC Curve	Precision	Recall	Sensitivity	Specificity	Kappa
SVM	0.9647	0.9964	0.9385	0.9682	0.9682	0.9626	0.9248
SVM-CFS	0.9647	0.9954	0.9524	0.9524	0.9524	0.972	0.9243
SVM-RFE	0.9637	0.9966	0.9534	0.9534	0.9534	0.962	0.9253
SVM-LDA	0.9872	0.9984	0.9871	0.9871	0.9871	0.9917	0.9448
NN	0.9746	0.9975	0.9733	0.9465	0.9465	0.9807	0.9463
NN-CFS	0.9606	0.9873	0.9431	0.9783	0.9783	0.9619	0.9472
NN-RFE	0.9724	0.9979	0.9739	0.9673	0.9673	0.9927	0.972
NN_PCA	0.9755	0.9947	0.967	0.9724	0.9554	0.9977	0.9592
NN-LDA	0.9782	0.9894	0.9741	0.9741	0.9741	0.9807	0.9748
NB	0.9218	0.976	0.885	0.8789	0.8789	0.9452	0.8015
NB-CFS	0.9276	0.9599	0.9052	0.8471	0.8471	0.9433	0.8711
NB-RFE	0.9118	0.986	0.875	0.8889	0.8889	0.9352	0.8115
NB-LDA	0.9724	0.9894	0.9739	0.9783	0.9583	0.9807	0.952

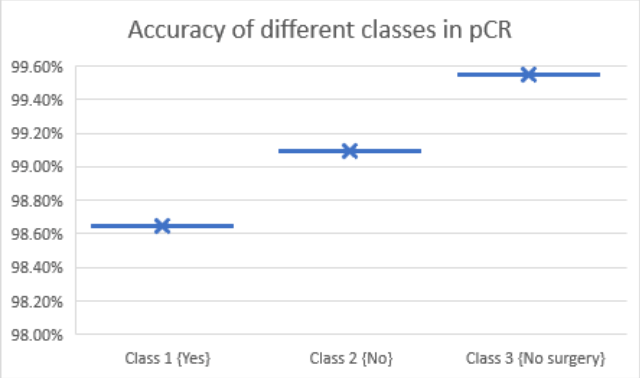


Fig. 4.2: Calculated accuracy of classes in pCR using proposed architecture

and get your information. Furthermore, highlight extraction lessens memory necessities and preparing time. Results are tested on different machine learning and neural network algorithms. From table 4.1 it is clear that the performance parameters of the machine learning algorithms are much lesser than that of the neural network algorithms. This is the reason behind selecting the neural network approach over machine learning.

Complete database is tested by using proposed architecture. It gives the around 99% accuracy for the prediction of pCR. The integrated 3 classes accuracy is visualized in figure 4.2. Accuracy for PCR classes of class 1, class 2 and class 3 are 98.62%, 99.10% and 99.55% respectively. As compared to other classes class 3 is having 99.5% accuracy which the highest. As the average accuracy of all the classes comes out to be 99%, it is highly reliable system.

Similarly, the same data is tested for RCB, and the accuracy for different classes is as them in fig 4.3. Accuracy for RCB classes of Class0, Class 1, Class 2, Class 3 and Class 4 are 99.55%, 100%, 98.20%, 99.10% and 99.55% respectively. It is shows that the class 1 is having the maximum accuracy while class 2 is having minimum accuracy. Except class 2, all the classes are having accuracy more than 99%.

The overall pCR and RCB classification accuracy come out to be 99.095022% and 99.276018% respectively. Precision value of pCR classification is higher than that of the RCB classification. In case of Recall and F1

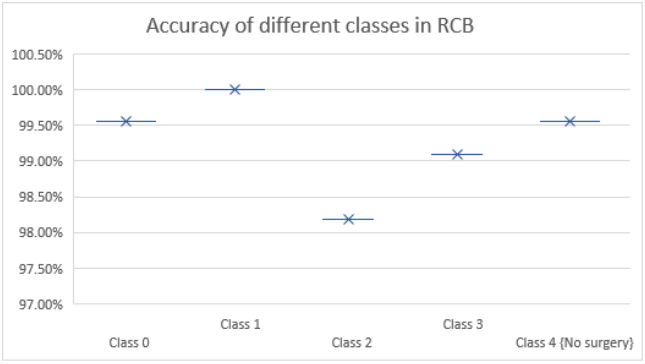


Fig. 4.3: Calculated accuracy of classes in RCB using proposed architecture

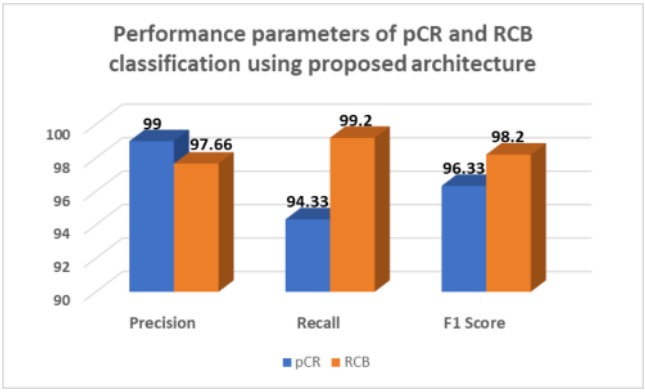


Fig. 4.4: Performance parameters of proposed architecture for classification of pCR and RCB

Table 4.2: Class wise performance parameters of pCR classification

Class	Precision	Recall	F1-Score
Class 1 {Yes}	99%	99%	99%
Class 2 {No}	98%	98%	98%
Class 3 {No Surgery}	100%	86%	92%

score value of RCB classification is higher than that of the pCR classification.

Table 4.2 and table 4.3 gives class wise performance parameters for pCR and RCB classification utilising the suggested architecture.

Precision, recall and F1 score of the class 1 is 100% it means changes of misclassification of class 1 is zero. Precision of class 1 and class 2 is maximum which is 100%. Except class 2 all the classes is having recall value is equal to 100%. The minimum value of F1 score is 97% for class 3 and class 4.

Same database is trained and tested using SVM and decision tree algorithm. Proposed system gives much higher accuracy than the SVM and decision tree algorithm.

Same database is trained and tested using different existing algorithms like naive bayesain, random forest, KNN, SVM and decision tree. But, in figure 4.5 presented comparsion with svm and DT because among all these two gave better performance. Proposed system gives much higher accuracy than the other existing algorithms.

Table 4.3: Class wise performance parameters of RCB classification

Class	Precision	Recall	F1-Score
Class 0	98%	100%	99%
Class 1	100%	100%	100%
Class 2	100%	96%	98%
Class 3	95%	100%	97%
Class 4 {No Surgery}	95%	100%	97%

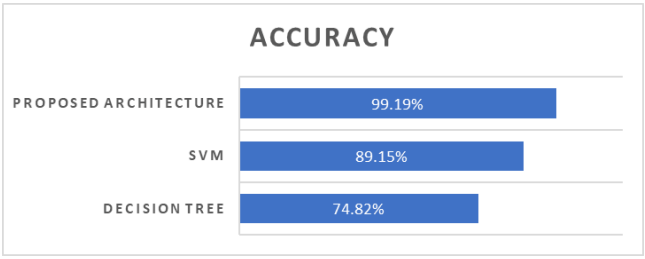


Fig. 4.5: Comparison of existing and proposed architectures

Table 4.4: The duration required to obtain data from different hardware systems

Platform	Time required to get result (in seconds)
CPU, i3 processor, 8GB RAM	4.395
CPU, i5 processor, 8GB RAM	4.163
CPU, I7 processor, 8GB RAM	3.124
GPU, Nvidia K80	0.962

Time complexity of the proposed system architecture is tested on different hardware platforms [35]. The time required to get the result is tabulated in table 4.4.

5. Conclusion. From all the results, one can easily conclude that the proposed architecture is much accurate to detect the effect of chemotherapy treatment on the patient. It gives the necessity of chemotherapy or not. If the patient does not respond to the chemotherapy, then the proposed system itself provides suggestion to go for surgery. Moreover, system also correctly predicts the criticality level so proper medication can be suggested by the doctors. In this article, we examined the WDBC dataset utilizing dimensionality decrease strategies and three normal ML calculations to order dangerous and harmless cancers. Trial work has shown that arrangement execution relies upon the ML order strategy picked. Reproduction results show that SVMLDA and NNLDA are better than other ML classifier models. By the by, SVMLDA is picked over NNLDA on the grounds that NNLDA requires longer calculation time. Hence, this paper proposes a keen methodology that coordinates straight discriminant investigation and backing vector machines (utilizing the RBF piece) for bosom malignant growth finding. This chose approach has shown great and promising outcomes for approval datasets. Accomplished arrangement exactness of 98.82%, responsiveness of 98.41%, explicitness of 99.07%, and region under the recipient movement bend of 0.9994. This study shows that highlight determination and component extraction can assist with working on the finding of harmless and dangerous cancers utilizing AI procedures. Hence, this work presumes that coordinating critical dimensionality decrease strategies with ML grouping methods gives a superior way to deal with clinical determination (bosom malignant growth conclusion utilized as a contextual analysis). The primary thought is to consolidate dimensionality decrease with the advantages of the ML calculation. Future work might plan to form the picked approach into a potential viable

method for helping doctors in diagnosing bosom disease and aid a fast second assessment. Future work may likewise consider contrasting other ML calculations utilized with analyze bosom disease. Other illness choices may likewise be considered in ongoing examinations.

REFERENCES

[1] J. R. Benson and I. Jatoi., *The global breast cancer burden*, Future oncology, 2012, vol. 8, no. 6, pp. 697–702.

[2] L. J. Fallowfield, S. K. Leaity, A. Howell, S. Benson, and D. Cella., *Assessment of quality of life in women undergoing hormonal therapy for breast cancer: validation of an endocrine symptom subscale for the fact-b*, Breast cancer research and treatment, 1999, vol. 55, no. 2, pp. 187– 197.

[3] A. R. Costa, F. Fontes, S. Pereira, M. Gonc¸alves, A. Azevedo, and N. Lunet., *Impact of breast cancer treatments on sleep disturbances– a systematic review*, The Breast, 2014, vol. 23, no. 6, pp. 697–709.

[4] J. E. Carroll, K. Van Dyk, J. E. Bower, Z. Scuric, L. Petersen, R. Schiestl, M. R. Irwin, and P. A. Ganz., *Cognitive performance in survivors of breast cancer and markers of biological aging*, Cancer, 2019, vol. 125, no. 2, pp. 298–306.

[5] P. Minicozzi, L. Van Eycken, F. Molinie, K. Innos, M. Guevara, R. Marcos-Gragera, C. Castro, E. Rapiti, A. Katalinic, A. Torrella et al., *Comorbidities, age and period of diagnosis influence treatment and outcomes in early breast cancer*, International journal of cancer, 2019, vol. 144, no. 9, pp. 2118–2127.

[6] J. Teng, A. Abdygametova, J. Du, B. Ma, R. Zhou, Y. Shyr, and F. Ye., *Bayesian inference of lymph node ratio estimation and survival prognosis for breast cancer patients*, IEEE Journal of Biomedical and Health Informatics, 2019, pp. 1–1.

[7] P. V. C. Souza, A. G. dos Reis, G. R. R. Marques, A. J. Guimaraes, V. J. S. Araujo, V. S. Araujo, T. S. Rezende, L. O. Batista, and G. A. da Silva., *Using hybrid systems in the construction of expert systemsin the identification of cognitive and motor problems in children and young people*, in 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), June 2019, pp. 1–6.

[8] C.-W. Han., *Fuzzy neural network-based fetal health monitoring using cardiotocography data*, in AIP Conference Proceedings, vol. 2104, no. 1. AIP Publishing, 2019, p. 030008.

[9] Y. Liu, S. Wu, Kuang-Pen Chou, Y. Lin, Jie Lu, Guangquan Zhang, Wen-Chieh Lin, and C. Lin., *Driving fatigue prediction with pre-event electroencephalography (eeg) via a recurrent fuzzy neural network*, in 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), July 2016, pp. 2488–2494.

[10] P. V. d. C. Souza and A. J. Guimaraes., *Using fuzzy neural networks for improving the prediction of children with autism through mobile devices*, in 2018 IEEE Symposium on Computers and Communications (ISCC), June 2018, pp. 01 086–01 089.

[11] A. J. Guimaraes, V. J. S. Araujo, V. S. Araujo, L. O. Batista, and P. V. ˜ de Campos Souza., *A hybrid model based on fuzzy rules to act on the diagnosed of autism in adults*, in Artificial Intelligence Applications and Innovations, J. MacIntyre, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds. Cham: Springer International Publishing, 2019, pp. 401–412.

[12] P. V. de Campos Souza, A. J. Guimaraes, V. S. Araujo, T. S. Rezende, and V. J. S. Araujo., *Using Fuzzy Neural Networks Regularized to Support Software for Predicting Autism in Adolescents on Mobile Devices*, Singapore: Springer Singapore, 2019, pp. 115–133.

[13] V. J. Silva Araujo, A. J. Guimar aes, P. V. de Campos Souza, ˜ T. Silva Rezende, and V. Souza Araujo., *Using resistin, glucose, age ´ and bmi and pruning fuzzy neural network for the construction of expert systems in the prediction of breast cancer*, Machine Learning and Knowledge Extraction, 2019, vol. 1, no. 1, pp. 466–482.

[14] Uddaraju, Susmitha, and M. Narasingarao., *A survey of machine learning techniques applied for breast cancer prediction*, International Journal of Pure and Applied Mathematics, 2017, 117.19, pp. 499-507.

[15] Susmitha, Uddaraju., *a review of machine learning frameworks for early and accurate prediction of neoadjuvant chemotherapy responses*, European Journal of Molecular & Clinical Medicine, 2020, 7.4, pp. 1040-1050.

[16] Uddaraju, Susmitha, and M. R. Narasingarao., *Predicting the Ductal Carcinoma Using Machine Learning Techniques—A Comparison*, Journal of Computational and Theoretical Nanoscience, 2019, 16.5-6, pp. 1902-1907.

[17] S. L. Bangare, S. T. Patil et al., *Implementing Tumor Detection and Area Calculation in MRI Image of Human Brain Using Image Processing Techniques*, Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 4, (Part -6) April 2015, pp.60-65.

[18] K. Gulati et. al., *Use for Graphical User Tools in Data Analytics and Machine Learning Application*, Turkish Journal of Physiotherapy and Rehabilitation; 32(3), ISSN 2651-4451, e-ISSN 2651-446X.

[19] LMI Leo Joseph et. Al., *Methods to Identify Facial Detection in Deep Learning Through the Use of Real-Time Training Datasets Management*, EFFLATOUNIA - Multidisciplinary Journal, ISSN:1110-8703, Volume 5, Issue 2, pp.1298 -1311.

[20] G. Awate et. al., *Detection of Alzheimers Disease from MRI using Convolutional Neural Network with Tensorflow*, https://arxiv.org/abs/1806.10170.

[21] Shachi Mall et. al., em Implementation of machine learning techniques for disease diagnosis, Materials Today: Proceedings, 2021, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.11.274.

[22] *Breast cancer in young women by Australian government*, Breast cancer in young women statistics | Breast Cancer in Young Women (canceraustralia.gov.au).

Edited by: Vinoth Kumar
Received: Jun 21, 2022
Accepted: Dec 2, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 225–232, DOI 10.12694/scpe.v23i4.2022

MULTIMODAL MEDICAL IMAGE FUSION USING HYBRID DOMAINS

A.RAJESH NAIDU, D.BHAVANA *

Abstract. In a variety of clinical applications, image fusion is critical for merging data from multiple sources into a single, more understandable outcome. The use of medical image fusion technologies to assist the physician in executing combination procedures can be advantageous. The diagnostic process includes preoperative planning, intra operative supervision, an interventional treatment. In this thesis, a technique for image fusion was suggested that used a combination model of PCA and CNN. A method of real-time image fusion that employs pre-trained neural networks to synthesize a single image from several sources in real-time. A innovative technique for merging the images is created based on deep neural network feature maps and a convolution network. Picture fusion has become increasingly popular as a result of the large variety of capturing techniques available. The proposed design is implemented using deep learning technique. The accuracy of the proposed design is around 15% higher than the existing design. The proposed fusion algorithm is verified through a simulation experiment on different multimodality images. Experimental results are evaluated by the number of well-known performance evaluation metrics

Key words: Image fusion, PCA, CNN, VGG 16, VGG19

AMS subject classifications. 68U10

1. Introduction.

Medical imaging may provide information in a number of ways. For clinical diagnosis, it is vital to keep an eye on the distinguishing aspects of different medical images. Image fusion is a method used to produce a single image by integrating the features of numerous image sensors. This information may come from a single source across many time periods or from a number of sensors during a single period. Fusion is the process of joining two or more things together to form a single entity[4],[10]. To improve the use of medical images and assist clinicians in interpreting image content, medical image fusion attempts to extract as much useful information as possible from source images[8]. Image fusion is the process of fusing two or more images into a series of images that incorporates the data from the individual images. The result is an image with higher information content than any of the input images. The purpose of image fusion is to create images that are more acceptable and understandable to both humans and machines.

MRI scanners employ magnetism, high magnetic fluctuations, and radio waves to generate images of the inside of the body's tissues[5]. An MRI scan that detects structural issues before they become irreparable may help to avoid nerve damage. Neuropsychological testing is often used to identify neural injury [18]. Medical image fusion was used to merge images from several imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), and proton emission tomography (PET) [6].

To eliminate superfluous data, source register images are joined with other important data in an image fusion process. To do quantitative analysis, entropy and other statistical approaches are applied. Liguo Zhang has developed a complete deep learning model for extracting features, merging features, and restoring pictures that does not involve the development of complex feature matching and fusion criteria [1]. According to the Associated Press, Jamesa provides a comprehensive list of possibilities as well as an explanation of the several scientific obstacles that medical image fusion faces [2].

Deep learning has been used to address some of the most difficult challenges in medical imaging. Simultaneous PET/MR imaging combines the soft-tissue contrast of MR imaging with the molecular sensitivity and specificity of PET. Obtaining a valid photon attenuation correction (AC) map, which is critical for accurate PET quantification, is still a difficulty [3] [19-22]. Because traditional MR imaging-based AC (MRAC) techniques lack direct bone measurement, a number of new methodologies have been used.[15] PET and SPECT imaging

*Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur District, India (bhavanaece@kluniversity.in)

applications of AI-based algorithms range from low-level electronic signal formation and processing to high-level internal dosimetry and diagnostic/prognostic modelling. Deep learning techniques have mostly been employed to improve incident photon timing resolution and localization accuracy with the objective of enhancing overall spatial and time-of-flight (TOF) resolutions in PET. The purpose of the research being done on quantitative SPECT and PET imaging right now is to get rid of the effects of noise, artifacts, and movement [16].

A wide variety of computer vision problems, including identification, fragmentation, and very low-level recognition, have been improved through deep learning. It's called image fusion when two or more separate images are combined to create a new image. To enhance the original multi-spectral image's resolution, it takes data from several sources while still conserving spectral information. Radiology uses computed tomography, or CT, to produce detailed images of the inside of the human body without the need for any incisions [9]. It's impossible to view your bones, muscles, and other biological parts in the same way that this image does. Tumors may be identified on CT scans by their shape, size, and location.

Utilize contourlet transforms and adaptive weighted PCA in the frequency domain to create a multimodal fusion. Describe the fusion mechanism for the DTCWT and SOFM modalities. Describe the use of NSCT-based image fusion for clinical research. For the purpose of improving input photos, the hybrid fusion method is introduced. Review and elucidate the fusion algorithms overview. developed a hybrid approach for accurate treatment analysis based on a blend of curvelet and wavelet transforms. To combine structural and functional information from the source images, introduce the fusion method. A-PCNN has been developed, and NSST has been suggested for multimodality medical pictures [8]. Describe the NSCT for multimodality medical image fusion. a curvelet-based fusing algorithm was invented [10]. An algorithm for fusing multimodal medical images has been created [11]. The DFRWT technique is described here for effective image fusion [12]. Create and modify the surface of the fusion system [13]. Describe the fusion technique for medical pictures based on SWT and NSCT [14].

2. Proposed Methodology. In this paper, the proposed method is a combination of CNN and PCA. Radiological images were collected, followed by normalization, which enhances both computation efficiency and execution. The normalised images were then fed into the CNN algorithm for fusion, and the fused images were then gathered and submitted to the PCA method for dimensionality reduction. the whole process of proposed model is shown in figure 2.1. As a result, the technique is now known as a CNN + PCA combine model, which produces superior results than other techniques. The normalized image is submitted to VGG16, VGG19, SqueezeNet, and PCA to verify the accuracy of alternative algorithms for image fusion, although the results are less accurate than the combination model of CNN+PCA. For image fusion, several models of radiological imaging such as MRI, CT, PET, and SPECT were collected. Two different format photographs are used in this image fusion.

2.1. Convolutional Neural Network. Convolutional networks, also known as convolutional neural networks, are a type of neural network that uses a grid-like design to process input (LeCun, 1989). In real-world applications, convolutional networks have had a lot of success. A "convolutional neural network" is a network that uses the convolutional mathematical approach as shown in figure 2.2. It's a very specialized linear procedure. Simple neural networks with a minimum of one layer that simply use convolution instead of generic matrix multiplication are used in these networks. Convolutional networks were among the first neural networks to solve key business problems, and they continue to be at the core of commercial deep learning applications today.[14] A convolutional neural network consists of several layers which are classified as the Input layer, Convolutional layer, Max Pooling layer, flatten layer, Output layer which helps us to extract features for the fusion of images. A tiny portion of the visual field reacts to stimuli in the Receptive Field which are known as neurons. A ring of comparable fields forms around the entire field of vision.

2.2. VGG16. It is a deep Convolutional Neural (CNN) architecture with several layers. VGG means Visual Geometry Group.VGG-16 is a deep convolutional neural network with 16 layers. It has a very appealing architecture due to its uniformity. It features only 3x3 convolutions, but a lot of filters, similar to AlexNet [19]. It may be taught for two to three weeks on 4 GPUs. It is now the most popular method for extracting characteristics from images in the communityIn ImageNet, The top-5 precision of the VGG16 model was 92.7 percent. The VGG16 Architecture was invented and presented by Karen Simonyan with Andrew Zisserman of

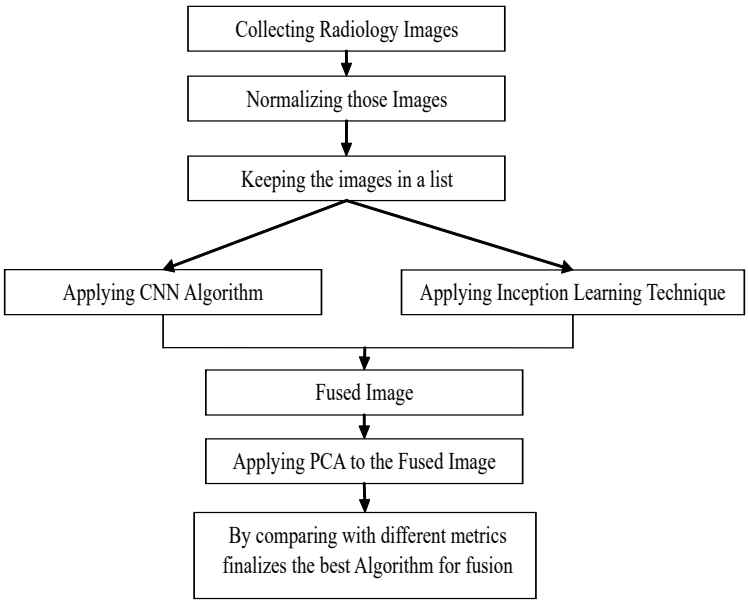


Fig. 2.1: Proposed Model

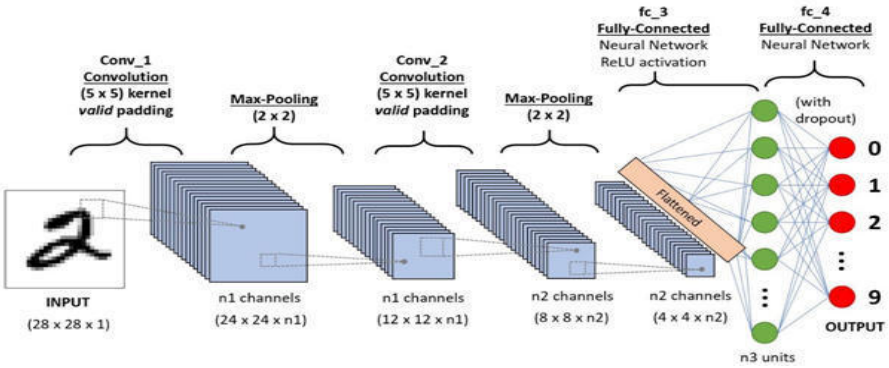


Fig. 2.2: CNN sequence to classify handwritten digits

Oxford University in their article "Very Fully Convolutional Networks for Large Object Recognition" in 2014.

2.3. VGG19. The VGG19 model is a VGG version with nineteen layers which is a combination of sixteen convolution layers, three fully linked layers, five max pool layers, and one Soft Max layer. The VGG19 (also known as VGGNet-19) is a model similar to VGG16 only the difference is it contains 19 layers [11,13]. The weight layers of the model are represented by the digits "16" and "19". In comparison to VGG16, VGG19 has three more convolution layers. VGG19 is a complex CNN having pre-trained phases and a deep understanding of how a picture is described in terms of appearance, color, and structure. VGG19 is a neural network that was learned on millions of images to solve difficult classification issues. For a wide range of images, the system has learned rich feature representations.

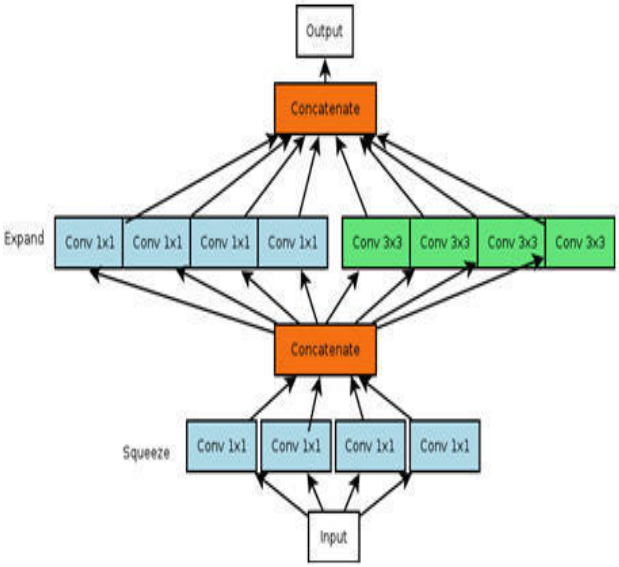


Fig. 2.3: CNN sequence to classify handwritten digits

2.4. SqueezeNet. SqueezeNet is a deep convolutional neural network with eighteen layers. SqueezeNet is a form of neural network that uses design tactics to minimise dimensionality, most notably the usage of flame components, which use 1x1 CNNs to "squeeze" variables. SqueezeNet was created with the goal of creating a compact neural network with fewer elements that could fit into memory storage and be conveyed more easily over a computer network. SqueezeNet is a small network designed to be a more compact version of AlexNet. It has more than 50 times less parameters than AlexNet yet is three times faster [12]. The SqueezeNet design is made up of the "squeeze" and "expand" layers. In a squeeze convolutional layer, only one filter is used. In an extended layer, these are routed via a mix of 1 1 and 3 3 convolution filters. SqueezeNet was trained with a learning method of 0.04, which lowers linearly over time. SqueezeNet makes the setup process easier because of its small size. This network was created in Caffe, but it has since gained popularity and has been adopted by a number of platforms.

2.5. Principal Component Analysis. Principal Component Analysis is an unsupervised learning approach for reducing dimensionality in machine learning. It is a mathematical method that uses orthogonal information to turn correlated attribute outputs into a collection of uncorrelated, linear qualities. The Principal Components are the newly amended features.

It is necessary to limit the number of variables and interpret linear combinations of data in order to understand data with a large number of variables in a meaningful way. PCA is a method for reducing a large number of potentially associated variables to a smaller number of variables known as principal components [7]. A method for condensing a large number of samples into a smaller set that preserves the majority of the data from the larger set is principal component analysis. We can utilize the principal component analysis technique to create and use a smaller set of variables called primary factors. It's much easier to research and interpret a smaller set of data.

3. Performance Evaluation. The source images are obtained using several radiological techniques from a normal brain. The fuse output is formed again by integrating its inverted output of the resultant connection with actual U and V segments, using real brain MRI and PET input images as a source for testing. For source images, the process is simple to implement; nevertheless, the fusion result is not flawless. The fusion image is unstable because the information in the fused image changes from one algorithm method to the next, while the data image has many modalities depending on the algorithm utilized. The hue of the PCA image changes when

compared to other fused images. Because we are using PCA, the generated fused image will be in grayscale mode. In this section, we will discuss the comparative analysis of different algorithm methods using various metric indexes. The comparative analysis is based on source images and fused image with help of comparing algorithm methods including the add-weighted average of them. The fused images were created utilizing several methods such as VGG-16, VGG-19, Squeeze Net, PCA, and CNN+PCA.

The following metrics can be used to assess the performance of image fusion algorithms:

Mean Square Error [17] (MSE):

$$MSE = \frac{1}{X * Y} \sum_{m=0}^{X} \sum_{n=0}^{Y} (S_1(m,n) - S_f(m,n))^2 \tag{3.1}$$

where $X * Y$ image size. $S_1(m,n)$ is the input image and the fused image.

Structural Similarity Index Measure [17 (SSIM)]

$$SSIM = \frac{(2\mu_{S_r}\mu_{S_f} + K_1)(\sigma_{S_r}\sigma_{S_f} + K_2)}{(\mu_{S_r}^2 + \mu_{S_f}^2 + K_1)(\sigma_{S_r}^2 + \sigma_{S_f}^2 + K_2)}) \tag{3.2}$$

where $\mu(S_r)^2 + \mu_(S_f)^2$ is close to 0, the constant K_1 is used to prevent instability and when $\sigma_(S_r)^2 + \sigma_(S_f)^2$ is close to 0, the constant K_2 is used to prevent instability

$\mu_{S_r)} = \frac{1}{XY} \sum_{m=0}^{X} \sum_{n=0}^{Y} S_r(m,n)$

$\mu_{S_r)} = \frac{1}{XY} \sum_{m=0}^{X} \sum_{n=0}^{Y} S_f(m,n)$

$\sigma_{S_r}^2 = \frac{1}{(XY-1)} \sum_{m=1}^{X} \sum_{n=0}^{Y} (S_r(m,n) - \mu_{(S_r)})^2$

$\sigma_{S_f}^2 = \frac{1}{(XY-1)} \sum_{m=1}^{X} \sum_{n=0}^{Y} (S_f(m,n) - \mu_{(S_f)})^2$

$\sigma_{S_r S_f} = \frac{1}{(XY-1)} \sum_{m=0}^{X} \sum_{n=0}^{Y} (S_r(m,n) - \mu_{(S_r)})(S_f(m,n) - \mu_{(S_f)})$

Visual Information Fidelity [17] (VIF): It is the parameter for evaluating image quality that is based on information aspects of natural scenes statistical (NSS).

$A(p_{i,k}, q_{i,k})$= actual image details

$A(p_{i,k}, r_{i,k})$ = unbalanced image details

$p_{i,k}, q_{i,k} and r_{i,k} = p, q$, and r in images frames in a variety of sizes

M_k= scale's amount of image pixels k

s = the variety of categories in which the image is split

Mutual Information [17] (MI) Mutual information is used to measure the quality of the fused image, it is given as:

$$MI = I(S_X; S_f) + I(S_Y; S_f) \tag{3.3}$$

where $(S_X; S_f) = \sum_{m=1}^{M} \sum_{n=1}^{M} h_{R,F}(m,n) \log_2 \frac{(h_{R,F}(m,n))}{(h_R(m) h_F(n))}$, where $h_R(m), h_F(n)$, are S_R and normalised grey level histograms, which correspond to the reference and fused images, respectively. When the MI value of a fused image is higher, it means the image has more features and texture information.

Entropy The expected value along with the random variable is measured using entropy [17]. The image's color histogram can be thought of as a probability density function. The detailed entropy of a picture is represented with $h(i)$ where $h(i)$ reflects the quantized color's % of pixel i in the overall image.

$$Entropy = \sum_{i=1}^{n} h(i) \log_2 h(i) \tag{3.4}$$

4. Results. The comparative analysis of different algorithm methods using various metric indexes. The obtained results indicates the proposed algorithm is the best algorithm method among the tested algorithms. How to accurately evaluate the efficiency of fusion methods is an essential topic in picture fusion. In recent years, several other assessment indexes have been proposed. There are two types of assessment methods now in use,

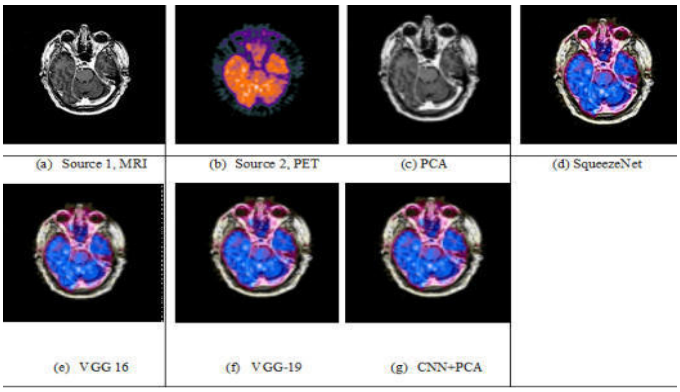


Fig. 4.1: Normal brain MRI,PET medical images fusion results

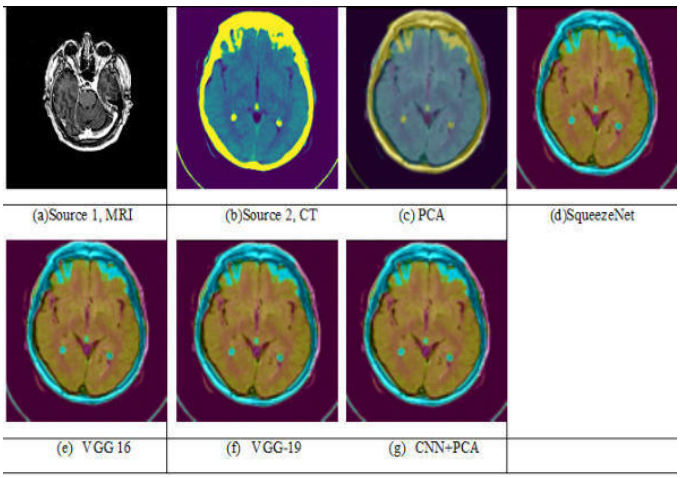


Fig. 4.2: Normal brain MRI and CT medical images fusion results

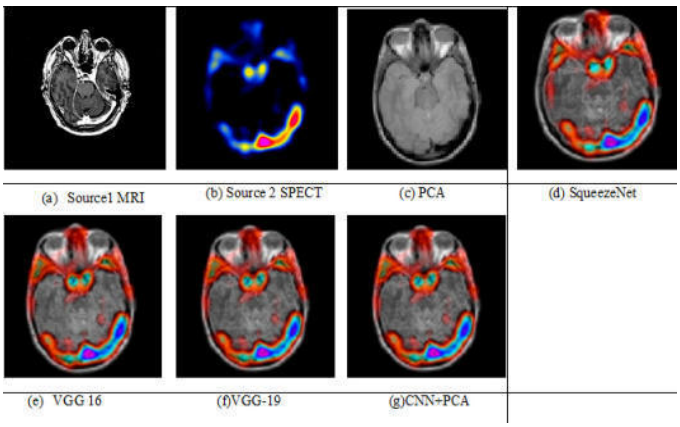


Fig. 4.3: Normal brain MRI and SPECT medical images fusion results

Table 4.1: MRI-PET Metric Score [19]

	VGG-16	VGG-19	SQUEEZE NET	PCA	CNN+PCA
MSE	38.81896868	43.26268603	52.48530215	38.41147947	37.95749468
SSIM	0.634070678	0.599964187	0.576918209	0.593477189	0.631366731
VIF	0.020115451	0.026862346	0.017607955	0.022060916	0.082619463
Entropy	3.770389411	4.070330749	4.112055998	3.990689894	3.001083163
MI	0.436537877	0.435251605	0.371655424	0.467336177	1.145331026

Table 4.2: MRI-CT Metric Score

	VGG-16	VGG-19	SQUEEZE NET	PCA	CNN+PCA
MSE	52.56273155	46.16454714	53.09027707	22.221134702	20.85981113
SSIM	0.338170304	0.373294043	0.353296692	0.428951084	0.060058943
VIF	0.03505622	0.053515519	0.045902515	0.078706307	0.015024352
Entropy	6.420616688	6.279529215	6.173356001	5.989346358	4.957725674
MI	0.658167886	0.724539243	0.718558411	0.878390975	1.265098822

Table 4.3: MRI-SPECT Metric Score

	VGG-16	VGG-19	SQUEEZE NET	PCA	CNN+PCA
MSE	42.7168407	42.27611837	45.56879791	57.79010585	34.55778519
SSIM	0.486359572	0.473912872	0.470555961	0.591496644	0.48038565
VIF	0.047101677	0.044480608	0.037290363	0.107424537	0.048830219
Entropy	4.729863707	4.815029541	4.693647639	4.380235834	3.202091852
MI	0.362361713	0.324649206	0.308850748	0.586975232	1.573748002

Table 4.4: MRT1- MRT2 Metric Score

	VGG-16	VGG-19	SQUEEZE NET	PCA	CNN+PCA
MSE	51.78371531	54.17560435	30.95919525	47.35199752	26.68707336
SSIM	0.17952148	0.167316343	0.478195945	0.515009571	0.4094382486
VIF	0.031350142	0.024646476	0.062285785	0.075168443	0.0615285748
Entropy	5.695356778	5.763374805	5.290079636	5.277351219	3.9653502677
MI	0.493445079	0.459940614	0.531062598	0.626701992	1.119491017

the quality evaluation using a source image and a non-source image. Evaluation of image quality, we choose five metrics, for the accurate understanding of their outcomes Mean Square Error (MSE) be a part of the objective evaluation index that must be consulted the Structural Similarity Index (SSIM), Information Entropy (EN), Normalized Mutual Information (Qmi) and visual information fidelity (VIF) pixel, all of which are objective evaluation indicators. a few pairs of radiology medical images taken as the test data. For test data, applied filters for better understanding. The comparing algorithms are based on VGG-16, VGG-19, SQUEEZENET, Principal Component Analysis (PCA), and CNN+PCA. Among them, we randomly choose two source images with diverse modalities. The input image is transformed from RGB to HSI space at the time of testing, and various another format like grayscale, YCbCr etc.

5. Conclusion. The method extracts further information from source images using weights that are not shared. CNN is used with PCA in the proposed method for image reconstruction, and it also aids during

lowering the computational complexity and dimensionality of the original images. MSE loss is the basis for the proposed method loss function. Minimizing MSE loss raises the objective assessment index, however, this does not guarantee acceptable image fidelity. Its focus on the future discussion will be refining the loss function as well as enhancing its effectiveness images that were fused to get a bunch more durable fusion method according to CNN. The proposed model's fusion image can effectively fully preserve the useful information in the source images possible. The proposed method outperforms other comparison methods in terms of evaluating image fusion metrics. The proposed method is stable in terms of durability and may be applied to image fusion.

REFERENCES

[1] X. Liang, P. Hu, L. Zhang, J. Sun and G. Yin, "MCFNet: Multi-Layer Concatenation Fusion Network for Medical Images Fusion," in IEEE Sensors Journal, vol. 19, no. 16, pp. 7107-7119, 15 Aug.15, 2019

[2] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art", Inf. Fusion, vol. 19, pp. 4-19, Sep. 2014.

[3] Rajalingam, B., and Priya, R. (2018). Multimodal medical image fusion based on deep learning neural network for clinical treatment analysis. International Journal of ChemTech Research, 11(06), 160-176.

[4] E. Daniel, "Optimum Wavelet-Based Homomorphic Medical Image Fusion Using Hybrid Genetic–Grey Wolf Optimization Algorithm," in IEEE Sensors Journal, vol. 18, no. 16, pp. 6804-6811, 15 Aug.15, 2018, doi: 10.1109/JSEN.2018.2822712.

[5] R. Zhu, X. Li, X. Zhang, and M. Ma, "MRI and CT Medical Image Fusion Based on Synchronized-Anisotropic Diffusion Model," in IEEE Access, vol. 8, pp. 91336-91350, 2020.

[6] R. Nie, J. Cao, D. Zhou, and W. Qian, "Multi-Source Information Exchange Encoding With PCNN for Medical Image Fusion," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 3, pp. 986-1000, March 2021

[7] He, C., Liu, Q., Li, H., and Wang, H. (2010). Multimodal medical image fusion based on IHS and PCA. Procedia Engineering, 7, 280-285.

[8] Hermessi, H., Mourali, O., and Zagrouba, E. (2021). Multimodal medical image fusion review: Theoretical background and recent advances. Signal Processing, 183, 108036.

[9] Tan, W., Tiwari, P., Pandey, H. M., Moreira, C., and Jaiswal, A. K. (2020). Multimodal medical image fusion algorithm in the era of big data. Neural Computing and Applications.

[10] Bai L, Xu C, Wang C (2015) A review of fusion methods of multi-spectral image. Optik-Int J Light Electron Optics 126(24):4804–4807

[11] Thepade, S. D., and Dindorkar, M. R. (2022). Fusing deep convolutional neural network features with Thepade's SBTC for land usage identification. Engineering Science and Technology an International Journal, 27(101014), 101014.

[12] Ucar, F., and Korkmaz, D. (2020). COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Medical Hypotheses, 140(109761), 109761. doi:10.1016/j.mehy.2020.109761

[13] Victor Ikechukwu, A., Murali, S., Deepu, R., and Shivamurthy, R. C. (2021). ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. Global Transitions Proceedings, 2(2), 375–381. doi:10.1016/j.gltp.2021.08.027

[14] Gai, D., Shen, X., Chen, H., and Su, P. (2020). Multi-focus image fusion method based on two stage of convolutional neural network. Signal Processing, 176(107681), 107681.

[15] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsub sampled shearlet transform domain," IEEE Trans. Instrum. Meas., vol. 68, no. 1, pp. 4964, Jan. 2019.

[16] Venkatesan, B., and Ragupathy, U. S. (2022). Integrated fusion framework using hybrid domain and deep neural network for multimodal medical images. Multidimensional Systems and Signal Processing. doi: 10.1007/s11045-021-00813-9

[17] Naidu, V. P. S., M. Divya, and P. Mahalakshmi. "Multi-modal medical image fusion using multi-resolution discrete sine transform."Control and Data Fusion e-Journal 1, no. 2 (2017):13-26.

[18] B, R., and R, P. (2018). Multimodal medical image fusion using various hybrid fusion techniques for clinical treatment analysis. Smart Construction Research. doi: 10.18063/scr.v0.594

[19] Rajalingam, B., Priya, R., and Bhavani, R. (2019). Hybrid multimodal medical image fusion using combination of transform techniques for disease analysis. Procedia Computer Science, 152, 150–157. doi: 10.1016/j.procs.2019.05.037

[20] Yadav, Satya Prakash, and Sachin Yadav. "Image fusion using hybrid methods in multimodality medical images." Medical and Biological Engineering and Computing 58.4 (2020): 669-687.

[21] Alseelawi, Nawar, Hussein Tuama Hazim, and Haider TH Salim ALRikabi. "A Novel Method of Multimodal Medical Image Fusion Based on Hybrid Approach of NSCT and DTCWT." International Journal of Online and Biomedical Engineering 18.3 (2022).

[22] Singh, Sneha, and Radhey Shyam Anand. "Multimodal medical image fusion using hybrid layer decomposition with CNN-based feature mapping and structural clustering." IEEE Transactions on Instrumentation and Measurement 69.6 (2019): 3855-3865.

Edited by: Vinoth Kumar
Received: Jun 30, 2022
Accepted: Nov 11, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 233–249, DOI 10.12694/scpe.v23i4.2024

COMPUTER-AIDED DIAGNOSIS APPLIED TO MRI IMAGES OF BRAIN TUMOR USING SPATIAL FUZZY LEVEL SET AND ANN CLASSIFIER

VIRUPAKSHAPPA S*, SACHINKUMAR VEERASHETTY†, AND AMBIKA N‡

Abstract. The most vital organs in the human body are the brain, heart, and lungs. Because the brain controls and coordinates the operations of all other organs, normal brain function is vital. Brain tumour is a mass of tissues which interrupts the normal functioning of the brain, if left untreated will lead to the death of the subject. The classification of multiclass brain tumours using spatial fuzzy based level sets and artificial neural network (ANN) techniques is proposed in this paper. In the proposed method, images are preprocessed using Median Filtering technique, the boundaries of the Brain Tumor are obtained using Spatial Fuzzy based Level Set method, features are extracted using Gabor Wavelet and Gray-Level Run Length Matrix (GLRLM) methods. Finally ANN technique is used for the classification of the image into Normal or Benign Tumor or Malignant Tumor. The proposed method was implemented in the MATLAB working platform and achieved classification accuracy of 94%, which is significant compared to state-of-the-art classification techniques. Thus, the proposed method assist in differentiating between benign and malignant brain tumours, enabling doctors to provide adequate treatment.

Key words: Median Filter, Spatial Fuzzy Level Set, GLRLM, Gabor features, Artificial Neural Networks, Benign, Malignant and Computer-aided Diagnosis (CAD) system.

AMS subject classifications. 68T05

1. Introduction.

Various Medical Imaging technique are used to produce the images of human body. The medical experts provide the treatment by observing and analysing those images [1]. Different medical imaging techniques being used in daily life are Magnetic Resonance Imaging (MRI), Computerized Tomography, Ultrasonic and X-ray [2, 3] which applied individually through variety of sensory systems. Medical imaging also helps doctor in viewing inner portions of the body and diagnose properly. Doctors also gets help in making keyhole surgery where interior parts can be operated easily without opening the body more.

The neuroscience field has seen a significant increase in demand for medical image analysis over the past 20 years. Higher resolution and contrast ratio images of the internal organs and tissues are produced by MRI imaging. The anatomical structural details are also provided by the nuclear magnetic resonance theory. The contrast ratio between various interstitial fluids will vary depending on the variety of imaging sequences. The signals from magnetic particles spinning in the body in accordance with its magnetic tune are picked up by a powerful computer during an MRI scan, which transforms the scanned data into visual representations of the internal body parts. Medical imaging results can be modified to take advantage of the imaging techniques used in the development of the analysis of remote sensing data and to easily analyse the symptoms of patients [4].

All neutrons, electrons, and protons move in a circular motion because they all have an angular moment around their respective axes. The magnetic field that surrounds all charged objects is created by their odd mass number and circular spin. When an external force is acting on these tiny magnetic fields that are randomly aligned, the MRI machine uses these magnetic fields to generate signals. Tiny magnetic fields automatically align with stronger ones if they are contained by larger homogeneous magnetic fields. This is how MRI works. In fact, the majority of the atoms in the human body are hydrogen, and each one of these atoms contains a single proton in the nucleus. MRI machines are therefore more sensitive to hydrogen.

Together with advances in medical imaging technology, neuroscientists are becoming more interested in techniques or methodologies for identifying normal tissues in the brain in order to improve the effectiveness

*Department of CSE, Sharnbasva University, Kalaburagi, Karnataka, India (virupakshi.108@gmail.com).
†Department of CSE, Sharnbasva University, Kalaburagi, Karnataka, India.(sveerashetty@gmail.com).
‡Department of CSE, Sharnbasva University, Kalaburagi, Karnataka, India (ambika.umashetty@gmail.com).

of therapy for brain disease [5]. It is difficult to diagnose and perform surgery on the brain, which is located inside the cranium. A non-obtrusive approach to diagnosis has become the standard practice, and its use has swept the globe. A brain tumour is one of the most dangerous diseases. The primary issue in brain tumour treatment is its location and its capacity to spread rapidly. Thus, the critical steps in addressing this issue are segmenting the image and identifying them [6, 7].

Medical image segmentation is one of the most important branches of image segmentation, and it is the first and most important step in medical image analysis. The advancement of magnetic resonance imaging in 3D space allows for the imaging of tissues with high contrast and resolution. The neuroscientist observes three important parts of the brain: cerebrospinal fluid, white matter, and grey matter. Morphological structures are the fundamentals of these studies, which are used to segment medical images. For testing purposes, the technique of medical image segmentation provides automatic and semiautomatic tumour extraction methods.

2. Literature survey. Yash Sharma et al [8] proposed a novel boundary box method for brain tumor detection named as Fast Bounding Box (FBB). This method helps in locating the tumors in the brain from MRI image of patient's brain. This is used to find out most dissimilar brain region called as the Region of Interest (ROI) i.e. tumor location. The FBB make use of novel score function known as Bhattacharya's Coefficient (BC) for axes points of the bounding box in which tumor region is inscribed and is different from other regions of the image. The FBB finds out the most dissimilar portion properly but sometimes fails to find out some tumour's tissue which is easily available and it might be the part of tumor region. In order to reduce this, the region of ROI is scaled up to 20 pixels which lead to enhanced ROI. This increases the size of the bounding box by 20 pixels on each side increasing the area so that if tumor part is left, it can also be covered in this process. The steps of this work are locating the tumor region using FBB and enhancing of ROI by scaling the bounding box. The ten features considered for the ANN training and GLCM feature provides 8 features.

Tianming Zhan et al [9] made use of fully automatic method in order to obtain the segmentation of tumor from multispectral MR images of the brain. They used different intensity patches in the input image for representing the abnormal and normal tissues. Dictionary is generated for further classification of tissues. The brain tumor and the normal brain tissue is classified by using the sparse representation classification in the complete image. Finally, regularisation of Markov Random Field (MRF) is used.

Kalyani A et al [10] presents neural network based classification of MRI brain images. The work consisted of three stages which are feature extraction, stage of feature reduction and finally classification. Initially, options with respective to tomography victimized images are obtained by Discrete Wavelet Transformation (DWT). The second stage deals with reduction of victimized principles part analysis to more essential options using PCA technique. In the classification stage, the two supervised machine learning based classifiers are used, the first one is feed forward neural network and second one is the back propagation neural network.

Apurva et al [11] has concentrated on detecting and classifying the tumor types like meningiomas, gliomas, nerve sheath and pituitary adenomas from MRI images of brain. The images in training phase and testing phase are pre-processed and segmented using adaptive K-means clustering. The segmentation process is followed by feature extraction using GLCM and Gabor wavelet methods. The classifier accuracy is improved by selecting the features using PCA algorithm. Multilevel Proximal Support Vector Machine (PSVM) classifier for detecting the different tumor types automatically from MRI images of brain.

V Sheejakumari et al [12] introduced Histon based technique for segmentation of the image. Before segmentation, noise is eliminated at the preprocessing stage by applying median filter. Then noise eliminated image is passed to the process of feature extraction. The features extracted are correlation, area, covariance and mean of images. The image classification is associated with neural network at final stage. Here neural network is modified by optimizing the values of weight using optimization technique called Artificial Bee Colony (ABC).

Jokin et al [13] proposed a hybrid method for classification of brain tumor in the MRI images. They removed the Noise using the median filter, extracted the texture features, used CART and SVM for hybrid classification.

P. Kumar et al [14] includes four stages in the work. They are preprocessing of the image, segmentation, extraction of the feature, feature reduction and finally classification. The Wiener filter is applied in preprocessing to reduce noise and further to extract features. Segmentation based on modified region growing is applied. Co-occurrence matrix and Histogram methods are utilized to extract edge and texture features. The feature space

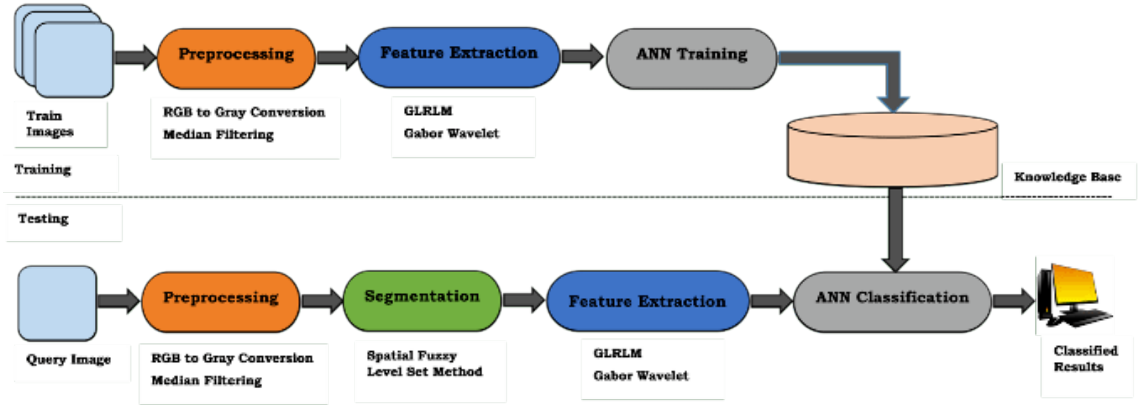


Fig. 3.1: Block Diagram of Proposed System

dimension is reduced by PCA to get efficient and better classification results. Kernel based SVM is used to classify them into Benign and Malignant. The similarity index, extra fraction and overlap fraction are measured for performance analysis.

Nithyapriya et al [15] used multifractional Brownian motion (mBm) to analyse the MRI brain images. Using mBm helps to extract spatially varying features. Depending upon multifractal features, segmentation process is carried. This is carried out using modified AdaBoost algorithm which is referred as Adaboost Support Vector Machine (SVM).

Marco Alfonse et al [16] used brain images of DICOM standard format to classify them into benign and malignant. Different phases considered are acquisition of data, preprocessing of data, Expectation Maximization and adaptive thresholding based segmentation, FFT based Features for MRI images, Minimal-Redundancy-Maximal-Relevance (MRMR) criteria based feature selection for valuable feature selection and the SVM for classification stage is used.

Deepa et al [17] used MEM approach for segmentation. The brain has different regions controlling different functions like speech, balance and movement. If any region gets damage it affects functionality of respective functions and may lose coordination of all activities. Brain tumor is nothing but tissue's abnormal growth in brain. If proper treatment to that tissue is failed, it spreads it to other tissue and results in cancer. Modified Expectation Maximization (MEM) is implemented in proposed method for segmentation and image is classified by SVM classifier. Correlation and Energy are calculated for the proposed work and are compared with existing methods.

3. Methodology. The Proposed work uses the Spatial Fuzzy Level Set Algorithm for Tumor Segmentation and GLRLM and Gabor wavelet features are used for feature extraction and ANN classifies into benign and malignant.

In pre-processing step, in order to reduce the computation complexity, input image is converted into monochrome image by using RGB to Gray conversion method and median filter is applied to eliminate noise present in the MRI image.

3.1. Median Filter. This filter is one of the non-linear filters used to eliminate the noise from the image while preserving the edges. This filter is applied by moving pixel by pixel in the image, its value is substituted by the value of median to its neighbors. The pattern of the neighbor considered is referred to as window which slides over the entire image on pixel by pixel basis. Values of window pixels are first arranged into numerical then median is calculated, this replaces the value of pixel being considered.

The median filter is similarly to mean filterwhich decreases the noise of the image. The median filter preserves the useful information in the image. This filter also comes under edge preserving smoothing filters.

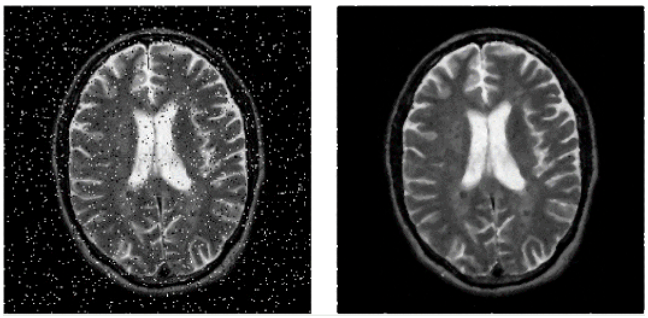


Fig. 3.2: Example for Median Filtering

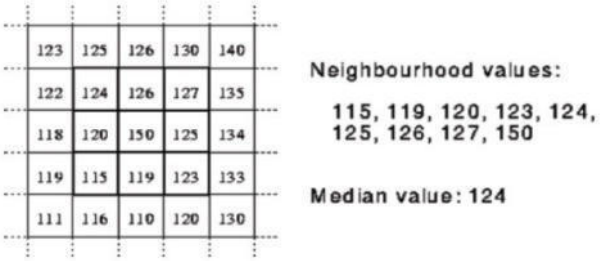


Fig. 3.3: Median value estimation for each pixel neighbourhood.

For image $A(x)$ and the image $B(x)$, the constraint of median is as in the Equation (3.1).

$$median\ \ [\ A\ (x)\ +\ B\ (x)]\ \neq\ median\ [\ A\,(x)]\ +\ median\ [\ B\ (x)] \tag{3.1}$$

The filter smoothens the data retaining the small and sharper information. The median value is nothing but middle value of the neighbor pixel values. It is necessary to notice that this is equal to average or the mean value, rather the median value has half of the neighbourhood values greater than its value and rest half lesser than median value. It is stronger middle representative than that of average. Thus it is effective in eliminating various noises from the image [18].

In the above window, the central pixel has different value from the neighbouring pixels; therefore, median value is substituted. The above window's median value is 124. The Window of size 3×3 is used in this case.

3.2. Spatial Fuzzy Level Set Algorithm. Fixed cost function is reduced by adaptive estimation of scope of subclass and centroid in the fuzzy clustering, so it is adaptive kind of thresholding. Among fuzzy clustering, Fuzzy c-means (FCM) is popularly used algorithm and is widely used in the medical field.

A membership function $\mu_{\mathrm{m,n}}$ is used in FCM to represent membership degree from m^{th} cluster to n^{th} object. it is mainly used in medical image segmentation as physiological tissues are non-homogeneous. FCM's cost function is given by Equation (3.2):

$$\mathrm{j} = \sum_{\mathrm{m}=1}^{\mathrm{K}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \mu_{\mathrm{mn}}^{\mathrm{l}} \|\mathrm{i_n} - \mathrm{v_m}\|^2 \tag{3.2}$$

The fuzziness of segmentation is controlled by the parameter $\mathrm{l} > 1$. The following are constraints must have

subjected by the membership functions as given in Equation (3.3):

$$\sum_{m=1}^{c} \mu_{mn} = 1; \quad 0 \leq \mu_{mn} \leq 1; \qquad \sum_{n=1}^{N} \mu_{mn} > 0 \tag{3.3}$$

The centroids v_m and μ_{mn} must be iteratively updated as shown in the Equation (3.4) and (3.5):

$$\mu_{mn} = \frac{\|i_n - v_m\|^{-2/(l-1)}}{\sum_{k=1}^{c} \|i_n - v_k\|^{-2/(l-1)}} \tag{3.4}$$

$$v_i = \frac{\sum_{n=1}^{N} \mu_{mn}^{l} i_n}{\sum_{n=1}^{N} \mu_{mn}^{l}} \tag{3.5}$$

Later spatial FCM was proposed by Chuang and others in which the fuzzy membership functions can be incorporated by the spatial information directly by the following Equation (3.6):

$$\mu'_{mn} = \frac{\mu_{mn}^{p} h_{mn}^{q}}{\sum_{k=1}^{C} \mu_{kn}^{p} h_{kn}^{q}} \tag{3.6}$$

The parameters which control the respective contribution are p and q. The h_{mn} will add the spatial information as follows in given by the Equation (3.7):

$$h_{mn} = \sum_{k \in N_n} \mu_{nk} \tag{3.7}$$

N_n represents the local window with image pixel n as circle. The centroid v_m and weight μ_{mn} are updated according to the Equations (3.4) and (3.5). In pixel classification based on FCM, image is segmented using Level set method which employs dynamic variational boundaries. Active contour method is also very popular technique for image segmentation. But here, level set method embeds active contours to time dependent function of PDE $f(t, x, y)$rather than the parametric characterization [19]. Then the active contour evolution can be approximated by orbiting at zero level set C(t) as in the Equation (3.8):

$$C(t) = \begin{cases} \phi(t, x, y) < 0 & (x, y) \ is \ inside \ \Gamma(t) \\ \phi(t, x, y) = 0 & (x, y) \quad is \ at \ \Gamma(t) \\ \phi(t, x, y) > 0 & (x, y) \ is \ outside \ \Gamma(t) \end{cases} \tag{3.8}$$

Series of zero contours are comprised in the implicit interface C. Therefore the image segmentation is converted into $\cup S_k \cup \Gamma = 1$.

The Equation (3.9) determines the ϕ evolution using the equation of level set.

$$\begin{cases} \frac{\partial \phi}{\partial t} + F \ |\nabla \phi| = 0 \\ \phi(0, x, y) = \phi_0 \quad (x, y) \end{cases} \tag{3.9}$$

The normal direction is denoted by $|\nabla \phi|$, initial contour is represented by $\phi_0(x, y)$ and comprehensive forces is represented by F which also includes gradient image's external force, geometrical interface's internal force and optional artificial momentum.

The F advancing force need regularization using edge representative function g in the Equation (3.10) to pause the level set evolution till optimal solution is obtained.

$$g = \frac{1}{1 + |\nabla (G_\sigma * I)|^2} \tag{3.10}$$

where $G_\sigma * I$ represents the image convolution using Gaussian smoothing kernel G_s and image gadient is represented by ∇. The function g is approximal to zero in variational boundaries, else it will be positive.

In this medical image segmentation novel fuzzy based level set method is presented where it starts with clustering technique method called spatial fuzzy, then the results are further used to start the segmentation of the image, estimation of controlling parameters are carried out and evolution of the level set is regularized. The parameter initialization and configuration is automated by this new algorithm using spatial fuzzy. The spatial restrictions are employed with FCM for interested approximated contours determination in medical images [20]. The enhanced function of level set segmentation uses FCM output for evolution directly. Considering the interested component from FCM as $R_k : \{ r_k = \mu_{nk}, \ n = x \times N_y + y \}$. Now the level set will start conveniently as Equation (3.11):

$$\phi_0 (x, y) = - 4 \varepsilon (0.5 - B_k) \tag{3.11}$$

Here ε denotes a constant to regulate Dirac function. The Equation (3.12) defines this function as

$$\delta_\varepsilon (x) = \begin{cases} 0, & |x| > \varepsilon \\ \frac{1}{2\varepsilon} \left[1 + \cos \left(\frac{\pi x}{\varepsilon} \right) \right], & |x| \leq \varepsilon \end{cases} \tag{3.12}$$

Binary image B_k is obtained from the Equation (3.13):

$$B_k = R_k \geq b_0 \tag{3.13}$$

where $b_0 (A (0, 1))$ represent adjustable threshold. Given the initial level set function ϕ_0, the length and the area can be estimated using spatial fuzzy clustering

$$l = \int \delta (\phi_0) \, dxdy;$$

$$\alpha = \int H(\phi_0) dxdy$$

$H(\phi_0)$ is Heaviside function is defined according to Equation (3.14):

$$H (\phi_0) = \begin{cases} 1, & \phi_0 \geq 0 \\ 0, & \phi_0 < 0 \end{cases} \tag{3.14}$$

This segmentation algorithm considers the membership degree for every pixel of the image μ_k as the length for particular interested component R_k. The dynamic interface is pushed or pulled adaptively by the enhanced balloon force towards the interested object

$$G (R_k) = 1 - 2R_k \tag{3.15}$$

The result of balloon force $G (R_k) \ (\in [-1, 1])$ denotes a matrix having variable for pushing and pulling force to every pixel of the image. The interested object is attracted by the level set function irrespective to the initial position. As shown in the Equation (3.16), the evolution equation can be transformed into

$$\xi (g, \phi) = \lambda \delta (\phi) \ \text{div} \left(g \frac{\nabla \phi}{|\nabla \phi|} \right) + g \ G (R_k) \ \delta (\phi) \tag{3.16}$$

3.3. Feature Extraction. The features used in this work are grey level run length matrix and Gabor Features.

3.3.1. Gray-level Run-length Matrix (GLRLM). The features for the texture analysis of the brain can be collected using GLRLM. The texture is analyzed by the pattern of grey value at certain angle in the direction of the reference pixels. The count of adjacent pixels having the similar grey value at certain direction is referred as Run length [21].

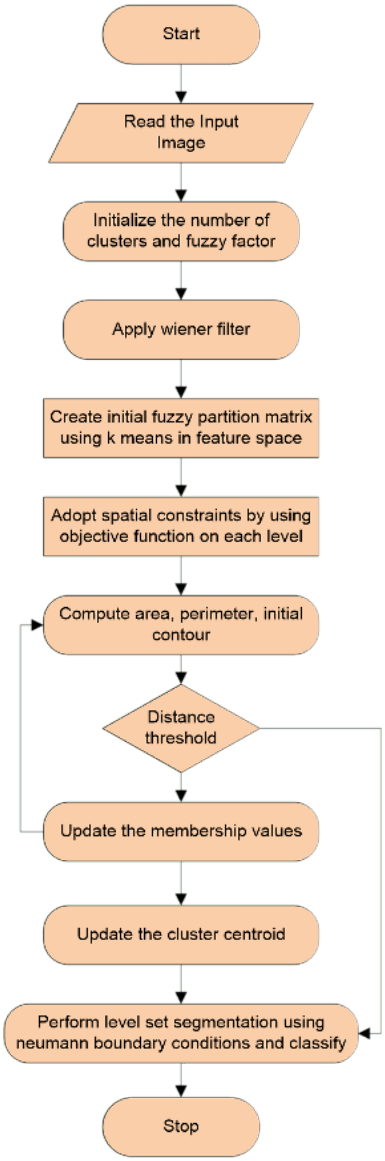


Fig. 3.4: Flow Chart of Fuzzy C Means With Spatial Constraints

GLRLM is a two dimensional matrix where each value is represented as $p(i, j|\theta)$ which gives the count of the elements j having the gray intensity i, in direction θ. Considering an example, figure 3.6 gives a 4×4 matrix of the pixel image having 4 different gray levels. The matrix given in figure 3.7 represents GLRLM with the direction of angle 0° $[p(i, j|\theta = 0^\circ)]$.

The GLRL matrix can be constructed in the following orientations 0^o, 45^o, 90^o as well as 135^o. This is depicted below.

The texture features extracted from GLRLM [22] are

1. Shot Runs Emphasis(SRE):

$$SRE = \frac{1}{n} \sum_{i,j} \frac{p\ (\ i, j\ )}{j^2} \tag{3.17}$$

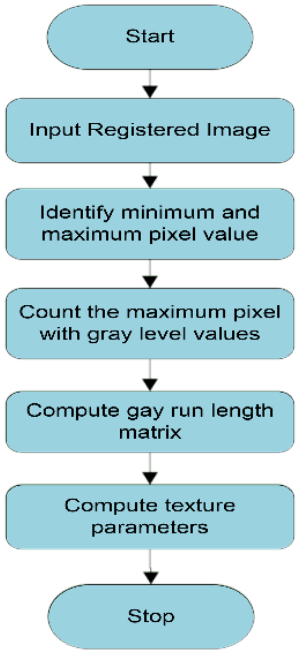


Fig. 3.5: Flowchart for GLRLM Feature Extraction

1	2	3	4
1	3	4	4
3	2	2	2
4	1	4	1

Fig. 3.6: Matrix of Image 4×4 pixels

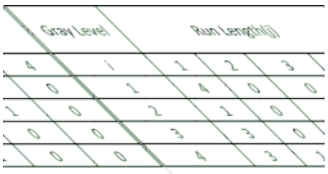


Fig. 3.7: GLRL Matrix

2. Long Runs Emphasis(LRE):

$$LRE = \frac{1}{n} \sum_{i,j} j^2 \;\; p \;\; (\, i, j) \tag{3.18}$$

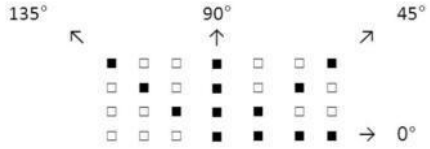


Fig. 3.8: Run Direction

3. Gray Level Non-uniformity(GLN):

$$GLN = \frac{1}{n}\sum_{i}\ (\sum_{j}\ p\ (\ x, y\ ))2 \tag{3.19}$$

4. Run Length Non-uniformity(RLN):

$$RLN = \ \frac{1}{n_r}\sum_{j=1}^{N} p_r\ (i)^2 \tag{3.20}$$

5. Run Percentage(RP):

$$RP\ = \ \sum_{i,j}\frac{n}{p\ \ (i,j)\, j} \tag{3.21}$$

6. Low Grey Level Run Emphasis(LGRE)

$$LGRE = \ \frac{1}{n}\sum_{i,j}\frac{p\ \ (i,j)}{i^2} \tag{3.22}$$

7. High Gray Level Run Emphasis(HGRE)

$$HGRE = \ \frac{1}{n}\sum_{i,j} i^2\ p\ \ (i,j) \tag{3.23}$$

3.3.2. Gabor Feature. The 2-D Gabor filter is represented by following Equation (3.24) and Equation (3.25):

$$h\,(x, y; f, \theta) = \ \frac{1}{\sqrt{\pi\sigma_1\sigma_2}}\exp\left(-\frac{1}{2}\left(\frac{R_1^2}{\sigma_1^2}+\frac{R_2^2}{\sigma_2^2}\right)\right)\ .\exp\left(i\left(f_x x + f_y y\right)\right) \tag{3.24}$$

$$H\,(u, v; f, \theta) = \ \ 2\sqrt{\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\left(\sigma_1^2(S_1 - f)^2 + \sigma_2^2 S_2^2\right)\right) \tag{3.25}$$

here $R_1 =$ x cos θ +ysin θ , $R_2 =$ $-$x sin θ +ycos θ , $\sigma_1 = \frac{c_1}{f}$, $\sigma_2 = \frac{c_2}{f}$, $f_x =$ fcos θ , $f_y =$ fsin θ , $S_1 =$ u cos θ + v sin θ , $S_2 =$ $-$u sin θ + v cos θ , c_1 and c_2 are constants.

The coefficient $\sqrt{\pi\sigma_1\sigma_2}$ guarantees that different Gabor filter in the family have their energy equal to 1 i.e., $\|h\|^2 = \int\int hh^* dxdy = 1$.

The spatial sinusoids which are localized using Gaussian window are referred by Gabor filters and band pass filters, which are frequency sensitive and orientation. The x and y terms used in Eq. (3.24) are digital image pixel co-ordinates. The 2-D Gaussian standard envelope are represented by σ_1 and $\sigma_2$parameters. The f indicates the pass band central frequency and µ is the spatial orientation [23].

Considering $\sigma_1 = \sigma_2 = \sigma$, Equation (3.24) and Equation (3.25) can be simplified as Equation (3.26) and Equation (3.27):

$$\mathrm{h}\ (\mathrm{x},\mathrm{y};\mathrm{f},\theta) = \frac{1}{\sqrt{\pi}\sigma}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right).\exp\ (i\ (f_x x + f_y y)) \tag{3.26}$$

$$H\ (u,v;f,\theta) = 2\sqrt{\pi}\sigma\exp\left(-\frac{\sigma^2\left((u-f_x)^2+(v-f_y)^2\right)}{2}\right) \tag{3.27}$$

The Gabor filter is complex type with two components: real and imaginary

— The real part is $\frac{1}{\sqrt{\pi\sigma_1\sigma_2}}\exp\left(-\frac{1}{2}\left(\frac{\mathrm{R}_1^2}{\sigma_1^2}+\frac{\mathrm{R}_2^2}{\sigma_2^2}\right)\right)\cos(\mathrm{fR}_1)$

— The imaginary is $\frac{1}{\sqrt{\pi\sigma_1\sigma_2}}\exp\left(-\frac{1}{2}\left(\frac{\mathrm{R}_1^2}{\sigma_1^2}+\frac{\mathrm{R}_2^2}{\sigma_2^2}\right)\right)\sin(\mathrm{fR}_2)$

The family of Gabor filters takes both phase and amplitude of the frequency spectrum [24]. Gabor features are extracted by convolution of image I with all the Gabor filters family at each and every pixel (x, y) and is given by the Equation (3.28):

$$G\ (\ x,y,f_k,\theta_m)\ =\ \sum_{x'}\sum_{y'} I\ \left(\ x-x',y-y'\right)\ h\ (\ x',y';f_k,\theta_m\ ) \tag{3.28}$$

where $I(x, y)$ represents the intensity of the pixel.

The Gabor features used [24] are:

1. The phase feature of $\mathrm{G}\,(\mathrm{f_k},\theta_\mathrm{m})$ can be represented by edge location and some other information of the image I as depicted by the Equation (3.29).

$$\mathrm{F}_1\,(\mathrm{x},\mathrm{y};\mathrm{f_k},\theta_\mathrm{m}) = \ \mathrm{phase}\,(\mathrm{G}\,(\mathrm{x},\mathrm{y};\mathrm{f_k},\theta_\mathrm{m})) \tag{3.29}$$

2. The amplitude feature of $\mathrm{G}\,(\mathrm{f_k},\theta_\mathrm{m})$is represented as oriented frequency spectrum of image I in every location as in Equation (3.30)

$$\mathrm{F}_2\,(\mathrm{x},\mathrm{y};\mathrm{f_k},\theta_\mathrm{m}) = |\mathrm{G}\,(\mathrm{x},\mathrm{y};\mathrm{f_k},\theta_\mathrm{m})| \tag{3.30}$$

3. The sum of squares of various frequency responses having same orientation can be considered as feature and it also indicates the local energy of particular orientation given by Equation (3.31)

$$\mathrm{F}_3\,(\mathrm{x},\mathrm{y};\theta_\mathrm{m}) = \mathrm{F}_2\,(\mathrm{x},\mathrm{y};\mathrm{f}_0,\theta_\mathrm{m})^2+\frac{1}{4}\mathrm{F}_2\left(\left\lfloor\frac{\mathrm{x}}{2}\right\rfloor,\left\lfloor\frac{\mathrm{y}}{2}\right\rfloor;\mathrm{f}_1,\theta_\mathrm{m}\right)^2+\frac{1}{16}\mathrm{F}_2\left(\left\lfloor\frac{\mathrm{x}}{4}\right\rfloor,\left\lfloor\frac{\mathrm{y}}{4}\right\rfloor;\mathrm{f}_2,\theta_\mathrm{m}\right)^2 \tag{3.31}$$

In Eq. (3.31) $\mathrm{F}_2\,(\mathrm{f}_0,\theta_\mathrm{m})$ is a $\frac{\mathrm{X}}{2}\times\frac{\mathrm{Y}}{2}$ image, and $\mathrm{F}_2\,(\mathrm{f}_1,\theta_\mathrm{m})$ is a $\frac{\mathrm{x}}{4}\times\frac{\mathrm{y}}{4}$ image, and $\mathrm{F}_2\,(\mathrm{f}_2,\theta_\mathrm{m})$ is a $\frac{\mathrm{x}}{8}\times\frac{\mathrm{y}}{8}$ image. The $\mathrm{F}_3\,(\theta_\mathrm{m})$ is a $\frac{\mathrm{X}}{2}\times\frac{\mathrm{Y}}{2}$ image, thus each pixel of $\mathrm{F}_2\,(\mathrm{f}_1,\theta_\mathrm{m})$ is further partitioned to 4 pixels and each pixels of $\mathrm{F}_2\,(\mathrm{f}_2,\theta_\mathrm{m})$ is partitioned equally into 16 pixels.

4. The orientation can also be considered as feature which is maximum local energy.

$$\mathrm{F}_4\,(\mathrm{x},\mathrm{y}) = \mathrm{k} \tag{3.32}$$

Where $\mathrm{F}_3\,(\mathrm{x},\mathrm{y};\theta_\mathrm{k}) = \max_{\mathrm{m}=0\sim2}\{\mathrm{F}_3\,(\mathrm{x},\mathrm{y};\theta_\mathrm{m})\}$

Algorithm's Steps:

step-1: Select the value of F.

step-2: Convert the image to 2D matrix

step-3: Using the formula given below the impulse response is calculated

$$G\ (\ x,y\ )\ =\ g\ (\ x,y\ )\times\ \exp\{2\pi i f\ (\ x\cos\theta\ +y\sin\theta\ )\ \}$$

g is given as follows

$$g\ (\ x,y\ )\ =\ \left(\frac{1}{2\pi S^2}\right)\times\exp\{-\ \left(x^2+y^2\right)/\ 2S^2\}$$

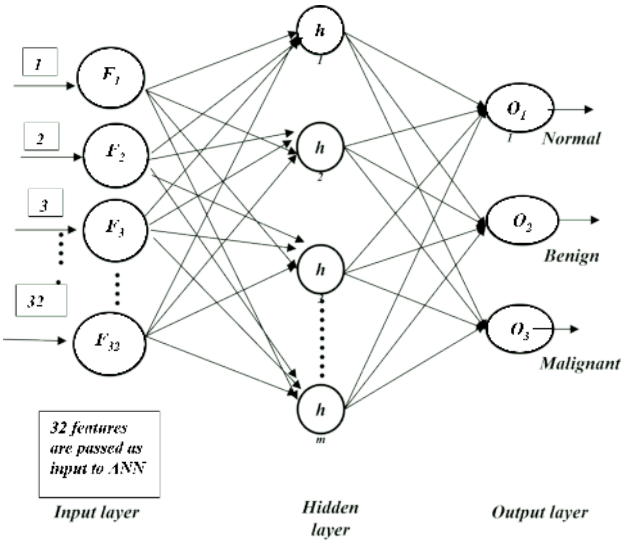


Fig. 3.9: Neural Network Structure

step-4: the image is passed to the filter having the impulse response as given in the step 3, filtered image is represented by $r(x, y)$.

step-5: The filtered image's energy is computed as $E = r2(x; y)$

step-6: From filtered image estimate the segmented image using.
$\varphi(x, y) = E \bigotimes m(x, y)$ mask window is denoted by $m(x, y)$. Using standard deviation Ss window size is estimated. we adopt Ss = S / Sigma. So,

$$m\ (\ x, y\ )\ =\ \left(\frac{1}{8\pi\sigma^2}\right) * \exp\left\{-\left(\frac{n12\ +\ n22}{4\sigma^2}\right)\right\}$$

step-7: Print $\phi(x, y)$, i.e. output image.

3.4. Artificial Neural Networks (ANN). A neural network (ANN) is made up of a large number of highly interconnected processing elements known as neurons. The ANN's operations are divided into two phases: training and testing. The ANN is trained with specific input patterns during the training phase. When an instructed input pattern is detected at the input during the testing phase, an output is generated.

Figure 3.9 shows architecture of proposed three output Feed Forward Back Propagation Neural Network (FF-BPNN). The nodes F_1, F_2…F_{32} represent the corresponding input features and the nodes O_1, O_2 and O_3 represent the output corresponding to the brain tumor class. A FF-BPNN uses a gradient descent learning algorithm, in which the network weights are moved along the negative of the gradient of the performance function.

The term backpropagation refers to the manner in which the gradient is computed for non-linear multilayer networks.

The proposed network consists of input, hidden and output layer. The input layer is represented by 32 nodes, which have been derived from feature extraction models namely; GLRCM and Gabor techniques. These features are normalized into the values in between 0 and 1. The outputs of each layer are the inputs to the next layer. Ten neurons are in the hidden layer and one neuron in the output layer. Initially the weight and bias are initialized to the values in between -1 and 1. The learning rate is assumed to be 1. The network is trained for the different training data set. The weights and bias are modified during learning process. Finally, the network becomes ready to classify the new data sample. The back propagation algorithm is given below.

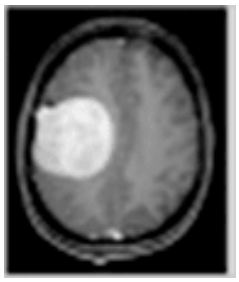


Fig. 4.1: Input Image

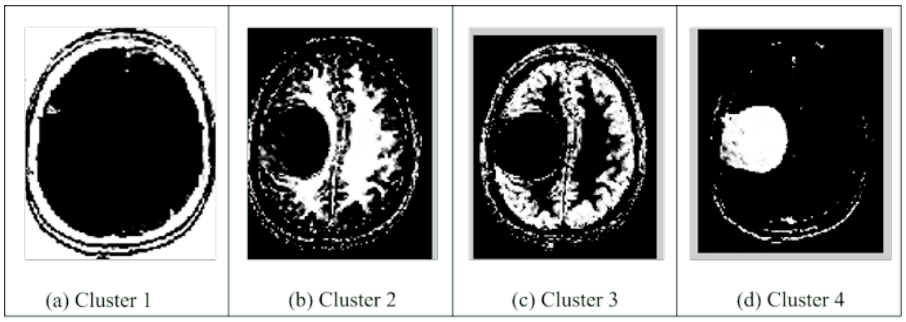


Fig. 4.2: Clusters of Input Image given in Figure 4.1.

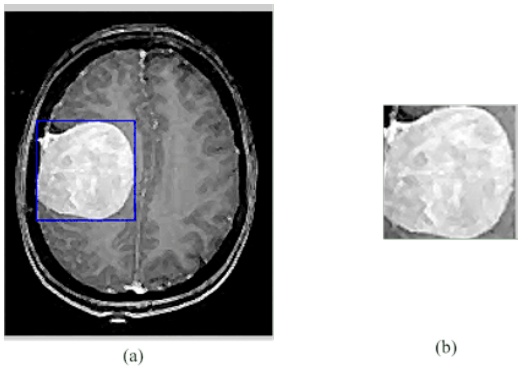


Fig. 4.3: (a) Tumor Detection & (b) Segmentation of Tumor Region

4. Results and discussion. The data collection in medical image processing [34, 35, 36, 37] plays an important role, So MRI images used in this work were collected from the local hospital. From the collected images, brain tumor is segmented by the method called spatial fuzzy level set segmentation, Gabor features and GLRLM features were extracted from the segmented images.

Considering the input image shown in the above figure, the result of the spatial fuzzy level set segmentation is shown in the figure 4.2.

The tumor region is segmented by using the cluster 4 among the entire four clusters shown in figure 4.3.

The GLRLM and Gabor features are extracted from the segmented image and classified using ANN. The classified result for the above image belongs to the class benign.

Similarly, to the above sample image, the result for other samples is shown in figure 4.4.

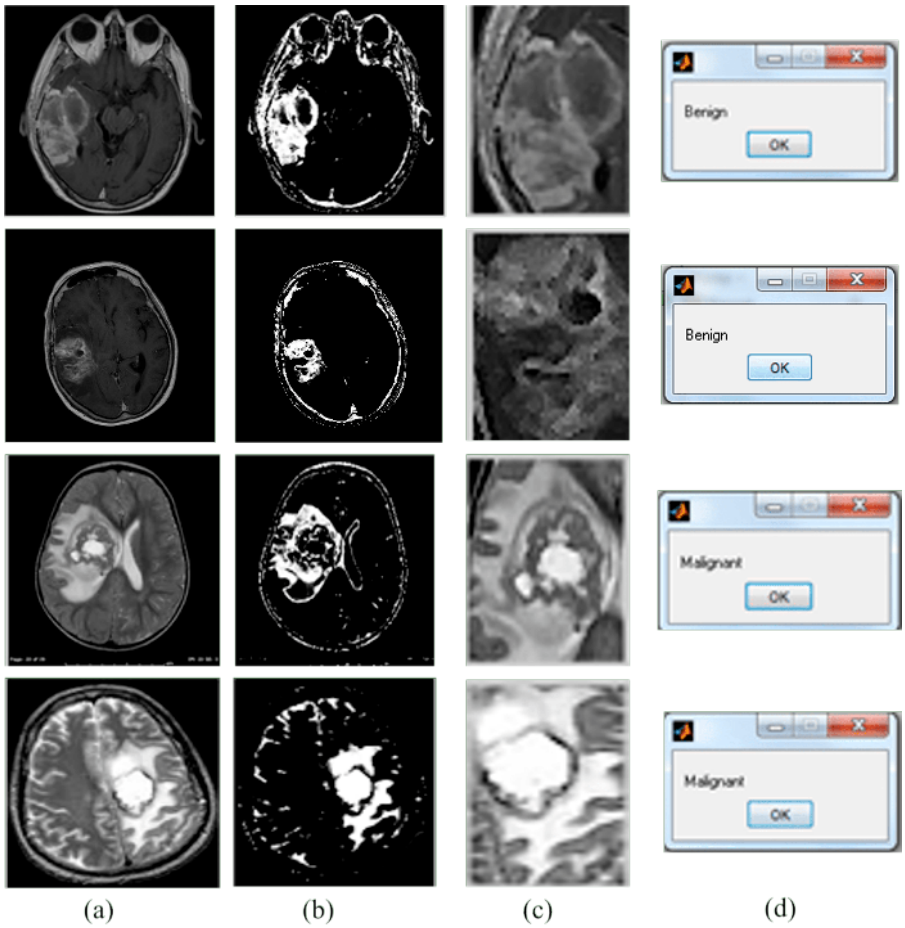


Fig. 4.4: (a) Original Image & (b) Selected Cluster & (c) Segmented Image & (d) Classified Result

Table 4.1: Table of Comparison of Accuracy for Segmentation

Sl.No	**Authors**	**Methods**	**Accuracy**
1	Ruchita A et. al [27]	*FCM*	92.67
2	Prof P.Senthil [26]	Level Set Segment	89.99
3	Jason J. Corso Et al [28]	Multilevel Bayesian segmentation	88
4	A R Kavitha and C Chellamuthu [29]	Adaptive Region growing based	46
5	Proposed	Spatial Fuzzy Level Set Segmentation	94

The Comparative Analysis for the Accuracy of Fuzzy based Level Set method with existing segmentation methods are carried out and results are tabulated in Table 4.1.

The Graph consisting of various segmentation methods is plotted with respect to the segmentation accuracy in Figure 4.5.

The bar chart in the Figure 4.5 shows the comparative analysis of five segmentation methods for Accuracy performance parameter. The obtained Accuracy values are 92.67, 89.99, 88, 46, 94 for FCM, Level Set Segment, Multilevel Bayesian segmentation, Adaptive Region growing based and Spatial Fuzzy Level Set Segmentation respectively. From the graphs, the Spatial Fuzzy Level Set Segmentation shows better Accuracy results as

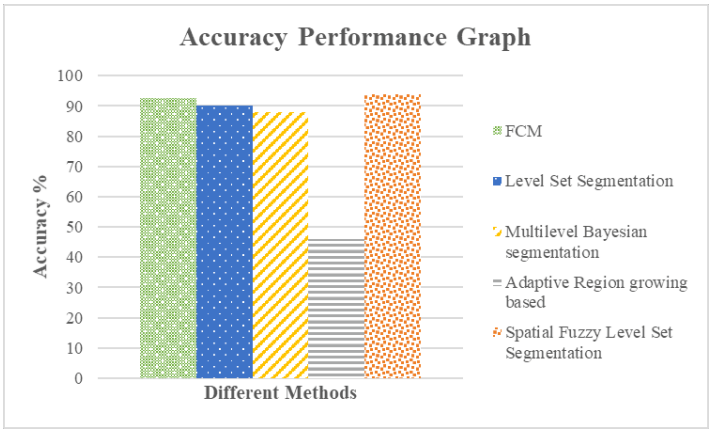


Fig. 4.5: Performance graph of Segmentation methods in terms of Accuracy

Table 4.2: Table for evaluating Sensitivity and Specificity of various Segmentation techniques.

Sl.No	Authors	Methods	Sensitivity	Specificity
1	Sheejakumari V, Gomathi [30]	HGANNA	0.83	1.00
2	Sheejakumari V, Gomathi [31]	IPSONN	0.96	0.93
3	Sheejakumari V, Gomathi [32]	Histon based segmentation and modified neural network	1.00	0.6
4	P.Senthil [26]	Level Set method	0.74	0.90
5	Uttamjeet Kaur [33]	Ant Colony Algorithm	0.94	1.00
6	Proposed work	*Spatial Fuzzy Clustering*	0.95	0.74

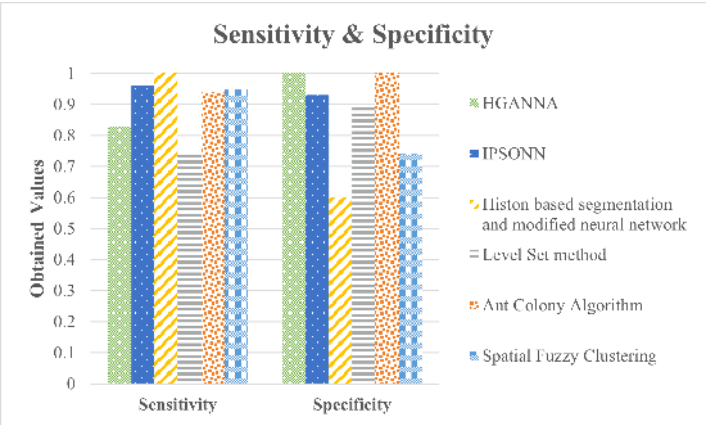


Fig. 4.6: Sensitivity and Specificity for Segmentation

compared with other segmentation methods.

The Table 4.2 gives comparison of Sensitivity and specificity for the various segmentation techniques.

Figure 4.6 shows a graph depicting the comparative analysis of various segmentation methods with respect to sensitivity and specificity for segmentation.

Comparative analysis of various classification methods is carried out and the results are tabulated in Ta-

Table 4.3: Comparison Table for existing and Proposed System

Sl.No	Authors	Methods	Accuracy	Sensitivity	Specificity
1.	Jokin Arul Raj and Sathees Kumar [13]	Textural Feature extraction, PSO based Feature Selection, CART and SVM	92.31	96.67	77.78
2.	P. Kumar and B. Vijayakumar [14]	Texture features extraction by Co-occurrence Matrix and Histogram then PCA and RBF Kernel Based Support Vector Machine	0.94	0.95	0.9
3	Proposed	Spatial Fuzzy Level Set, GLRLM, Gabor, ANN	0.94	0.75	0.95

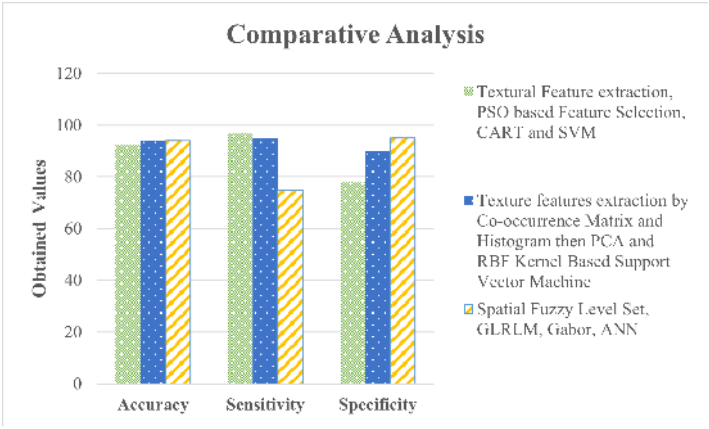


Fig. 4.7: Performance Graph for Existing and Proposed Methods

ble 4.3.

Finally, bar chart consists of all the parameters such as Accuracy, Sensitivity and Specificity versus three classification systems have been portrayed in Figure 4.7. From the above figure it shows that Proposed computer-aided diagnosis (CAD) system based on Spatial Fuzzy level set with ANN classifier performs well than the other CAD systems as compared in terms of Accuracy and Specificity.

5. Conclusion. The spatial fuzzy technique used in this paper will give more homogeneous regions compared to the other techniques which will eventually help in reducing blur and noisy spots. Level set segmentation is flexible in detecting the edges and has more control on movement of curves. Gabor features are of higher flexibility and represent the images more efficiently by surpassing the directional and Gaussian derivative performances. The higher order statistical texture can be extracted from GLRLM method. By combining these methods, the proposed work gives better results for segmentation and classification with overall classification accuracy of 94%. The accuracy of the proposed method can be improved further by extensive training. In future various state of the art optimization techniques can be used along with ANN classifier in order to increase the accuracy of classification. This CAD system is aimed for the beneficial of radiologist for localization and diagnosis of the brain tumors in the MRI images.

Acknowledgement. The authors would like to thank our mentor Late Dr. Basavaraj Amarapur, former HOD, Electrical & Electronics Engineering Department, PDA college of engineering Kalaburagi for their continuous guidance and support. Also authors would like to thank Dr. Nagendra Patil, Radiologist in Patil

Diagnostic Centre Kalaburagi, for providing MRI images and validating the obtained results.

REFERENCES

[1] Poongodi, M., Sharma, A., Hamdi, M. et al., *Smart healthcare in smart cities: wireless patient monitoring system using IoT*,J Supercomput., 2021.
[2] Yongfu Shao, Jue Wu, Hongping Ou, Min Pei, Li Liu, Ali Akbar Movassagh, Ashutosh Sharma, Gaurav Dhiman, Mehdi Gheisari, Alia Asheralieva, *Optimization of Ultrasound Information Imaging Algorithm in Cardiovascular Disease Based on Image Enhancement*, Mathematical Problems in Engineering, 2021, vol. 2021, Article ID 5580630, 13 pages.
[3] Dhiman, G., Vinoth Kumar, V., Kaur, A. et al, *DON: Deep Learning and Optimization-Based Framework for Detection of Novel Coronavirus Disease Using X-ray Images*, Interdiscip Sci Comput Life Sci, 2021, 13, pp. 260–272.
[4] Erlin L, Meng W, Jianfeng T, Jianjian L, *Automatic Segmentation of Brain Tumor Magnetic Resonance Imaging Based On Multi-Constrains and Dynamic Prior*, Int J Smart Sensing Intell Syst, 2015, 8, pp. 1031-1049.
[5] Virupakshappa AB, *An approach of using spatial fuzzy and level set method for brain tumor segmentation*, International Journal of Tomography & Simulation, 2018, 31 (4).
[6] Uplaonkar DS, Patil N, *Modified Otsu thresholding based level set and local directional ternary pattern technique for liver tumor segmentation*, International Journal of System Assurance Engineering and Management, 2022 Jan 13: pp. 1-1.
[7] Sameena Banu, Ambika, *Segmentation of Brain Tumor Lesion Using the Mumford-Shah Model*, International Journal of Advanced Scientific and Technical Research, 2014, 3 (4), pp. 415-423
[8] Sharma Y, Chhabra M. , *An Improved Automatic Brain Tumor Detection System*,Int J Adv Res Comp Sc Soft Eng., 2015, 5, pp. 11–15.
[9] Zhan T, Gu S, Feng C, Zhan Y, Wang J., *Brain Tumor Segmentation from Multispectral MRIs Using Sparse Representation Classification and Markov Random Field Regularization*, IJIGSP., 2015; 8, pp. 229-238.
[10] Bhawar KA, Bhil NK., *Brain Tumor Classification Using Neural Network Based Methods*, IJESRT., 2016; 5, pp. 721-727.
[11] Apurva YN, Nanda S., *MRI Brain Tumor Segmentation and Classification based on Multilevel PSVM Classifier*, IRJET., 2016; 3.
[12] Sheejakumari V, Gomathi S., *Brain tumor detection from MRI images using Histon based segmentation and modified neural network*, Biomed Res., 2016; S1-S9.
[13] Raj JA, Kumar S. *An Enhanced Classifier for Brain Tumor Classification*, IJCTA., 2016; 9, pp. 325-333.
[14] Kumar P, Vijayakumar B., *Brain Tumour MRI Segmentation and Classification Using by PCA and RBF Kernel Based Support Vector Machine*, Middle-East J Sci., 2015; 23, pp. 2106-2116.
[15] Nithyapriya G, Sasikumar C., *Detection and Segmentation of Brain Tumors using AdaBoost SVM.*, IJIRCCE 2014; 2.
[16] Alfonse M, Badeeh A., *An Automatic Classification of Brain Tumors through MRI Using Support Vector Machine*, Egyptian Computer Science Journal., 2016; 40.
[17] Deepa T, Muthalagu R, Chitra K., *MEM Based Brain Image Segmentation and Classification Using SVM*, JEAS., 2015; 10.
[18] Nag M, Shanker NV., *Image De-Noising By Using Median Filter and Weiner Filter*, IJIRCCE., 2014; 2, pp. 5641-5649.
[19] Li BN, *Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation*, Comput Biol Med., 2011; 41, pp. 1-10.
[20] Uma Devi N, Poongodi R., *Integration of Spatial Fuzzy Clustering with Level Set for Efficient Image Segmentation*, IJCSCN, 2015; 3, pp. 296-301.
[21] Tang X., *Texture Information in Run-Length Matrices*, IEEE Trans Image Process., 1998; 7.
[22] Wibawanto H., *Discriminating Cystic and Non Cystic Mass using GLCM and GLRLM-based Texture Features*, IJEER., 2010; 2, pp. 569-580.
[23] Zheng D., *Features Extraction Using Gabor Filter Family*, International conference on signal and image processing., 2004; pp. 139-144.
[24] Soni K, Kuma U., *A New Gabor Wavelet Transform Feature Extraction Technique for Ear Biometric Recognition*, IEEE Conference on Power, 2014.
[25] Gurleen K, Harpreet K., *Efficient Facial Recognition Using PCA-LDA Combination Feature Extraction with ANN Classification*, Int J Adv Res Comp Sci Soft Eng., 2016; 6, pp. 258-263.
[26] Senthil P., *Image Mining Base Level Set Segmentation Stages To Provide An Accurate Brain Tumor Detection*, IJECSC., 2016; 6, pp. 8294-8299.
[27] Banchpalliwar RA, Salankar SS., *A Review on Brain MRI Image Segmentation Clustering*, IOSR-JECE., 2016; 11, pp. 80-84.
[28] Corso JJ, Sharon E, Dube S, El-Saden, S, Sinha U, Yuille A., *Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification*, IEEE Trans Med Imaging., 2011.
[29] Kavitha AR, *Chellamuthu C. Detection of Brain Tumour from MRI Image using Modified Region Growing and Neural Network*, Imaging Sci J., 2012.
[30] Sheejakumari V, Gomathi S., *Healthy and pathological tissues classification in MRI Brain Images using Hybrid Genetic Algorithm-Neural Network (HGANNA) Approach*, EJSR., 2012; 87, pp. 212-226.
[31] Sheejakumari V, Gomathi S., *MRI Brain Images healthy and pathological tissue classification with aid of Improved Particle Swarm Optimization and Neural Network (IPSONN)*, Comput Math Methods., 2015; pp. 1-12.
[32] Sheejakumari V, Gomathi S., *Brain tumor detection from MRI images using histon based segmentation and modified neural*

network, Biomed Res., 2016.

[33] Kaur U, Sharma R, Dosanjh M., *Cancers Tumor Detection using Magnetic Resonance Imaging with Ant Colony Algorithm*, IJAREEIE, 2016; 4, pp. 58-62.

[34] Veerashetty, S., Virupakshappa & Ambika , *Face recognition with illumination, scale and rotation invariance using multiblock LTP-GLCM descriptor and adaptive ANN*, Int J Syst Assur Eng Manag, 2022.

[35] V.D. Ambeth Kumar, *Medical data mining process using efficient clustering and classification approaches*, International Journal of Innovations in Scientific and Engineering Research, 2020, Vol 7, Issue 8, pp. 106-114.

[36] Virupakshappa, Amarapur B, *An improved segmentation approach using level set method with dynamic thresholding for tumor detection in MRI images*, HELIX. 2017 Jan 1; Vol.7, Issue 5, pp. 2059-66.

[37] T. Jemima Jebaseeli and C. Anand Deva Durai, *A robust ecg signal processing and classification methodology for the diagnosis of cardiac health*, International Journal of Innovations in Scientific and Engineering Research, 2020, Vol 7, Isssue 8, pp.115-123.

Edited by: Vinoth Kumar
Received: Jun 27, 2022
Accepted: Nov 3, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 251–262, DOI 10.12694/scpe.v23i4.2026

A NOVEL CSINR TECHNIQUE FOR ACCURATE AND PRECISE GPS COMMUNICATION BY GEOGRAPHICAL CENTRIC SELF-LEARNING NODES

N. SWAROOP KUMAR *, K.S. RAMESH†, A. MAHESWARY‡, AND R. REVATHI§

Abstract. Since its introduction, the Global Positioning System (GPS) is finding many countless, useful, and emergency applications, focused mainly on track. As the technology is advancing day by day and the best feature of GPS, which does not, relies on mobile signal to work, making it a feasible feature to incorporate into other devices as functionalities. By adopting GPS to a system, accurate mapping and geographical labeling can be obtained. GPS works better with the coordination of nodes and requires centralized monitoring and a reporting system. As it is a known fact that a demerit follows merit anywhere else, In GPS also, the major attention is required to make the nodes to success in mapping the intermediate space between agent node used for reporting and the remaining nodes of a cluster, where satellite and node coordination can be possible integer ambiguity technique. Many researchers have proposed solutions to the aforementioned problem; unfortunately still today the proposed methods are weaker in achieving lesser time delay of Total Electron Content (TEC). The proposed Centric Self-Learning Interconnected Nodes Reading (CSINR) technique is novel in terms addressing the intermediate nodes failing to label the inter-connected object spaces between reporting agents and nodes using integer ambiguity technique for node co-ordination and using a dedicated GPS prediction-based clock system, which predicts precise and accurate mapping between interconnected nodes. Based on the information shared between among the network managers a separate pseudo-connected network will be formed and further this network will be considered an interconnected nodes network. From the information calculated from temporal factors and clock offset the separate pseudo network is extracted by using the proposed CSINR technique. Add-on self-improvement is introduced to the proposed method by a self–learning feature to an individual join extract the principal rate of partisan neighbouring join to sustain accuracy in order consistent basis. An evaluation ratio of 97.43%, sensitivity of node occurrence is resulted as 92.78% and accuracy of 97.43% & 97.12% is achieved among a cluster of 32 & 64 nodes respectively.

Key words: Mapping of GPS Nodes, Position Accuracy, Precision of GPS Nodes, Self-Learning nodes, co-relation.

AMS subject classifications. 68T05

1. Introduction.

Peer-peer communication among nodes is a fundamental and primary requirement for a stable network that works for multi-dimensional attributes and functions. It is inferred that the distance between the satellites and receiver is calculated by considering the product of the time delay (τ) and the velocity of light (c). A net of 24 satellites positioned into an track forms a GPS navigation system and it is finding its profound role in achieving network stability and thereby making track able. At the time of its introduction, it was intended for only military applications, but later on, it spread its wings to other civilian applications. GPS would have not been a successful technological development applied for a reliable communication environment for multiple applications, services, and techniques until unless without node coordination between each and every node in the network. Under a dynamic communication system, GPS provides paperless transportation logs and records [1], by means of continuously tracking the movement and monitoring of nodes via a given. GPS works by rotating its satellites around the earth double a day and transmits signal in sequence precisely to the earth. Then Receivers on earth determine the user's literal site by using triangulation, compare time delay between a signals transmit by a satellite through the time it was established, and estimate distance of a specific satellite. Further with the distance difference between the remaining satellites in the orbit, the receiver be able

*Research Scholar, Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. (swaroop.proff@gmail.com)

†Professor, Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. (ks.ramesh1712@gmail.com).

‡Associate Professor, Department of ECE, Sri Venkateswara College of Engineering andTechnology, Chittoor, India (maheswaryswaroop@gmail.com)

§Associate Professor, Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. (revathimouni@gmail.com)

to settle on the user's spot and put on show from its 2D and 3D spot (longitude, latitude and altitude) and mark it on the user's electronic map. This information is further used to know the exact speed, trip distance, total distance between transmitter and receiving node, day and night time duration calculation. As the GPS communication depends on the nodes, nodes should be formed as a cluster, and any one of the node will be selected as a cluster node for monitoring the trustiness of every node in the cluster and the number of nodes in the is fixed in number and it is fixed. Clusters can be made as per the area of authorization. Clusters can be making any number. But if it will be fixed number then accuracy will be better. More number of clusters are there, then accuracy will be good.

Along with the advantages GPS has profound challenges in terms of achieving exact and accurate location of GPS coordinates. The author of [2] has presented an elaborated survey report of multiple techniques involving fundamental computations defined to achieve precision and accuracy of GPS coordinates. Further, the difficulties have been addressed in [3] with Smartphone tracking-based evaluation. Dealing with the smartphone based GPS node coordination is much more complex, as it involves mobile nodes, and dynamic nodes, which may leave or enter the network at any time. Under such scenarios, GPS is vulnerable to polling of crowd node, controlling of conjunction of GPS node difficulty, and issues with management nodes. Assurance of overall coordination among GPS nodes can be easily achieved by mapping to a particular antenna. The above-mentioned issues were addressed in the proposed methodology by means of a systematic approach and evaluated in terms of reliable solutions and frameworks for accuracy prediction of GPS nodes. In the following section various existing research solutions were elaborated and in III section the detailed methodology of planned technique is introduced. In IV section Statistical model and design and proof of the findings were included further results and discussion and comparison with the existing methodologies along with future scope are included in Section-V.

2. Literature Survey. As it is known fact that for a reliable tracking and monitoring of source node movement and location is accurately achieved by the Global Positioning System (GPS) based on geometric coordinates. Unfortunately, GPS System is more vulnerable to man in middle attacks, black-hole attack, Spoofing and unauthenticated node tracking issues. Author [4] have proposed a reliable framework solution in terms of inactive coordinates of GPS and the false nodes coordination log information of duplicate nodes of the network. The proposed method also deals with such type of fake nodes and corrects the long time travel behavior of unexpected node of the network. The mentioned methodology provides support in terms of node traceability and verifies authenticity of accessing. Handling of large volume of data, generated by the process is addressed in [5] proposed an optimization technique adopted for data collection in big-data management environment, thereby achieving accurate mapping of GPS nodes.

A more detailed method used to process and analysis the information related to mapping and reliability is proposed [6] which is used to generate a secure GPS location accessing. Author named as the Standard Positioning System (SPS), provides a generalized representation of overall coordinates associated for communication and hence a- centralized monitoring is un assessed, to overcome the challenge a dedicated monitoring and positioning system is required for evaluation. A customization and processing approach for improved behavioural mapping in GPS based outset is evaluated. These customizations are predominant and have influence of coordinate recommendations for modern technological assistance. The geometric parameters of information collection via GPS coordinates are evaluated under a timely manner based paradigms to assure the dependency of time and coordinates under the operations.

The overall coordination required for essential communication and hence a centralized monitoring is unassessed. Further the security threats of central monitoring system and the following challenges of a dedicated monitoring and positioning system were addressed in [7] and a solution for GPS based outset is proposed in terms of a customized behavioral mapping among the nodes. Modern technological assistance can be achieved by a network formed with such a customized and predominant coordinates in the network. The author [8] have proposed an interlocked data transfer solution for handling information breeching in multi-node environment named as in TelMED [8].

Authors of [9] have proposed timely mannered paradigms to collect the information related to the geometric parameters via GPS coordinates. The proposed method works with the assumption of dependency of time on its coordinators operation. It is a known fact that the information transmission via wireless network is

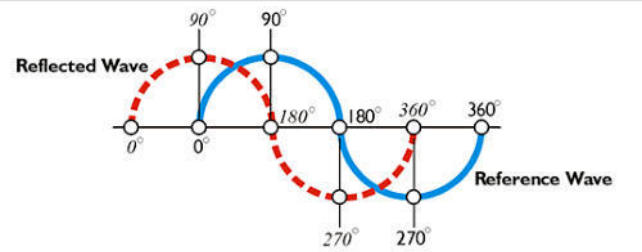


Fig. 3.1: Integer Ambiguity Problem Illustration

more vulnerable to security attacks and a solution is proposed to procure the dependency of data in wireless network. A draft of all over analysis report of dependency and growth rate of GPS nodes is presented in [10] and also covered the current time challenges of GPS processing followed with the road map of development towards the near future. An agile and low cost GPS tracking method is proposed in [11] and resulted a more reliable customized system for allotting permission for incoming request and also accessing of locations. A self-learning system with the aim of accurate and reliable polling and mapping of GPS coordinate for a secured communication. A detailed report of ICT along with its role in customized network following with its challenges was presented in [12]. Authors have proposed a methodical idea for valuing the situation of GPS coordinates system mapping.

3. CSINR: Centric Self-Learning Interconnected Nodes Reading technique. As we know that in GPS,the measurement of the difference in the phase of the incoming signal and the phase of the internal oscillator in the receiver reveals the small distance at the end of a range.

By comparing the phase of the signal returned from the reflector with the reference wave it kept at home, an EDM can measure how much the two are out of phase with one another. However, this measurement can only be used to calculate a small part of the overall distance. It only discloses the length of a fractional part of a wavelength used. This leaves a big unknown, namely the number of full wavelengths of the EDM's modulated carrier between the transmitter and the receiver at the instant of the measurement. This cycle ambiguity is symbolized by N. Fortunately; the cycle ambiguity can be solved in the EDM measurement process. The key is using carriers with progressively longer wavelengths. For example, the sub meter portion of the overall distance can be resolved using a carrier with the wavelength of a meter. This can be followed by a carrier with a wavelength of 10 meters, which provides the basis for resolving the meter aspect of a measured distance. This procedure may be followed by the resolution of the tens of meters using a wavelength of 100 meters. The hundreds of meters can then be resolved with a wavelength of 1000 meters, and so on as shown in fig.3.1.

Here is that comparison, the reference wave in blue with the reflected wave with the dashed red line. The reflected wave came back out of phase by a quarter of a wavelength. With an EDM, wavelengths of varying length can be sent out. For example, if the EDM sends out a wavelength of 100 meters, then by looking at the fractional part of that 100-meter wavelength, it would be possible to determine the tens of meters in the distance. The hundreds of meters of the overall distance could be resolved by sending out a wavelength of 1,000 meters and looking at the fractional part (by phase comparison). This method depends on the fact that the EDM survey can send out waves of different wavelengths and have them return to where they came from. That makes comparison of the returned wave with the reference wave possible. By comparing phase angles, the fractional part of the wavelength that went out can be determined. The components of the total distance are built up by sending on wavelengths of different size; first the thousands of feet, then the hundreds of feet, then the single feet, and finally the decimals of feet. However, this whole method is not possible in GPS surveying. There are only three carriers; L1, L2, and L5. They have constant wavelengths. Therefore, while it's possible to determine the fractional part of the wavelength, that one small component of the distance, from a single measurement, knowing the number of full wavelengths between the receiver and satellite is more difficult. This is known as the integer ambiguity problem.

Another important parameter to consider in the GPS communication is Total Electron Content (TEC),

which is the time delay of Total Electron Content Total electron content (TEC) is an important descriptive quantity for the ionosphere of the Earth. TEC is the total number of electrons integrated between two points, along a tube of one meter squared cross section, given in equ(3.1) i.e., the electron columnar number density. It is often reported in multiples of the so-called TEC unit, defined as $TECU = 1016el/m2 \approx 1.66 \times 10 - 8m$. The TEC is path-dependent. By definition, it can be calculated by integrating along the path ds through the ionosphere with the location-dependent electron density ne(s):

$$TEC = \int n_e(s)ds \tag{3.1}$$

TEC is significant in determining the scintillation and group and phase delays of a radio wave through a medium. Ionosphere TEC is characterized by observing carrier phase delays of received radio signals transmitted from satellites located above the ionosphere, often using Global Positioning System satellites. TEC is strongly affected by solar activity.

The CSINR technique aims at improving the reliability in intermediate node space which is helpful in improving the accuracy of GPS nodes mapping. The process of GPS coordinate system definition and information gathering starts with defining Node Attribute Extraction (NAE), which captures the information of active node in terms of its initial parameters through the antenna. This information technically includes, the location of users *(U_n)* and a stream of nearst locations will be formed.

A matrix which extracts the inter-related node primary to NAE is framed with many attributers and association ratio information of the network. As shown in Fig.3.3 a novel clustering method is proposed and named as the Static User Grouping (SUG), which provides an accurate customizable and reliable solution for node allocation and defining the space relationship to improve the network performance further. This is an internally evaluated node base invention of co-relationship matrix. Another improvement of self proposed method is withdrawal of inter-reliant medium of GPS coordinates is proposed, evaluated by categorizing nodes separately from the activity of each and every node in the coordinate system.

The novelty of the proposed technique is a dedicated GPS clock prediction is proposed to predict precision and mapping among the interconnected nodes. The interconnected nodes are dependent on the information shared via network managers and hence, CSINR technique extracts the nodal relationship of interconnected nodes to build a pseudo connected network based on temporal factors and clock offset. The Fig.3.4 shows the above said node extraction and categorization combined with the proposed method for point to point node interaction and for the selection of cluster node. Further, a cluster head node will performs the inner auditing of nodes standards *(N_{01} to N_{0n})* in every categorized prevailing conditions *(Y_i)*, which can be further customized. The collected data from the cluster node is further processed a matrix called summation matrix defined for ensuring the synchronization between each and every node of the network. A threshold value is fixed from the information calculated from the timepiece pre-set and equalize coordinates of GPS site denoted as a f_o. Information related to node existing from the network with respect to reporting agent can be easily obtained by a static user grouping technique as shown in Fig. 3.2. CSINR Spatial Representation method is used to yield the worth of the template coordinates as in Fig.3.4, and it is a work of fiction method as a region tagging where R_X is the inner zone dependency value and Z_X is the outer zone dependency value. The nodes of the proposed method are self-learning and extracts extract the principle value of supporting neighbouring nodes to maintain reliable source accuracy information. The proposed method generates the co-relationship matrix that is important because Based on the co-relationship matrix we may know the nodes which are fixed and which are movable. With this matrix we can validate the nodes. With this matrix we can find the suitable path to communicate the data.

The proposed technique is improvised by adopting filtering method to overcome the Integer Ambiguity problem. In filtering approach, independent measurements are averaged to find the estimated position with the lowest noise level. A third use a search through the range of possible integer ambiguity combinations from which it calculates the one with the lowest residuals. These approaches can't assess the correctness of the particular answer, but they can provide the probability with certain conditions, that the answer is within given limits.

A clear understanding of CINR technique under each node of GPS tracking is clearly presented by the 3D model of spatial representation as shown in Fig. 3.5. The process under node extraction, by the customizing

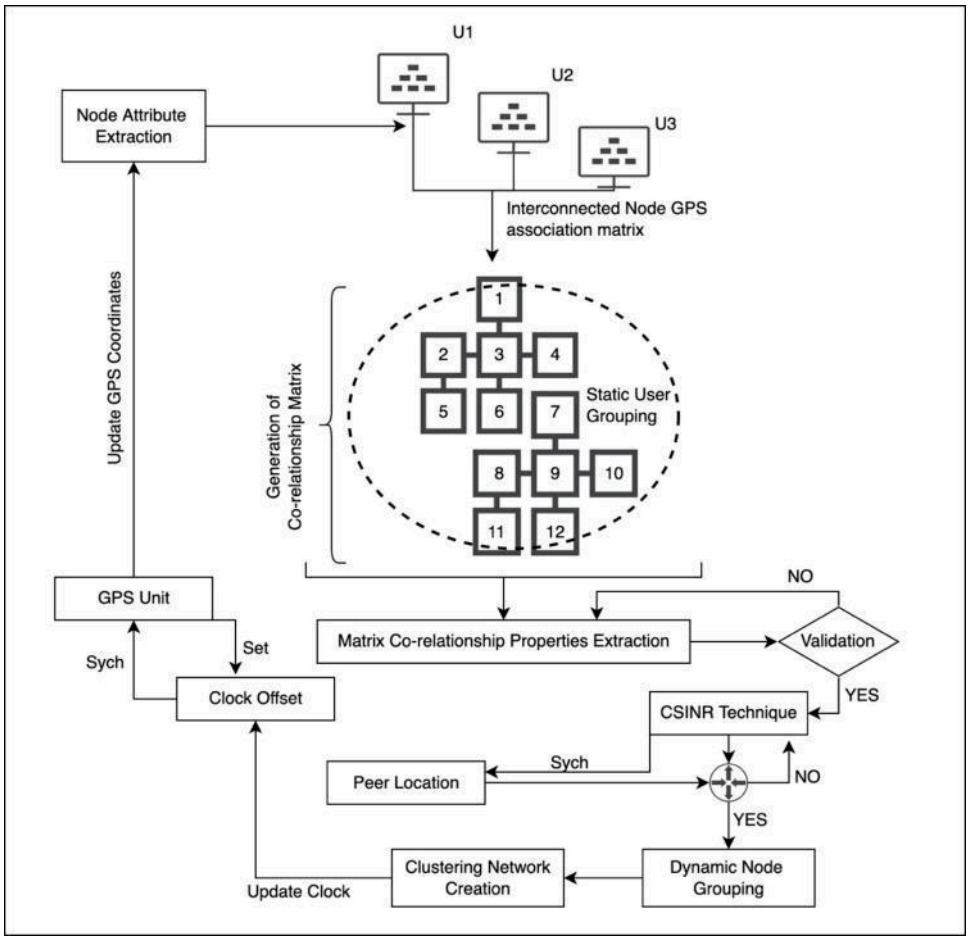


Fig. 3.2: Block representation of CSINR technique

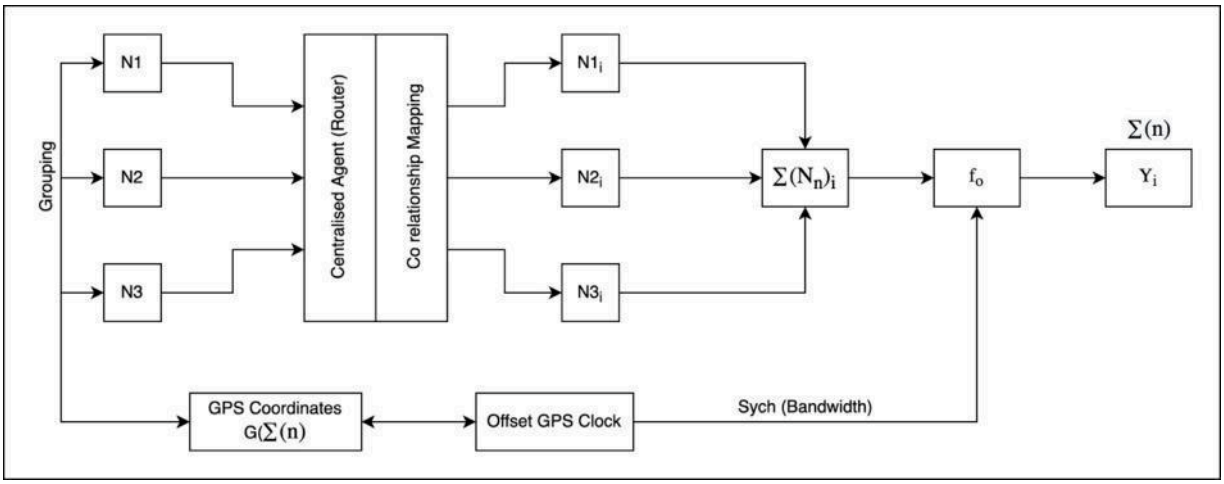


Fig. 3.3: Static Clustering CSINR Technique

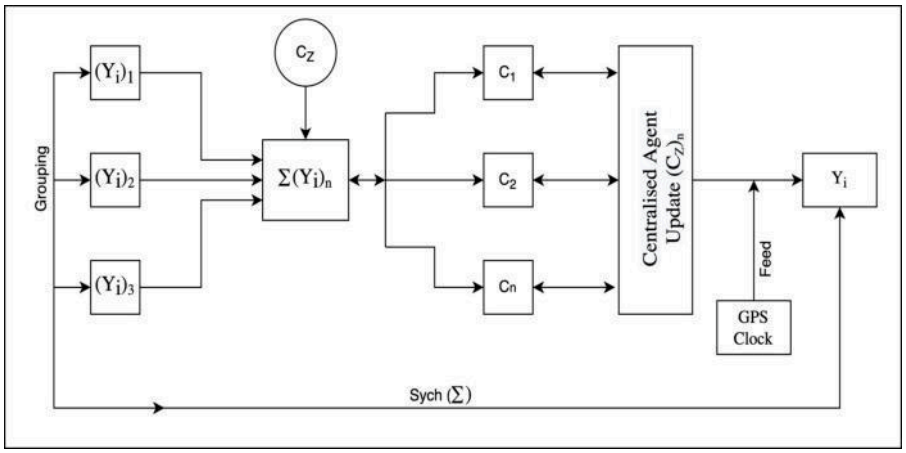


Fig. 3.4: Dynamic Clustering CSINR Technique

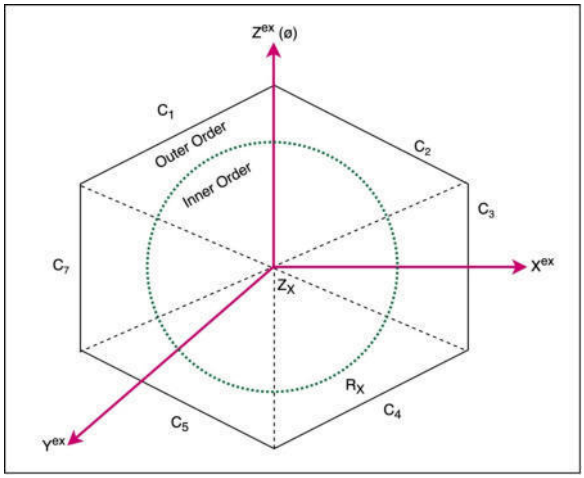


Fig. 3.5: Cluster Representation

the inner zone and outer zone categorized as static nodes and the outer zone is as lively nodes. As values exceeding the doorstep and hence extract parameter also restricted so there by projection observation curve will also be maximized. The nodes information related to entry and exit from the network and the information of node capturing and categorization will taken care by the cluster head.

4. Statistical Approach. Another challenging task of GPS is formation of a hexagonal cell as the GPS coordination is completely dependent on node interconnection. Let $G = \{G_1, G_2, G_3 ... G_n\}$ is the GPS factor of a device and is authenticated to monitoring server$\{D_n\}$such that $\{G_i \Rightarrow D_n\}$under an order of$[r, s, t]^i$; Receiver (*'r'*), Sender(*'s'*) and bit of transmit value('t'). Normally, it shows below

$$\{D_n\} = (P_i + G_i) \Rightarrow \left(\mathrm{P}_i + [r, s, t]^{G_i}\right) \tag{4.1}$$

$$D_S = \left(\sqrt{\left(\sum_{i=0}^{\infty} D_{Si}\right)}\right) - \Delta G \tag{4.2}$$

Where $\Delta G$is an aligned and optimized value of a GPS system with $\{D_S\}$and arriving GPS standards (G*)* as $\{G \subseteq \Delta G\}$by means of position to $\{D_S \Rightarrow\} L(\Delta G)$, here *'L'* means the place of ΔG and habit. Precisely, the sort of sites (L) by a underneath issue of $[r,s,t]^i$is represented as below.

$$r = \sqrt{(r-r_i)^2 + (r-r_{i+1})^2 + (r-r_{i+2})^2 + ...} \tag{4.3}$$

$$S = \frac{1}{\sqrt{r\,(r-r_i)^2}} \overrightarrow{\Delta G} \tag{4.4}$$

$$t = \sum_{i=0}^{\infty} \left(r_i \rightarrow \left(\sum_{j=0}^{n} \Delta L_i \frac{\Delta G_j}{\partial t} \right) \right) \tag{4.5}$$

The value of each parameter information is communal in the provide $[r,s,t]^i$sector is evaluated and factorized and GPS Coordinates is evaluated by considering the incoming factor *(I_f).* Equation 4.6 represents the room harmonization of every sign rate is represented.

$$[I_f] = \begin{vmatrix} \cos\theta^{\dagger}\sin\varphi & \left(\sin\theta^{\dagger}\cos\varphi + \sin\varphi\cos\theta^{\dagger}\right) \\ -\left(\cos\theta^{\dagger}\sin\varphi\right) & \left(-\sin\theta^{\dagger}\cos\varphi - \sin\varphi\cos\theta^{\dagger}\right) \\ \sin\varphi & -\sin\theta + \cos\varphi \end{vmatrix} \tag{4.6}$$

From the event occurrence and alignment of GPS coordinates the cyclic spaces are evaluated and filtered. The factorial value $[r,s,t]^i$of all joint by orientation to bandwidth din provincial values is represented as (U_{acc}) and $(U_{inc})$to arranging again an reliable join below the provided procedure room.

$$(U_{acc}) = \lim_{n\to\infty} \left(\int_{i=0}^{n} \frac{\delta\,(I_f)_i}{\delta t} \right) - \Delta G_i [r,s,t]^i \tag{4.7}$$

$$(U_{inc}) = \lim_{n\to\infty} \left(\frac{\Delta G\left[\delta\,(I_f)_i\right]}{\delta t} - \sum_{j=0}^{n} \left(\frac{\delta\,(\Delta G_i)_j}{\delta t} \oplus \Delta G \right) \right) \tag{4.8}$$

$$U = U_{acc} \oplus U_{inc} \tag{4.9}$$

U_{acc} and U_{inc} are inserted in U_i .

$$\delta U\,(t) = \frac{\delta\,(U_{acc})}{\delta t} \oplus \frac{\delta\,(U_{inc})}{\delta t} \tag{4.10}$$

$$\delta U\,(t) = \sum_{i=0}^{\infty} \sum_{j=1}^{n} \left\{ \frac{\delta\,(U_{acc})_i}{\delta t} \oplus \frac{\delta\,(U_{inc})_j}{\delta t} \right\} \tag{4.11}$$

$$\delta U\,(t) = \sum_{i=0}^{\infty} \sum_{j=1}^{n} \left\{ \frac{\delta\,(U_{acc})_i \oplus \delta\,(U_{inc})_j}{\delta t} \right\} \otimes G_i \tag{4.12}$$

$$\overrightarrow{\delta U\,(t)} = \lim_{n\to\infty} \left(\frac{\Delta G_i [r,s,t]^i}{\Delta \mathrm{T}\,(0.248)} \Delta G_i \left[\frac{\delta\,(Ut)}{\delta t} \right]_0^{\infty} \right) \tag{4.13}$$

$$\overrightarrow{U(t)} = \left(\frac{\Delta G_i [r,s,t]^i}{\Delta \mathrm{T}\,(0.248)} \left\{ \lim_{n\to\infty} \left(\sum_i^{\infty} \sum_j^{n} \left(\frac{\delta U(t)_i}{\delta t} \oplus \frac{\delta \overrightarrow{U(t)_j}}{\delta t} \right) \right) \right\} \right) \tag{4.14}$$

$$\overrightarrow{U(t)} = \left(\frac{\Delta G_i [r,s,t]^i}{\Delta \mathrm{T}} \left\{ \lim_{n\to\infty} \left(\sum_{i,j}^{n} \left(\frac{\delta U(t)_i \oplus \delta \overrightarrow{U(t)_j}}{\delta t} \right) \Delta G_i \right) \right\} \right) \tag{4.15}$$

$\delta U_i \Rightarrow \delta U(t)$ are primary ideals of GPS coordinators of utilize (U) underneath a agreed atmosphere ofΔG. The same are additional implemented by a balanced calculation on or after old user labels $U_{acc}$and next to him jointU_{inc}. All joints are consistent and mold of next-door GPS truth is calculated and processed. It show in Eq. (4.9) and will be shown as Eq. (4.16):

$$e(t) \Rightarrow \delta \left(eU_{acc} \oplus eU_{inc} \right)^{\mathrm{T}} \tag{4.16}$$

Where, $eU_{acc}$and eU_{inc} are the ecological changeable that are linked to the join or consumer room (T) through $(\forall \mathrm{T} \subseteq \Delta G)$and $(\Delta G \in eU)$as all (U) customers are allied among near join values in the defined environment of operation.

The essential situation of consumer action $\left(\overrightarrow{U}\right)$beneath the process rule of$\left(\delta U \Rightarrow \delta \left(\overrightarrow{U(t)}\right)\right)$ the self knowledge gaining method re give the track and breathing space of action from side to side which the accepted consumer is below genuine process. Think the acquiring knowledge by own label (S_P) beneath the power of background *(e)* uneven as $\left(S_P \in e^{\mathrm{T}}\right)$here '$T$' is maximum location given with in specified area. The joint (n) is collaborated as $S_P = [(n_i)_1, (n_i)_2, (n_i)_3 \ldots (n_i)_n]$such that $\left(\forall S(n_i) \Rightarrow \forall e^{\mathrm{T}}\right)$with a underneath variable of self-governing characteristic.

Based on Eq. 4.15 & 4.16, the basics of consumers $(\delta(u))$site ascription is equated with an independent factor values (V_f)such that each of $(V_f) \Rightarrow \forall e^{\mathrm{T}}$in user $U_{acc}$and $U_{inc}$is coordinated. The medium of customer's contains with own learning is shown in Eq. 4.17 .

$$S_P = \begin{cases} \left(\frac{\delta(U_{acc} \oplus U_{inc})}{\delta t}\right) \otimes \Delta G_i \\ \left(\frac{\delta(U_{acc} \oplus U_{inc})}{\delta t}\right) \otimes \left[(V_f)_i\right]_0^n \end{cases} \tag{4.17}$$

$$S_P = \begin{cases} \left(\frac{\delta(U_{acc} \oplus U_{inc})}{\delta t}\right) \otimes [r,s,t]^{\mathrm{T}} \Delta G_i \\ \left(\frac{\delta(U_{acc} \oplus U_{inc})}{\delta t}\right) \otimes \sum_{i=0}^{n} \left[(V_f)_i\right] \Delta G_i \end{cases} \tag{4.18}$$

The reliance reason of several standards attributes in consumer $U_{acc}$and $U_{inc}$is reliant on $[r,s,t]^{\mathrm{T}}$aspect, the starting arrangement sort to remove and build up a factorial terms to $[\Delta G_i]$where$[\forall G \subseteq \Delta G_i \cup \Delta L_i]$ gives all stations covered as pieces and needy to consumers site coordinates. Believe place $[\Delta L]$getaway represents in Eq.4.19 .

$$[L] = \left(\frac{\delta(U_{acc} \oplus U_{inc})}{\delta t} \right) \cong \left(\frac{\delta(r)^{e\mathrm{T}}}{\delta t} - 0.728\Delta \mathrm{T} \right) \tag{4.19}$$

$$[\Delta L] = \left(\frac{\delta(U_{acc} \oplus U_{inc})}{\delta t} \right) \cong \left(\sum_{i=0}^{n} \frac{\delta(r)^{e\mathrm{T}}}{\delta t} - 0.728\Delta \mathrm{T} \right) \tag{4.20}$$

Equation (4.20) is a learning paradigm of location and orientation of a node. The proposed methodology achieves ode reliability as user to consumer communication parallel as given by Eq. (4.19) & Eq. (4.20) in that order. The GPS nodes are monitored and recorded from the orientation mapping information of GPS neighboring nodes. Recording can also includes a simulated environment also named as node_name for easier identification and location extraction. From the line of scope *(Ios)* information the node attribute, occurrence and, precision of node mapping can be easily evaluated.

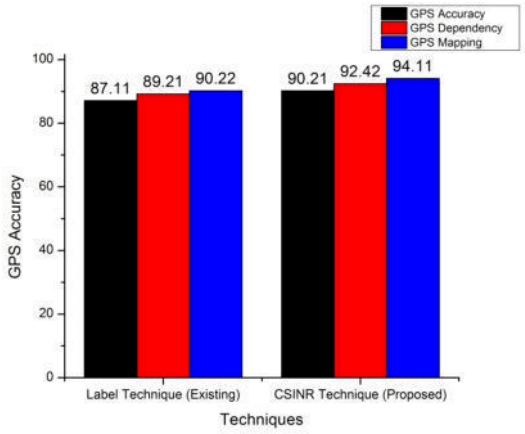


Fig. 5.1: Comparison of CSINR with existing methods

Table 5.1: Comparative validation of proposed

	Plot method	**Label procedure**	**Suggestion method**	**CSINR method**
GPS potency (hz)	2.872	2.8933	2.922	2.9421
label standards (%)	87.22	88.22	92.34	94.93
GPS standards (arbb)	81.88	92.00	90.32	97.64
Nominal space (m)	3.30	4.211	4.23	4.90
Changeable space (m)	0.46	0.46	0.49	0.91
Outcrop arc (m)	3.69	3.89	3.93	3.93

5. Results and Discussions. The reason why to prefer the proposed method over the traditional geometric method is because thisis the strategy requires a significant amount of satellite motion to succeed, and, therefore, takes time to converge on a solution. It works pretty well, but requires satellite motion and takes time to converge.

By proposing a clustering method a higher order user grouping can be achieved from the information of in progress site of Global position system coordinates as explained the aforementioned sections. The collected in turn are further processed and evaluated and updated to the central monitoring reporting agent. Further, comparative analyses of all existing methods are presented in Fig. 5.1 and it is clear that the proposed (CSINR) technique is showing better performance in terms of accuracy over others. Proposed technique is also differing from the existing methods from the join way of showing and site baggage.

The agenda of information breeching through multi-agent information sharing systems has to be optimized and generate a system interlocked approach for data transfer is proposed for secure and reliable data communication using inter-domain information systems. The fundamental concept behind this approach is to secure the data dependency under open channel communication line.

Table 5.1 summarizes the existing systems with the proposed system in terms of GPS potency, Labeled standards, GPS standard, nominal distance, and Variable distance and Project Curve parameters. From the table.5.1, it is inferred that the proposed methodology is showing better performance over the existing methodologies. The proposed methodology is showing improved projection curve over the other techniques and a percentage of 5 over the Plotting technique and 2 over the Labeled Technique and further it is used for an accurate and reliable prediction of GPS coordinates. Table 5.2 shows a comparison between the proposed and existing systems by means of TEC for a cluster of 5, 9, 17, 33, 65 nodes respectively and delay comparison graph is shown in Fig. 5.3 and with respect to Signal to Noise Ratio is shown in Fig. 5.4 respectively. The

Table 5.2: Total Electron Content v/s CSINR

Nodes	TEC Delay (ms)	CSINR Delay (ms)	TEC (SNR)	CSINR(SNR)	TEC (Accuracy %)	CSINR (Accuracy %)
5	23.69	21.52	44.22	43.20	88.36	93.22
9	33.93	32.89	48.22	47.20	89.29	93.95
17	42.10	39.19	49.33	47.30	90.33	94.05
33	47.00	43.10	53.33	48.34	90.92	94.76
65	50.34	48.20	55.84	49.44	91.45	97.33

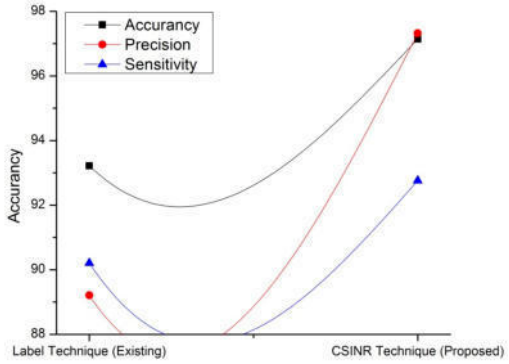


Fig. 5.2: CSINR with old methods

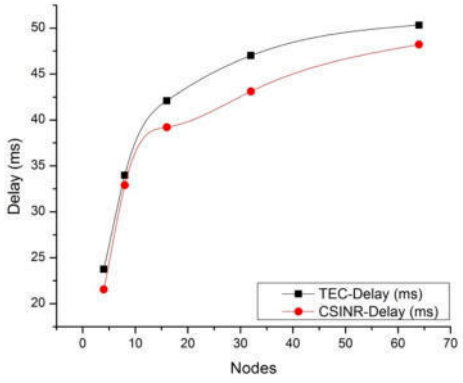


Fig. 5.3: CSINR comparison with TEC in terms of Time Delay

proposed method provided an improved accuracy of 97.32% for 64 bits as Fig. 5.5.

6. Conclusion. By adopting a strong complex and secure encryption technique along with proposed technique achieves secure data transmission under multiple cyber physical attacks such as black-hole monitoring, spoofing and unauthorized coordinates tracking between the nodes can be easily achieved.

A committed and a good number steadfast methodology was proposed accurate finding and attaching of GPS coordinates. The experimental setup of proposed method includes MATLAB2016 has required and expected instrument package. Results show that proposed modus operandi completes an improved accuracy

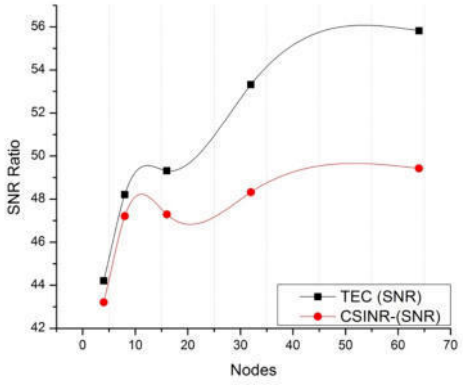


Fig. 5.4: CSINR comparison with TEC in terms of SNR

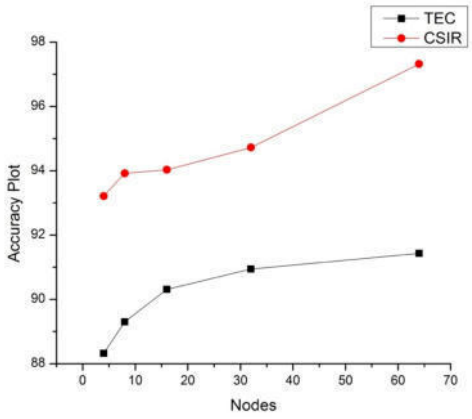


Fig. 5.5: CSINR comparison with TEC in terms of Accuracy

with 97.32% over the Total Electron Content (TEC). Proposed technique can be employed for location extraction under attribute based environment and also for inter-connected co-relationship mapping. A reliable communication is possible by developing a imitation matrix hexagonal system provides consistent announcement environment. The technique of self-learning provides the data dependency and precision mapping for GPS coordination. The CSINR based method is providing an accuracy of 97.32% and 96.82% in inter-domain GPS coordinate system. The proposed method can be used in future for lively consumer tagging and alliance.

REFERENCES

[1] Shen, Li, and Peter R. Stopher., *Review of GPS travel survey and GPS data-processing methods*, Transport reviews, vol. 34, no. 3, 2014, pp. 316-334.

[2] Berber, Mustafa, Aydin Ustun, and Mevlut Yetkin., *Comparison of accuracy of GPS techniques.*, Measurement, vol. 45, no. 7, 2012, pp.1742-1746.

[3] Merry, Krista, and Pete Bettinger. , *Smartphone GPS accuracy study in an urban environment.*, PloS one Vol.14, no. 7, 2019, pp.e0219890.

[4] Marra, Alessio D., Henrik Becker, Kay W. Axhausen, and Francesco Corman., *Developing a passive GPS tracking system to study long-term travel behavior.*, Transportation research part C: emerging technologies, vol.104, 2019, pp.348-

368.

[5] Liu, Haitao, Yew-Soon Ong, Xiaobo Shen, and Jianfei Cai. , *When Gaussian process meets big data: A review of scalable GPs.*, IEEE transactions on neural networks and learning systems, vol.31, no. 11, 2020, pp.4405-4423.

[6] Renfro, Brent A., Miquela Stein, Nicholas Boeker, and Audric Terry. , *An analysis of global positioning system (GPS) standard positioning service (SPS) performance for 2017*, 2018.

[7] Hegarty, Christopher J. , *The global positioning system (GPS).*, In Springer Handbook of Global Navigation Satellite Systems, 2017, pp. 197-218.

[8] Ahmed, Syed Thouheed, Maheshwari Sandhya, and Sharmila Sankar. , *TelMED: dynamic user clustering resource allocation technique for MooM datasets under optimizing telemedicine network.*, Wireless Personal Communications, 2020, pp.1-17.

[9] Lombardi, Michael. , *An Evaluation of Dependencies of Critical Infrastructure Timing Systems on the Global Positioning System (GPS).*, NIST Technical Note 1289, 2021, pp. 57.

[10] Renfro, B., Jessica Rosenquest, Audric Terry, and Nicholas Boeker. , *An analysis of global positioning system (GPS) standard positioning system (SPS) performance for 2015*, Space and Geophysics Laboratory. Applied Research Laboratories, The University of Texas at Austin, 2017.

[11] Knight, C. W., D. W. Bailey, and D. Faulkner. , *Low-cost global positioning system tracking collars for use on cattle.*, Rangeland ecology & management, vol.71, no. 4, 2018, pp.506-508.

[12] J. Lekha., *Secured layer cluster key management in MANET adopting virtual grid routing*, International Journal of Innovations in Scientific and Engineering Research, vol.7, no. 11, 2020, pp. 206-216.

Edited by: Vinoth Kumar
Received: Jun 28, 2022
Accepted: Nov 24, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 263–274, DOI 10.12694/scpe.v23i4.2029

A NEW IMPROVED BINARY CONVOLUTIONAL MODEL FOR CLASSIFICATION OF IMAGES

PUTTA HEMALATHA*, SHANKAR G† AND DEEPAK RAJ D.M‡

Abstract. There are numerous image classification strategies are developed in deep learning. However, due to the complexity of images, conventional image classification strategies have been incapable to meet real application needs. As the amount of pixel information rises, the classification becomes more difficult. However, CNN is widely used method for object identification in picture due to its simple and accurate, but still, it remains hazy which strategies are most supportive for analysing and distinguishing the objects in pictures. In this paper we introduced a CNN network and clustering-based technique called IBCNN to perform classification based on patch extraction. The proposed method can accomplish their goals in the following four different ways: a) Automatic Kernel selection; b) resilient patch size selection; c) CNN layer; and d) pooling layer modification. In addition, it also modifies the pooling layer with average value and calculate the pixel size. The proposed method was applied on ten different image datasets. Finally, the proposed model is compared to three benchmarking models: such as WCNN, MLP, and ELM-CNN to estimate its performance. The obtained results shows that the proposed method gives competitive results compared to the other models.

Key words: classification, machine learning, deep learning, image classification, convolution network

AMS subject classifications. 68T05

1. Introduction.

To gain a complete image interpretation, we should not only focus on classifying different images, but also try to precisely estimate the concepts and locations of objects contained in every image. In general, pictures contain numerous connected data, which causes troubles within the classification errand and permits opening on diverse investigate challenges. In addition pictures may contains commotions and peculiarities, the separation between these substance is regularly troublesome and can lead to a tall rate of misinterpretation. The extraction of objects and characteristics is exceptionally troublesome and vital for an exact classification [1]. Deep learning-based methods have grown quite prominent among many Artificial Intelligence methodologies. It's utilised to solve different challenges in the fields of object detection, image classification and computer vision. Deep learning has become a very promising technology as a result of technological advancements and the availability of large datasets. Convolution Neural Network (CNN) is one of the prominent models widely used in image classification process [2]. CNN is the main deep learning technique, have outperformed most machine learning algorithms in a variety of real-world image classification or object identification applications. It is well known that the architecture of CNNs has a significant impact on their performance [3]. Images are large-scale representations of the object, made up of several pixels and shape data. The goal of convolution is to detect a specific object in an image while ignoring undesirable pixels or image features to handle a variety of problems such as security, cartography, pixel monitoring, shape, edges, and so on [4]. In general, images are large and include a bunch of associated information, which makes classification complex and allows you to open them on multiple research channels. The task of autonomously capturing these large images is as challenging as it is vital. Furthermore, multiple object images may contain noises and irregularities, making it difficult to distinguish between these contents, perhaps leading to misinterpretation [5]. Extraction of patterns, pixels, edges, and textural features is difficult but crucial for accurate classification [6], [7]. In the disciplines of identification and detection, deep learning algorithms have exhibited real-world

*Institute of Aeronautical Engineering, Department of Information Technology, India (p.hemalatha@iare.ac.in)
†R.M.D. Engineering College, Department of Computer Science and Engineering, India (gs.cse@rmd.ac.in)
‡Vel Tech, Rangarajan, Dr. Sagunthala, R&D, Institute of Science and Technology, Department of Computer Science and Engineering, India (drdeepakrajdm@veltech.edu.in)

performance [8]. There are a number of options, including Autoencoders, Stacked Denoising, weighted CNN (WCNN) [9], Multilayer Perceptron (MLP), extreme learning machine CNN (ELM-CNN), and others [10]. The number of neurons, the number of hidden layers, and other characteristics among CNN's, can have a substantial impact on categorization results [11]. The problem statement of object location and detection is to decide where objects are found in each image (image localization) and which category each images depends. So, the channel of conventional object detection methods can be basically distinguished into four stages: a) kernel selection, b)patch selection c) CNN layer, and d) pooling layer estimation.

In this paper, we proposed a combine clustering technique with a modified CNN to propose an improved binary classification approach; hence, the number of kernels is initialised using CKmeans. In addition, we proposed an adjusting max-pooling layers using the average number of patch size value. Finally, we apply a different patch size on each image to select the optimal pixel for better classification.

The article is organized by the following sections: related works described in section-2; section -3 explains the proposed approach, whereas, section-4 is experiment and discussion, section-5 shows the result, and article end up by conclusion in section-6.

2. Related Work. Generally, images are a big scene composed of several pixels. Each of the pixels brings such important information. Unlike other types of images, where each pixel is described by one value, the pixels of the multiple objects contain images are described by a vector of values. This makes the classification of images somewhat cumbersome and complicated. Deep learning has the advantage of extracting data, which allows classification more precisely. To help readers better understand the associated research and the suggested method, this part introduces 12 [13]. CNNs and skip connections, which are regarded the background of the proposed algorithm. The work on discovering the architectures of CNNs is then examined. CNN, a multilayer perceptron, was inspired by the visual neural mechanism. A CNN usually has three layers: an input layer, a hidden layer, and an output layer. The input layer calculates the mean value and uses PCA/whitening to normalise the data. The convolution, excitation, and pool layers are commonly included in the hidden layer, which is made up of numerous neurons and connections between the input and output layers. Numerous activation functions should be employed if the CNN has multiple hidden layers [12].

The ReLU and sigmoid functions are two often used activation functions. Because of the problem of strongly correlated pixels, and change in suggested CNN architecture. For instance, Discriminatory abilities, can be increased by using the middle layer and increasing stall rates, as well as adding convolutional layers. During the convolutional operation, the filter slides horizontally (with a specific step size), then vertically (with a different step size) for the next horizontal slide, and so on until the entire image has been scanned. The feature map is a new matrix formed from the filter outputs. The breadth and height of a stride are the horizontal and vertical step sizes. In our research, we have discussed WCNN [9], MLP [10], and ELM [11].

The weight parameter affects input data inside the network's hidden layers, making WCNN a neural network. A neural network is made up of a series of nodes (also known as neurons) that connect together to form a network. Each node is made up of a collection of inputs, a weight, and a bias value. A node receives an input multiple times, and the same it multiplies by a weighting factor to observes or either sends the result to the next hidden layer of the network [14]. The hidden layers of a neural network are typically used to store the network's weights.

A neural network's input layer accepts input signals and forwards them to the next layer. After that, the neural network has a series of hidden nodes that modify the data input. The weights are applied to the buried layer nodes' nodes. For instance, a single node may multiply the incoming data from a well before weighting factor and transferring things to a whole new layer, and then add a bias. The output layer of the neural network is indeed known as the final layer. The concealed layers' input is regularly tweaked by the output layer to create the needed numbers in a specific range represented in Fig. 2.1 (a). A feedforward artificial neural network called a multilayer perceptron (MLP). An MLP is made up of three layers of nodes: an input layer, a hidden layer, and an output layer. Except for the input nodes, each node is a neuron with a nonlinear activation function [15]. MLP employs backpropagation during training as a supervised learning technique. The multiple layers and non-linear activation of MLP separate it from a linear perceptron. It can distinguish between non-linearly separable data. Multilayer Perceptron (MLP) is a sort of computer vision technique that has been replaced by CNN's as seen in Figure 2.1(b). MLP is no longer considered adequate for advanced computer vision tasks.

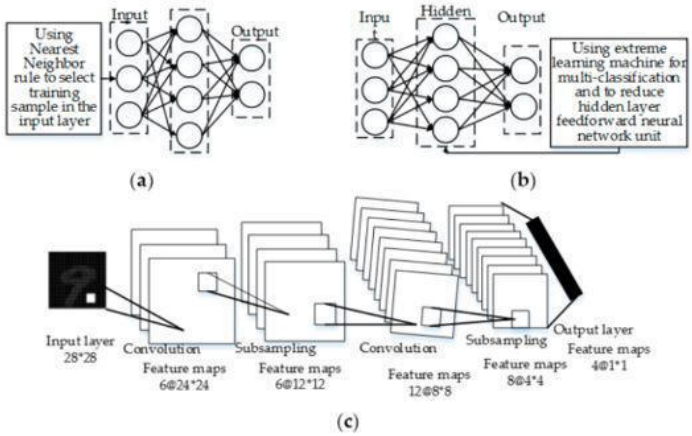


Fig. 2.1: Model of comparison algorithms using a convolution neural network structure a) WCNN, b) MLPCNN and C) ELM-CNN.

Every perceptron gets linked to all of the other perceptron's, resulting in fully connected layers [16].

The extreme learning machine (ELM), a revolutionary paradigm in the field of machine learning, has received a lot of academic attention in recent years. The predictive model is a generalised version of the single-layer feedforward neural network (SLFN), which only contains one hidden layer with randomly generated hidden neurons. SLFNs are networks that do not have any back or side connections between nodes. Fast computational time during training and testing is one of ELM's primary features. Furthermore, unlike typical neural networks, the classifier is capable of achieving strong generalisation without the need to tune network parameters [17]. Figure 2.1(c) shows how ELM has been applied in a range of different learning contexts. ELM, in comparison to earlier tuning-based approaches, necessitates a greater number of hidden nodes in practise. However, in such networks, a small number of hidden nodes may have a minor impact on the network's output, boosting the network's intricacy [18]. Furthermore, if indeed the current hidden node surpasses total training set and the ELM may experience a singularity problem, making the system unstable.

2.1. CNN – General Working Principle. CNN is a type of neural network that has shown to be particularly effective in areas like image recognition and classification. Applications include object detection, vision in robots, autonomous cars etc., and it also used in smart grid applications as well, because, smart grid generates huge amount of data. The general working procedure of CNN is shown in figure 2.2(c). Typically, CNN has three types of layers to build architecture such as convolutional layer with ReLU activation, Pooling layer and fully connected or dense layers. Depict from the figure 2.2, convolutional layer extract information from the input picture, the next layer is pool1 layer reduce extracted information from conv1 which is also called down-sampling layer [19], [20]. Next layer is pattern which is used to convert matrix into vector. Dense layer usually available at the end of convolutional layer, based on fig. 2.2 the process to classify the given image is cat or dog, and the output layer exactly says that the given image is cat or dog with SoftMax activation. Convolution's primary objective is to collect the significant features of an image. By learning input features, convolution retains the spatial link between input. Then the convolutional layer produces output by simply convolving input image with filter. The mathematic representation of CNN is defined as: where y→ output image, x→ input image, which is actually representing the input image with $N * N$ matrix, n represent the size of the input image, 'f' represents size of filter which was chosen for the given example. The size of the filter can change as per the requirement or problem description. Typically, computer interpret image in the form of grey scale (i.e., 0 and 1) and RGB (i.e., 0-255). Depict from figure 2.2, the given image convoluted and distinguished by a layer action in which image's pixels are extracted one by one in a given size patch x→ $n * n$ matrix, with each active pixel enclosed through group of nearby pixels. Second, each patch will traverse the

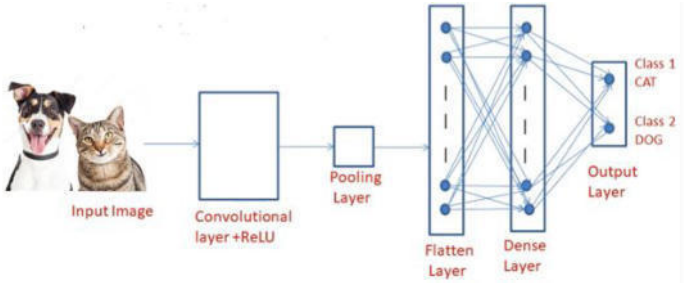


Fig. 2.2: General working procedure of CNN.

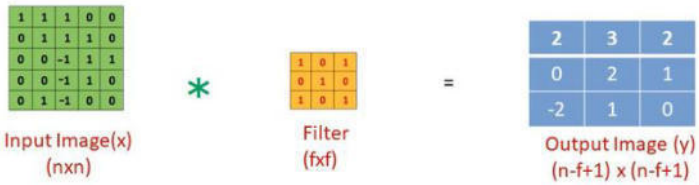


Fig. 2.3: Convoluting the input images 5*5 with 3*3 filter and produce output image 3*3 dimensions.

convolution and max - pooling several times. Third, it will be twisted using the various specified kernels at the convolution layer. According to the filter selected, it would be treated only at max pooling. Finally, the patch gets smaller and smaller until it's just a simplified data vector.

$$y = \sum_{n=1}^{n} x * f \tag{2.1}$$

The content of the pixel will be determined by comparing these values to the database [21], [22]. From the figure 2.3 and 2.4 the input image sample chosen as $5*5$ dimension with filter $3*3$ dimension. When the input image is convolved with $3*3$ filter produce output image $3*3$ dimension.

In order to find (1,1) element of this output with $3*3$, lets super impose this filter into input image such that (1,1) of this filter is super imposed on (1,1) of input image; (1,2) of this filter is super imposed on (1,2) of an input image and so on till (3,3) of this filter is super imposed on $3*3$ element as shown on in figure 3(a). Therefore, element wise multiplication $1*1$, $1*0$, $1*1$, $0*0$. $1*1$, $1*0$, $0*1$, $0*0$, $-1*1$ produce result 2. Similarly, if we want to find $1*2$ element of this output image matrix we have to shift this filter ones on the right side, so that (1,1) filter super imposed on (1,2) of an input image and (1,2) of this filter is super imposed on (1,3) of an input image and so on till (3,4) of this input image. And, once again the element wise multiplication and add all the values will be finding 3 and so on till finding 0 in the out image. To understand little more about the pixel calculation is broadly described in figure 2.4.

3. Proposed Method. We start by separating the cat image (named H) into multiple scales because the input image is 2D shaped. Each scale is dealt with independently. Kernels are extracted from each scale. However, the problem is figure ring out how to identify the number, position, and length of kernels, as well as the length of both the patches that encircled each pixel. To solve this problem, we proposed an algorithm called IBCNN (Improved Binary CNN) using the CK-means method with a few improvements. (1) Kernel selection (Number, Position, and Size), (2) Patch Extraction, (3) Hidden layer processing, and (4) Classification are the four primary procedures in our system as shown in algorithm 1.

3.1. Kernel Selection. The proposed algorithm begins making a feature map, which contains locations of all the important pixels (kernel centre). In this paper, the CK-means algorithm has been adopted to determine

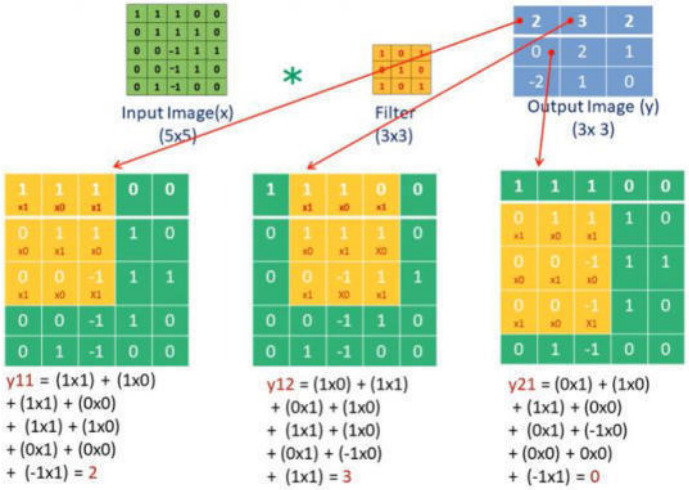


Fig. 2.4: Comprehensive convolution layer operation.

kernels. CK-means intends to stake the X= $x_1, x_2, x_3.x_n$ in vague pixel 'p' with 'μ_{ij}' represent cluster along with $X_i \rightarrow$ degree of membership in respective of ith clusters. The membership degree on the matrix represents the clustering outcome μ.

First and foremost, set the parameters: p $\rightarrow$ number of clusters n $\rightarrow$ number of training data m $\rightarrow$ fuzzy parameter used to define the width of p parameter

For the number of cluster p initialization, for the two relative integers n and p the Euclidean distance division was used to find the mean difference between patch size along with zero value in distinct p. Euclidean division connects two relative integers i.e., quotient $\leftarrow$ n, remainder $\leftarrow$ m, with respect to training sample n: n = pm + w, where $0 \leq$ w $\leq |\, p \,|$. This scenario could help to cover as much area as possible, whereas, 'w' should be decreased with respect to the predefined parameter m, n and p; and m ϵ (1.25,2). Procedure involved in the proposed algorithm as follows: To begin, split each pixel and construct x vectors of data one by one. For data ranging from 1 to n: 1) In order to generate number of clusters p, the fixed predefined fuzzy parameter m have to select random patch size in cluster. 2) Generate novel matrix Φ; after the membership degree φ has been determined.

$$\Phi_{ij} = max\Big(\frac{\Phi_{ij}}{\sum_{i=1}^{n} \Phi_{ij}} \mid \frac{\Phi_{ij}}{\sum_{i=1}^{n} \Phi_{ij}}\Big) \tag{3.1}$$

from eqn (3.1) follow the procedures to calculate the Φ_{ij} matrix: i) To begin, fill a random number between 0 and 1, to ensuring that the sum of each row or column is 1. ii) second, the largest integer in a place can be given the value 1 in a new matrix, while the others are given the value 0. iii) at last, create a vector that consisting total of the values in each column. Give a value in column as 1 if it includes or else 0, thus ultimately assist to find the largest value in vector. After that, calculate the centroid cluster '$Cent_j$' with respect 'j'.

1) Use eqn(3.2) to calculate the centroid value of the cluster can be defined as:

$$Cent_{ij} = \frac{\sum_{i=1}^{n} x_{ij}\Phi_{ij}}{\sum_{i=1}^{n} \Phi_{ij}} \tag{3.2}$$

2) Calculate initial value j_A using eqn(3.3):

$$J_A = \sum_{i=1}^{n}\sum_{j=1}^{n} \Phi_{ij}, d_{ij}(x_i, c_j)^2 \tag{3.3}$$

Data: set of input image datasets D for classification, Maximum pixel number generation
Result: classification of given images y=Cn (n=1,2,3…
Initialization $I_0 \leftarrow$ Using the proposed variable-kernel selection technique, initialise a pixel size with the provided image size.
step 1 Kernel selection
while $t \leftarrow$*0 the maximal generation* **do** $\Phi_{ij} = max\left(\frac{\Phi_{ij}}{\sum_{i=1}^{n}\Phi_{ij}} \mid \frac{\Phi_{ij}}{\sum_{i=1}^{n}\Phi_{ij}}\right)$
if Φ_{ij} *is even* **then**
estimate number, position, and size of the kernel with ckmeans, **then**, calculate the centroid value of the cluster
$Cent_{ij} = \frac{\sum_{i=1}^{n} x_{ij}\Phi_{ij}}{\sum_{i=1}^{n}\Phi_{ij}}$
elseif *threshold* $\theta > 0$
step 2 patch size estimation *calculate the average of the current pixel and its neighbours pixels using* $W_{i,j}$
then$W_{ij} = (\frac{x_{vi}+x_k}{n}, \frac{y_{vi}+y_k}{2})$
$p_i x_1(x_i, y_i) = (x_{ci} \colon y_{lu})$
step 3 CNN layer
for each cnn $\rightarrow C_i : Ci \leftarrow I_0 * W_1$ select active pixel size $C_i \rightarrow$ choose 'n' α and β to map convolution layer l calculate the average of the current pixel:
$y_n^{l-1}(i,j)\sum_{x=1}^{a=1}(y_n^{l-1}, (i-x, j-1), ...y_n^{l-1}(1-x, j), ...y_n^{l-1}(i, j-x).., y_n^{l-1}(i,j))$
step 4 Average_pooling layer
for each the filter size is a*a and
if a ≥ 1 **then** $y_n = f_1(z_n^{l-1} * w_n^l + \beta_n^1)$
end if
end if
end while

Algorithm 1: Proposed IBCNN Algorithm

3) Calculate fuzzy parameter $Cent_{ij}$ using eqn(3.4):

$$Cent_{ij} = \frac{(\frac{1}{d_{ij}(x-i,c_j)})^{\frac{2}{m-1}}}{\sum_{i=1}^{n}(\frac{1}{d_{ik}(x_i,c_k)})^{\frac{2}{m-1}}} \tag{3.4}$$

4) By using Euclidean distance find dij between x_i and c_j.

5) Check the finishing condition, and then set the positive parameter $\in > 0$ from the beginning.for better result we chosen $\in = 0.001$.

$$d_{ij}(J_{a-1}, J_A) \leq \in \tag{3.5}$$

If $J_A$is the previous iteration's objective function and $J_{(A-1)}$ is the current iteration's objective function, then move to eqn. (3.3) to (3.5) otherwise return to (2). We build a binary matrix from the clustering matrix to represent the kernel placements. To accomplish this, we've established a threshold parameter θ. The number of neighbouring pixels[3, 2], that are similar to centre pixel or active pixels $\rightarrow$ k. Active pixel calculating is illustrated in greater detail in the following diagram figure 3.1. Where, I $\in$ [2, 0], in the centre pixel or active pixel 'x', if $p_i x_i = =$ X, then the nearest pixel value could $N++$. However, if $N \geq \theta$ the centroid cluster X must be marked as 1 otherwise 0.

3.2. Patch size estimation. In patch calculation, initially we have to set a threshold θ to consider. Based on the size of the image the pixel value may assigned. For instance, let assume v be the nearest kernel centre neighbors i.e., v = v1, v2, v3, v4. The smaller value has been chosen by calculating the distance between each pixel and each kernel's centre. Then, calculate the average of the current pixel and its neighbours' pixels using

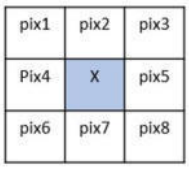


Fig. 3.1: Estimation of centre pixel 'x'

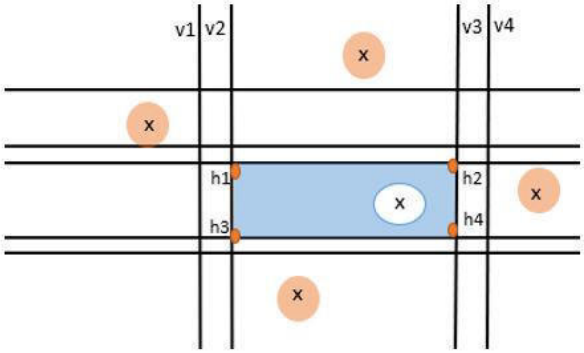


Fig. 3.2: Selection of kernel size in coordinates intersection

eqn 3.6.

$$W_{ij} = \left(\frac{x_{vi} + x_k}{n}, \frac{y_{vi} + y_k}{2} \right) \tag{3.6}$$

The ordinate of the graph is represented by x and y point respectively with denoting the active pixel k. We choose the equivalent outlines with respect to x and y axes, that cross on all the generated points after selecting the new points. Constant affine functions fc(x) is used for to estimate the neighbors kernel function i.e., fcx = ax+b, where, constant a=0, and co-ordinates of the neighbor kernels b. Four horizontal lines h1, h2, h3, h4 which represent rows and four vertical lines v1, v2, v3, and v4 represents columns are used to find the co-ordinates intersection values. The rectangle's coordinates to be retrieved or represented by intersection values, as shown in Fig 3.2. The active pixel's space '1' is then divided and shifted into four sections: up, down, left, and right. The new parameter position p can estimate as: $p_i x1 = m_n \wedge x_l$, $p_i x2 = m_n \wedge x_r$, $p_i x3 = m_d \wedge x_l$, $p_i x4 = m_d \wedge x_r$ with resection lines ln and column cn. The position of the active pixel can be evaluated with the build objective function p_i(x, y) = (xL n: d + xC r: l + yL u: d + yC r: l). Therefore, when applying pi(x, y) to calculate the neighbors pixel size the intersection coordinates y_C and x_l are seeming to be equal because they are parallel to the axes. So, some position of the parameter might change with respect to the initialization values and the appropriate steps taken to adjust as follows:

select_start_position_1

$$p_i x_1(x_i, y_i) = (x_{ci} : y_{lu}) \tag{3.7}$$

select_start_position_2

$$p_i x_2(x_i, y_i) = (x_{cr} : y_{lu}) \tag{3.8}$$

select_start_position_3

$$p_i x_3(x_i, y_i) = (x_{ci} : y_{ld}) \tag{3.9}$$

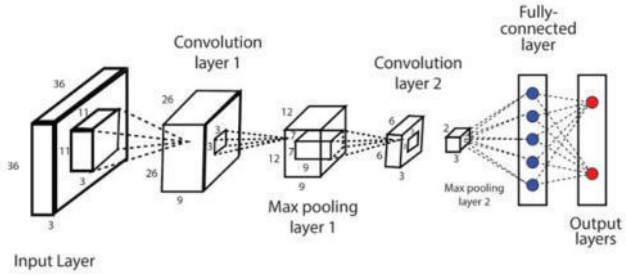


Fig. 3.3: Overview of CNN work cycle.

select_start_position_4

$$p_i x_4(x_i, y_i) = (x_{cr} : y_l) \tag{3.10}$$

Training, testing, and validation are all performed on the extracted sample from the input image. Therefore, the training samples will be subjected to the following treatment. The pixels in the centre of the kernel X are used to find the co-ordinates of pixel h1-h2:h3-h4 and vertex v1-v2:v3-v4 as shown in fig 3.2.

3.3. CNN Layer. The convolution layer weights the kernels chosen and adds local neighbours to every image pixel. The centre element of each kernel is positioned upon that active pixel as shown in fig 3.3. A weighted sum of the neighbours' pixels and itself will replace active pixel. Depict from eqn. 3.11 let y_n be given input image determines the image patch size Bi and number of layer 'n' α and θ are the two positive integers used to map the convolution layer 'l'.

$$y_n^l = f_i(\sum_{m \epsilon n} y_m^{l-1} + \beta_n^1) \tag{3.11}$$

However, a convolutional operator represented as ϕ and f_i is an activation function of the layer, in the average_pooling layer S_1, the bias for attributes map β_n^1 The W_n^l column displays a pixel group on each layer 'l-1' which related to the n-dimensional characteristic map. At last, the convolution kernel W_n^l, come into the picture to calculate and assign the different filters.

3.4. Average pooling layer. The average_pooling layer permits each average_pooling pixel's input size to be halved or greater (i.e., depending on the filter size). Estimate the average pixel size in term of a*a for each patch, which corresponds to the filter's length; before being added to the bias, such outcome would be increased from a tuneable weight. In every average_pooling layer if the filter size is a*a and if a ≥ 1 as well then, the output N*N matrix $\rightarrow z_n^{l-1}$(i,j) can be calculated as:

$$y_n^{l-1}(i,j) = \sum_{x=1}^{a=1}(y_n^{l-1}, (i-x, j-1), ...y_n^{l-1}(1-x, j), ...y_n^{l-1}(i, j-x).., y_n^{l-1}(i,j)) \tag{3.12}$$

Using eqn. 3.12, let $y_n^{l-1}(i,j)$, as a result of the average of such images in every patch and the features map 'n', the average_pooling layer 'l' is now computed with respect to y_n input shown in eqn 3.13.

$$y_n = f_1(z_n^{l-1} * w_n^l + \beta_n^1) \tag{3.13}$$

The output image map size $Z^l W^l$, where $H \rightarrow$ horizontal position of the pixel and $W \rightarrow$ represent vertical position of the pixels. In the final convolution layer, each pixel is linked to exactly 1 preceding features map. The input characteristic maps are convolution kernels from the same size. Sigmoidal neurons make up the output layer referred as 'y', whereas, 'n' specifies the amount of sigmoidal output neurons. Therefore, eqn. 3.12 is the equivalent equation for computing the output of a sigmoidal neuron. Here β seems to be the biased linked with both the 'L' layers and 'n' neuron, and 'w' is the weight of the last convolutional layer's characteristic map 'P' at the output layer's neurons 'N'. Finally, compute all of the sigmoidal neurons' outputs that lead to a network's results 'y':

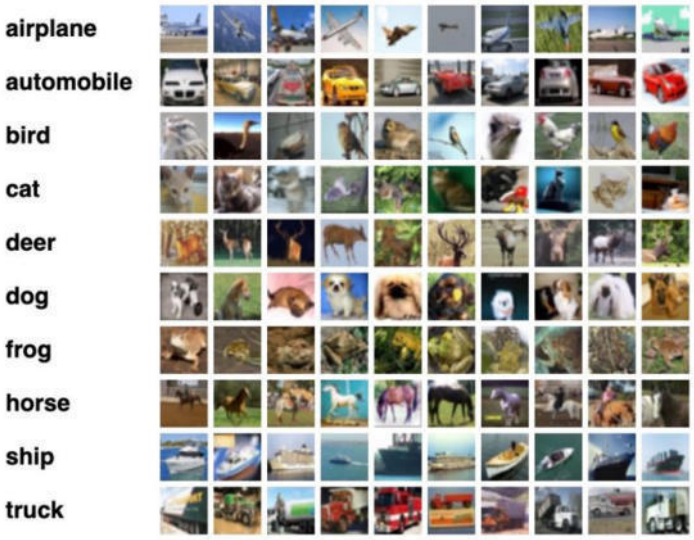


Fig. 5.1: Visual description of the dataset. However, column represent the dataset label and row represents the types of the dataset.

Table 5.1: Characteristics of the dataset

Name of the Dataset	Total number of images	pixel dimension
airplane	10028	227 * 227
automobile	45001	1080 * 1080
bird	35484	800 * 600
truck	14750	720 * 720
frog	17584	227 *227
deer	14856	1280* 1227

4. Experiment and Discussion. The quadratic CNN learning method is implemented on a 64-bit Ubuntu 16.04.4 using the Python programming language. TensorFlow 1.5.0 is used to evaluate the accuracy of the proposed method. The reliability of said dropout-based CNN algorithm and the SGD optimizer are assessed [23], examined, and associated to the other three models at diverse learning rates such as training loss or training accuracy and validation loss or validation.

5. Dataset Description. To demonstrate the efficiency of our suggested method, we evaluated the model on ten different image datasets. The characteristics of the datasets is shown in table 1 and visually in Fig 5.1. The method was compared with different models such as weighted CNN (WCNN), convolution with fully connected multilayer perceptron (MLP-CNN), external learning with support vector machine (ELM-SVM)[24]. However, the datasets are publicly available at(https://paperswithco
de.com/dataset/cifar-10).

6. Result. We retrieve the pixel size from each input image for experimental purposes. In tradition CNN various nonlinear functions such as Tanh, Sigmoid, ReLU and leaky ReLU are used as activation functions. However, in this work, we have used ReLU [25], as a common activation function for the entire process which help easily to maps a real number in to [1,0]. For each pixel, we estimate a cluster matrix, then create a new feature map (binary matrix) that specifies the locations and sum of kernels from each matrix which could help to retrieve the kernel size from the feature map. Typically, the proposed method conducted two different tests:

Table 6.1: setting patch size, pooling size, convolutional kernel size

Dataset	airplane	automobile	bird	apple	fish	baby	cats	dogs
size of the patch	36*36	36*36	24*24	21*21	36 * 36	21 * 21	9 * 9	36 * 36
Convolutional kernel size	12*12	12*12	6*6	5*5	5*5	9*9	5*5	5*5
Pooling filter size	9*9	9*9	5*5	3*3	3*3	6*6	3*3	3*3

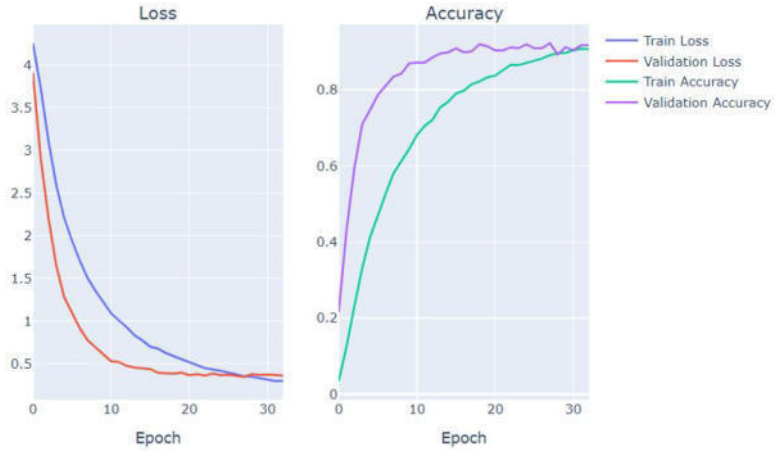


Fig. 7.1: Performance analysis in term of loss and accuracy

first, it involved applying a constant patch size (see Table 5.1). Secondly, we returned patches of varying sizes, the size of which is governed by the dimensions of the input image and the number of kernels produced.

To test if our proposed technique is effective, we first compare the performance to an arbitrarily pre - defined number of kernels. Each of the ten datasets is subjected to the tests. The proposed algorithm calculates the size of the kernel in the second testing, usually mean number of hidden layers completed on all datasets is between 6 and 12 depending on the number of kernels generated as well as the magnitude of such original pixel size. The suggested technique finds the number of kernels to be 50, 55, 60, 65, 70, 75, 80, 85, 90, and 95 for each of the ten datasets. Indeed, in the other variants, we manually chose the number of kernels: 35, 40, 45, 50, 55, 60, 65, 70, 75, and 80. The precision or number of kernels on every approach was set according to the values defined in columns, are shown in Table 6.1. This suggested methodology offers the highest categorization accuracy. As a result, we can conclude that the suggested method can locate a large number of kernels for various images.

7. Performance Analysis. The performance analysis of the proposed model evaluated in term of loss and accuracy of validation and train metrics. However, the validation accuracy range is increased from 0.8 to 0.95. The recognition rate of the model increases as the learning rate increases, and the learning and recognition rates have a positive relationship, as shown in Figure 7.1. When the SGD optimizer's training an accuracy learning rate reached 0.94, the model's recognition rate was relatively constant. The validation accuracy and train accuracy of the proposed model are currently growing faster than the previous model, The accuracy rate rises considerably whenever the learning rate goes up to more than 0.06. The proposed algorithm is executed ten times on each dataset. Table 6.1 describe the kernel size as well the filter size, and Figure 7.2 illustrate the accuracy of the proposed algorithm comparing with other algorithms.

According to [26], for the airplane dataset test, the proposed algorithm achieved 2% higher than MLP-CNN, however the parameter we selected here to test is 10.2M. Similarly, IBCNN achieved greater result on bird, apple, fish, baby, and dogs. However, MLP-CNN achieved 1% higher on cat's dataset than other models. WCNN and IBCNN performed 2% greater on apple dataset. ELM-SVM attained reasonable accuracy of 91% with 2.2M

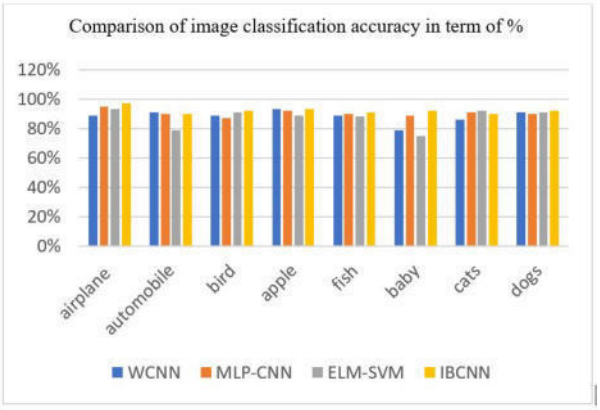


Fig. 7.2: Accuracy comparison with other image classification models

parameter, with 9*9 patch size, however, these accuracies are estimated by selecting the patch size and manual parameter [27], [28]. Another parameter: pooling filter size 5*5, 9*9, 3*3 are used to test WCNN, IBCNN, MLP-CNN. Similarly, the validation accuracy is obtained 0.95M which is 2% higher accuracy than another model. Depict from figure 7.2, the accuracy of classification is evaluated in %, based on training accuracy and validation accuracy IBCNN gained 97% on airplane dataset, 93% on apple, 91% on fish, 92% on baby, 92% on bird, and 92% on dogs which is greater than WCNN, MLP=CNN, ELM-SVM. As a result, we can conclude that the suggested method can locate a large number of kernels for various images and obtained highest accuracy in classification of object in image. The proposed technique has the highest accuracy in classification. As a result, we can conclude that the suggested method can locate a large number of kernels for various images. The classification accuracy of the proposed approach is marginally greater than that of existing CNN models for peer competitors in the second category.

8. Conclusion. Object detection in image is a cutting-edge technique that can be used in a variety of security, autonomous, and surveillance applications. Classifying and object identification in large picture or image in deep learning is achieved benchmarking result with the help of convolutional network. The use of CNN for image classification is a solution that involves data and parameter initialization. We developed a classification approach in this paper that combines clustering and convolutional neural networks. There are four significant contributions of the work as: (1) using a clustering algorithm to determine the number of kernels; (2) picking the kernel placements; (3) computing an adaptive patches size; and (4) softening the pooling layer. On ten different image datasets, the proposed approach was tested. The proposed model's performance was assessed in terms of validation and train metrics loss and accuracy. The validation accuracy range is extended from 0.8 to 0.95, the model's recognition rate increases as the learning rate increases, and the learning and recognition rates have a positive relationship. The results show that the proposed strategies are competitive with alternative approaches. The further work suggested on spatial images or hyperspectral image classifications.

REFERENCES

[1] Milan Tripathi, *Analysis of Convolutional Neural Network based Image Classification Techniques*, Journal of Innovative Image Processing (JIIP), Vol. 3 No.2, July 2021, pp 100-117.

[2] Y. Sun, B. Xue, M. Zhang, G. G. Yen and J. Lv , *The LaTeX Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification*, IEEE Transactions on Cybernetics, Vol. 50 No. 9, September 2020, pp. 3840-3854.

[3] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger , *Densely Connected Convolutional Networks* , IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261-2269.

[4] L. Xie and A. Yuille , *Genetic CNN*, In Proceedings of 2017 IEEE International Conference on Computer Vision, 2017, pp. 1388–1397.

[5] Y. Sun, B. Xue, M. Zhang and G. G. Yen , *Evolving Deep Convolutional Neural Networks for Image Classification*, in IEEE Transactions on Evolutionary Computation, Vol. 24 No. 2, April 2020, pp. 394-407.
[6] M. Liu, D. Cheng, and W. Yan , *Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images*, Frontiers in Neuroinformatic, Vol. 12 No 2, March 2018, pp. 35-43.
[7] Feng-Ping , *A Medical Image Classification Algorithm Based on Weight Initialization-Sliding Window Fusion Convolutional Neural Network* , Hindawi, Complexity, Vol. 2019 No.12, Article ID 9151670, 2019, pp. 1-15.
[8] A. Singh , *Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM* , In Proceedings of the 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 2015, pp. 98–102.
[9] Yang, Jing, and Guanci Yang , *Modified Convolutional Neural Network Based on Dropout and the Stochastic Gradient Descent Optimizer* , MDPI Algorithms, Vol. 3 No. 28, March 2018, pp.1-15.
[10] Singh P., Verma A., Chaudhari N.S , *Deep Convolutional Neural Network Classifier for Handwritten Devanagari Character Recognition* , Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing, Vol 434. Springer, 2016, pp. 551-561.
[11] Mahmood, Saif F., Marhaban, Mohammad H., Rokhani, Fakhrul Z., Samsudin, Khairulmizam and Arigbabu, Olasimbo Ayodeji , *SVM–ELM: Pruning of Extreme Learning Machine with Support Vector Machines for Regression* , Journal of Intelligent Systems, Vol. 25 No. 4, 2016, pp. 555-566.
[12] R.R. Vargas, B. Rene, C. Bedregal, E.S. Palmeira , *A Comparison between KMeans, FCM and CKmeans Algorithms* , Theoretical Computer Science, Vol 7 No.3, 2011, pp. 15-28.
[13] Zhou, F.Y., Jin, L.P. and Dong , *A Review of Convolutional Neural Networks. Journal of Computers* , World Journal of Engineering and Technology, Vol.9 No.4, 2021, pp. 1229-1251.
[14] ZhiFei Lai, HuiFang Deng , *Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron* , Computational Intelligence and Neuroscience, Vol. 2018 No.13, 2018, pp. 1-13.
[15] Zinkevich, M, Weimer, M., Li, L. Smola , *Parallelized stochastic gradient descent. In Advances in Neural Information Processing Systems* , Neural Information Processing Systems Foundation, 2010, Vol. 23, No.23, pp. 2595–2603.
[16] Luo, P, Li, H.F , *Research on Quantum Neural Network and its Applications Based on Tanh Activation Function* , Computer and Digital Engineering, Vol. 16, No.13, March 2012, pp.33-39.
[17] Günnemann, N, Pfeffer, J , *Predicting Defective Engines using Convolutional Neural Networks on Temporal Vibration Signals* , In Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications, Vol. 74, 2017, pp. 92–102..
[18] Shi, X.B.; Fang, X.J.; Zhang, D.Y.; Guo, Z.Q , *Image Classification Based on Mixed Deep Learning Model Transfer Learning*, Journal of System Simulation, 2016, 28, pp.167–173.
[19] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio , *Maxout networks* , In Proceedings of the 30th International Conference on Machine Learning, 2013, Vol. 41, pp. 1319–1327. .
[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L.Fei , *Image-net large scale visual recognition challenge* , International Journal of Computer Vision, Vol. 115, No. 3, pp. 211–252, 2015.
[21] K. Nag and N. R. Pal , *A multi objective genetic programming-based ensemble for simultaneous feature selection and classification*, IEEE Transactions on Cybernetics, 2015, Vol. 46 No. 2, pp. 499–510.
[22] G. Xuan, *Apple Detection in Natural Environment Using Deep Learning Algorithms* , 2020, Vol. 8, pp. 216772-216780.
[23] Behera, Santi Kumari; Rath, Amiya Kumar; and Sethy, Prabira Kumar, *Fruit Recognition using Support Vector Machine based on Deep Features*, Karbala International Journal of Modern Science, 2020, Vol. 6 No. 2, pp. 235-245.
[24] J. Samuel Manoharan, , *Capsule Network Algorithm for Performance Optimization of Text Classification* , Journal of Soft Computing Paradigm (JSCP), 2020, Vol.03 No.01, p.p1-9.
[25] M. Pan, Y. Liu, J. Cao, Y. Li, C. Li and C. -H. Chen , *Visual Recognition Based on Deep Learning for Navigation Mark Classification*, in IEEE Access, 2020, vol. 8, pp. 32767-32775.
[26] Díaz-Pernas FJ, Martínez-Zarzuela M, Antón-Rodríguez M, González-Ortega D , *Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural*, Network. Healthcare (Basel), 2021, vol.9 No.153, pp.1-14.
[27] Krizhevsky, A.; Sutskever, I.; Hinton, G.E., *ImageNet classification with deep convolutional neural networks. Comms*, ACM, 2017, vol.60, p.p 84–90.
[28] Chen, L.; Wu, C.; Fan, W.; Sun, J.; Naoi, S , *Adaptive Local Receptive Field Convolutional Neural Networks for Handwritten Chinese Character Recognition*, In Chinese Conference on Pattern Recognition, 2014, pp. 455–463.

Edited by: Vinoth Kumar
Received: Jun 30, 2022
Accepted: Dec 5, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 275–289, DOI 10.12694/scpe.v23i4.2030

A META HEURISTIC MULTI-VIEW DATA ANALYSIS OVER UNCONDITIONAL LABELED MATERIAL: AN INTELLIGENCE OCMHAMCV

SRINIVAS KOLLI *, PRAVEEN KRISHNA A.V †, AND M. SREEDEVI‡

Abstract. Artificial intelligence has been provided powerful research attributes like data mining and clustering for reducing bigdata functioning. Clustering in multi-labeled categorical analysis gives huge amount of relevant data that explains evaluation and portrayal of qualities as trending notion. A wide range of scenarios, data from many dimensions may be used to provide efficient clustering results. Multi-view clustering techniques had been outdated, however they all provide less accurate results when a single clustering of input data is applied. Numerous data groups are conceivable due to diversity of multi-dimensional data, each with its own unique set of viewpoints. When dealing multi-view labelled data, obtaining quantifiable and realistic cluster results may be challenge. This study provides unique strategy termed OCMHAMCV (Orthogonal Constrained Meta Heuristic Adaptive Multi-View Cluster). In beginning, OMF approach used to cluster similar labelled sample data into prototypes of dimensional clusters of low-dimensional data. Utilize adaptive heuristics integrate complementary data several dimensions complexity of computational analysis data representation data in appropriate orthonormality constrained viewpoint. Studies on massive data sets reveal that proposed method outperforms more traditional multi-view clustering techniques scalability and efficiency. The performance measures like accuracy 98.32%, sensitivity 93.42%, F1-score 98.53% and index score 96.02% has been attained, which was good improvement. Therefore it is proved that proposed methodology suitable for document summarization application for future scientific analysis.

Key words: Clustering, Document summarization, Data mining, Meta heuristic technique.

AMS subject classifications. 68T05

1. INTRODUCTION.

Large amounts data gathered from several study fields, such image processing, computer vision data fusion, natural language and processing in real time as result of fast computer-related technology deployment. A wide range of dimensions associated to a wide variety of properties are examined in these data, which include many high-dimensional features with complicated structures [1, 2]. High-dimensional data represents the abundance of data curse dimensionality, therefore managing high-dimensional data general concern big challenge for optimizing the dimension's dimensions. Using hidden data to represent low dimensionality and reducing dimensionality in relation to input data is an effective method for large amounts of data [3].

Theoretically optimized matrix factorization has emerged as the research hotspot with the easiest implementation for multi-labeled data reduction. It's possible extract low-dimensional attribute relations high-dimensional data relations using factorization matrix-related methodologies such ICA (Independent Component Analysis), PCA (Principal Component Analysis), & VQ (Vector Quantization). No components in matrices are decomposed; this implies that in order to maximize matrix representation, negative elements must be included in the low-dimensional representations of data [4]. Deep learning has recently proven exceptional performance in include representation projects [5]. Lattice factorization has been enriched by various analysts who have incorporated substantial learning into the process [6]. A multi-layer non-negative MF technique presented (MNMF) [7]. First, MNMF degraded the grid many times to produce the fundamental part-based representation that may remove profoundly different degrees of information from the original information. To propose

*Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India (kollisreenivas@gmail.com).,

†Associate Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India (praveenkrishna@kluniversity.in).

‡Professor, Department of CSE, Amrita Sai Institute of Science and Technology, Andhra Pradesh, India. (srivijaysujay@gmail.com).

thorough semi-non-negative grid factorization strategy [8, 9] employed Semi- non-negative MF (SNMF) and coordinated deep Factorization [10, 11]. But the deterioration of the coefficient network for preparation information in both MNMF and deep semi-NMF can only be seen as a profound decay in this network [12, 13]. The premise grid was used to minimize dimensionality new data problems construct new data issues [14, 15]. When premise framework applied to the deep representation, it had a direct impact on the outcome. Using factorization, investigated the accurate reduction of dimensions with depth of initial information framework, and they presented a profound NMF technique that relies on premise picture learning [16, 17].

Invariant data acquired from several data domains/sources may create this issue. Represent data several dimensions depending their representation features and relationships with multiple features, this challenge tackled using multi labelled clustering [18, 19]. Different techniques to multi-view clustering have been explored before, but they have not been analyzed in terms of dimensionality reduction in the representation of multi labelled data in supervised learning. In order to identify the text, the supervised learning technique defines the labelled information based on its features [20, 21]. It is thus crucial to find data sources for unsupervised learning with multiple labelling [22, 23]. Because finding quantifiable and realistic cluster outcomes multi-view labelled data still challenge, new technique called OCMHAMVC being offered representing Data as cluster with many kinds. To begin, first suggested methodology examines low-dimensional data using OMF model, clusters comparable labeled sample data prototype clusters data related several dimensions [24, 25].

The following are the main goals of the suggested method:

1. Unsupervised multi-labeled clustering method uses orthonormality matrix factorization (mix normality constraints and orthogonal constraints) is first proposed.
2. Objective model we implement an objective model, which provides and expresses the minimums of the suggested implemented model
3. In order to demonstrate efficiency of the suggested technique compared to standard approaches accuracy, other metrics multi-labeled cluster data sets, we conduct tests on numerous real time datasets.

2. Review of Related Work. Using multi-labeled data, this section explores the relationship between standard clustering methods and their results.

Prior to this time, a wide variety of single-view data grouping solutions had been discussed and implemented. Three typical single-sight collection strategies piece bunching [22, 23, 24], paranormal groups [26, 27] & subspace groups [28, 29]. Most part, bit-based methods employed build primary commitments High-Dimensional piece space where proffered grouping successfully [30]. employ Gaussian piece design commitments split region and wire pair- savvy constraints into part sorting out some methods co-ordinate pattern collecting [31] preshown bits used design data sources and improvement piece game plan better encourage bunching execution [31, 32], optimum portion space picked from social event predetermined portions. In the run-up to the terrible grouping, they normally produce a partiality diagram to describe information similitude and analyses Eigen structure this affection diagram acquire clustering conclusion. To create proclivity chart, existing clustering methods [33, 34] show efficient strategy.

As an example, [35] construct the inborn diagram and the punishment chart independently using pairwise requirement data. In [36], an incredible affection graph is fostered by using discriminative component subspaces. Mishandling multi-layer layers in the form of a pyramid-style structure by [37, 38, 39] creates a reformist bipartite diagram. Besides the foregoing, the collecting structure makes good use of a help vector machine (SVM). In [40] introduced Twin Help Vector Machine (TWSVC) framework, discover bundle plane close contrasting pack places avoiding characteristics other groups. Matrix factorization (NMF applied range MVC assessment methodologies non-threat impediments consider improved inter-predibility (Guan et al 2020; Trigeorgis et al 2018). Using non-negative structural factorization of multiple datasets, an average inert factor is discovered (Liu et al 2018; Zhang et al 2019, 2020). Semi-NMF has been projected to broaden NMF by relax the factorized premise structure to be actual performances. Semi-NMF is one of the most prevalent variants of Non-negative Matrix Factorization. Because of this preparation, semi-NMF may be employed in a far wider variety of applications than NMF [41, 42]. Additionally, our method offers many advantages over NMF-based MVC techniques, including the ability to analyze Semi-NMF in more depth [43, 44]. Data tests from a related class may be brought closer together using the Semi-NMF structure, as mentioned. This dish has a particular flavor since it is based on extensive education [45]. Even though our approach isn't exactly the same as the

current MVC auto-encoder-based solutions, we still have a major strategy (Andrew et al 2019; Wang et al 2020). When compared to previous research, we don't employ Canonical Correlation Analysis (CCA), restricted two cases. An Andrew et al 2020, as well as Wang et al, 2018. In this metaheuristic nature-based algorithm, an improved meta-heuristic methodology motivated by different researcher's studies are provided. One important model called Oppositional based Harris Hawk Optimizer-OHHO technique; it is an unsupervised algorithm. The following approach is constructed on the speculative Harris Hawk Optimize-HHO algorithm, which has no inherent dependent variables. The search space is later changed by integrating HHO with the Oppositional Based Learning-OBL method in order to provide better estimation for the dominant approach. A clustering strategy is also been discussed on unsupervised learning, known as OCMHAMVC [46]. The deep learning related scientific paper [47] has been taken as input files and applied various clustering techniques on it getting various clustered index values. The all literature survey section giving limitations of earlier studies related to document indexing. The survey which was analyzed has been taken as reference and proposed an advanced technology.

3. Preliminaries. This section explains fundamental preliminaries employed suggested strategy, along with the relevant procedures that are required.

3.1. Optimized Matrix Factorization (OMF):. This data is provided in $A = \{a_1, a_1, ...a_n\} \in M_+^{n\times d}$, where n indicates no. of various examples, and d indicates its dimension feature vector $a_j\,(1 \le q \le n)$ represented number of distinct samples in **(** $H = \{h_1, h_2, ..., h_d\} \in Q_+^d \& W = \{w_1, w_2, ..., w_d\} Q_+^d\,(k \prec\prec n \& d \prec D)$**).**

To find reduced rank matrices inside non-negative relationships, OMF investigates, i.e., which may be characterized $W = \{w_1, w_2, ..., w_d\}$ Input data examined as $a_j = \sum_{i-1}^{d} h_i w_{ji}$ stated using combination of linear matrix construction and impact factor wi after exploring matrix relations (i.e. H & W). Here is a breakdown of the purpose of non-negative matrix formation:

$$\min_{H,W} ||A - HW||_F^2 \;\; w.r.t \quad H \ge 0, W \ge 0 \tag{3.1}$$

$||||_F$ Fresenius standard procedure with variety functions
W and H are described as variables in the Karush-Kuhn-Tucker (KKT) ruling condition,

$$H_{il} = H_{il} \frac{\left(AW^T\right)_{il}}{\left(HWW^T\right)_{il}} \tag{3.2}$$

With associative parameters,

$$W_{ij} = W_{ij} \frac{\left(H^T A\right)_{ij}}{\left(W^T HW\right)_{ij}} \tag{3.3}$$

Factorization DL described as

$$\begin{gathered} W^m \simeq H_1^m W_1^m \\ H_1^m \simeq H_2^m W_2^m \\ ..... \\ H_{1-2}^m \simeq H_{l-1}^m W_{l-1}^m, \\ H_{l-1}^m \simeq H_l^m W_l^m \end{gathered} \tag{3.4}$$

where $H_1^m, H_2^m, ..., H_{1-1}^m, H_1^m \& W_1^m, W_2^m, ..., W_{1-1}^m, W_1^m$ Coefficient and basis matrices m-dimensionality referred matrices. As result of combining the two equations above,

$$\min_{W_l^m, H_l^m} ||A^m - H_l^m W_l^m W_{l-1}^m, ..., W_2^m W_l^m||_F^2 \;\; w.r.t \quad H_l^m \ge 0, W_l^m \ge 0 \tag{3.5}$$

It is defined as: based on numerous viewpoints with objective functionality

$$\min_{W_l^m, H_l^m} \sum_{m-1}^{M} ||A^m - H_l^m W_l^m W_{l-1}^m, ..., W_2^m W_l^m||_F^2 \;\; w.r.t \quad H_l^m \ge 0, W_l^m \ge 0 \tag{3.6}$$

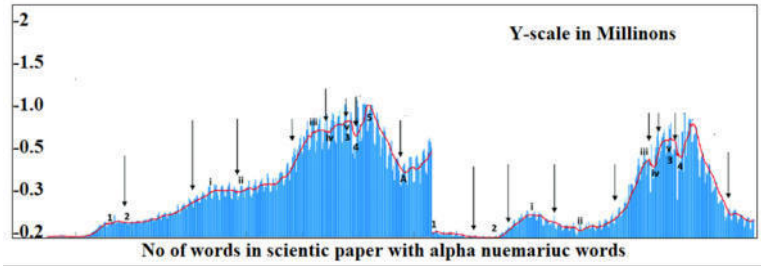


Fig. 4.1: Number of words scientific paper with unstructured scale

Objective functionality multiple attribute describes

$$\min_{H_l^m, W_l^m} ||A^m - H_l^m W_l^m W_{l-1}^m, ..., W_2^m W_l^m||_F^2 \;\; w.r.t \quad H_l^m \geq 0, W_l^m \geq 0 \tag{3.7}$$

3.2. In-depth Matrix Factorization Indices. Associative modularity's associative modularity may be used to design or investigate structure in complicated procedures. Deep learning is used to describe the process of discovering optimal matrix functions.

$$\begin{gathered} A \simeq Z_1 W_1 \\ A \simeq Z_1 Z_2 W_2 \\ ..... \\ A \simeq Z_1 Z_2 .... Z_m W_m \end{gathered} \tag{3.8}$$

$$Z_l \in Q^{K_{t-1} \times K_t} \;\; Be \quad l - th(l \leq m) \tag{3.9}$$

$$W_l \in \Re^{k_i \times n} \; (> 0) \tag{3.10}$$

Matrixes are clustered into clusters or factored out of a matrix with$Z_l, ..., (Z_1, Z_2, Z_t) \in \Re^{d \times k_i}$ m-dimensional layers by taking this connection into account, which may be done in a variety of ways. In this way, distinct data sets may be represented using the same group procedures, but from different viewpoints. For multi-labeled clustering numerous attribute relations, deep matrix factorization approach is appropriate. Our first findings lead propose new heuristic approach investigating multi-labeled data clustering enhanced matrix construction.

4. Proposed Method. A clustering strategy based on unsupervised learning, known as OCMHAMVC, is discussed in this section. It is associated restrictions related to orthogonal and combined frameworks with co-regularization, and this approach is referred to here. First define multi objective functions suggested and then examine optimum maximization strategy cluster multi labelled data. Lastly, efficient complexity analysis, computational analysis suggested approach. The scientific data is applied to proposed methodology interims of alpha numeric unstructured data.

Figure 4.1 clearly explains about scientific data analysis, millions of words are applied to proposed OCMHAMVC and get easy indexing. This analysis giving multi-view clustering with more accuracy and throughput. The multi objective function of an optimization issue with numerous objective functions is referred to as a multi-objective optimization problems. A multi-objective optimization problem might be described mathematically as

$$min_{x \epsilon X}(f_1(x), f_2(x), ..., f_k(x)) \tag{4.1}$$

4.1. Multi objective functions of OCMHAMVC.

$$\{A^u \in \Re^{m_u n}\}_{u=1}^{n_u} \tag{4.2}$$

For the sake of argument, let's say we have an efficient multi-labeled data collection, which is made up of n classes of samples each with a separate multi-label. Labeled constraints are represented in multiple perspectives as l samples, whereas unlabeled constraints are represented as l-1 samples.

$$X = \begin{pmatrix} C_{c*l} & 0 \\ 0 & I_{n-1} \end{pmatrix} \tag{4.3}$$

I^{th} and j^{th}-class attribute data C_{ij}=0 as a result, C_{ij} the data were given new labels, and $(n-1)(n-l)$ unlabeled sample l_n-1 was assigned to process of building identity matrix. If number samples unlabeled data exceeds predetermined threshold, then an identity matrix may be used to look for previously labelled constraint data. When a cluster is not identified, labelled matrix data generation is referred to as a "labelled matrix."

$$X = \begin{pmatrix} C_{c*l} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I_{n-l-1} \end{pmatrix} = \begin{pmatrix} C_{c*l} & 0 \\ 0 & I_{n-1} \end{pmatrix} \tag{4.4}$$

Define auxiliary matrix Z, which expresses dimensions using sample data and numerous goal functions for orthogonal matrix creation as follows.

$$MO_F = \sum_{u=1}^{n_u} \theta_u ||A^u - V^u (Z^u)^T X^T||_F^2 + \lambda \sum_{u=1}^{n_u} ||Fo(Z^u (Z^u)^T) - I||_F^2 + \sum_{u=1}^{n_u} \sum_{s=1}^{n_u} \frac{1}{2} \theta_{us} ||Z^u - Z^s||_F^2 \tag{4.5}$$

Explore low-dimensional feature representation executed data using dimension multi-objective function. Investigate representations of desired characteristics for each and every one of the dimensions. It is essential that the qualities across classes have an effective distinguishing factor, and that the scalability of all chosen features is the same. The introduction of orthogonal constraints in multi-labeled clustering ensures the desired feature representation is met. Associative clustering prototypes with various parametric notations are used in conjunction with low-dimensional feature representations under orthogonal constraint to effectively discriminate between classes and attributes. A joint constraint matrix is offered as a means to migrate the performance of orthogonal constraints. The padding of clustered information has been providing less congestion as well as getting fast dimensionality grouping. The following analysis helping to remove congestion and providing document analysis effectively.

$$F_{ij} \begin{cases} 1 & j = i \quad 1 \le j, i \le c \\ 0 & otherwise \quad 0 \end{cases} \tag{4.6}$$

As a result, the orthogonal constraint framework $\lambda \sum_{u=1}^{n_u} ||Fo(Z^u (Z^u)^T) - I||_F^2$

We can next determine whether or not an orthogonal constraint relation has controllability by looking at how o is represented in various notational systems. Under the clustering specification structure, representation many forms various dimensions include unity data.

4.2. Convex Feature Optimization. As a result objective non-convex functions based global minimum relations, variables examined applying them in conjunction. Using the most recent constraints, recalculate the optimization, such that one attribute relationship is linked to other attributes that are existent but are not changed in any way. To include non-negative matrix relations into Lagrange matrices, non-negative attribute relations with convex optimization are stated as

$$\begin{aligned} La\_r = {} & \textstyle\sum_{u=1}^{n_u} \theta_u ||A^u - V^u (Z^u)^T X^T||_F^2 + \lambda \sum_{u=1}^{n_u} ||Fo(Z^u (Z^u)^T) - I||_F^2 + \sum_{u=1}^{n_u} \sum_{s=1}^{n_u} \tfrac{1}{2} \theta_{us} ||Z^u - Z^s||_F^2 \\ & + \textstyle\sum_{u=1}^{n_u} tr(\beta^u (V^u)^T) + \sum_{u=1}^{n_u} tr(\alpha^u (Z^u)^T) \end{aligned} \tag{4.7}$$

Implemented multi-labeled characteristics are given the KKT treatment by way of non-negative matrix relations are very important in documents analysis as well as Lagrange matrices. The convex optimization methodologies have been briefly concentrating on attributes and providing accurate document analysis. The non-negative Lagrange matrix is providing computations very smoothly compared earlier matrixes

$$\left(A^u X Z^u - V^u (Z^u)^T X^T X Z^u\right)_{ij} v_{ij}^u = 0 \tag{4.8}$$

$$\begin{aligned}&\theta_u X^T (A^u)^T V^u - \theta_u X^T X Z^u (V^u)^T V^u - 2\lambda (Fo(Z^u (Z^u)^T) Z^u + 2\lambda Z^u - 4\lambda F(Z^u.Z^u.Z^u)\\ &+4\lambda F(Z^u.Z^u.Z^u) - \textstyle\sum_{s=1}^{n_u} \theta_{us} Z_{ji}^s Z_{ji}^u = 0\end{aligned} \tag{4.9}$$

Finally, the multi-objective function is said to as being achieved, Descriptor extraction & utilization has been involved in document clustering. The Word groups known as descriptors which are used to characterize, elements of a cluster. In general, documents clustering is viewed as a centralized procedure also web document clustering for customers of search engines is an example of Document clustering. Online & offline applications of document clustering could be distinguished, especially comparing to offline apps, performance issues typically limit online applications. Text clustering could be utilized for a variety of purposes, including gathering related documents (news, tweets, etc.), analyzing customer & employee feedback, as well as identifying significant implicit subjects in all scientific datasets.

$$v_{ji}^u \leftarrow v_{ji}^u \frac{(A^u X Z^u)_{ji}}{(V^u (Z^u)^T X^T X Z^u)_{ji}} \tag{4.10}$$

$$z_{ji}^u \leftarrow z_{ji}^u \frac{\left(\theta_u X^T (A^u)^T V^u + 2\lambda Z^u + 4\lambda F(Z^u.Z^u.Z^u) + \sum_{s=1}^{n_u} \theta_{us} Z^s\right)_{ji}}{\left(\theta_u X^T X Z (V^u)^u V^u + 2\lambda (Fo(Z^u (Z^u)^T)) Z^u + 4\lambda F(Z^u.Z^u.Z^u) + \sum_{s=1}^{n_s} \theta_{us} A^u\right)_{ji}} \tag{4.11}$$

Above, following description how objective algorithm created:

Algorithm for Multi Labeled Clustering

Input: multi labeled data set $\{A^1, A^2, ..., A^{n_u}\}$, No. of Clusters, No. of Samples, dissimilar variablesθ_v, θ_{vs}

1 Form constraint labeled matrix X
2 Form optimized constraint matrix F
3 For $u = n - 1$ then
  a. Normalization factors i.e. $A^u \left(||A^u(:, j)||^2\right)$;
  b. Update primary parameters V^u & Z^u plotted region [1, 0]
4 E-For
5 For u=1 n_u then
  a. execute matrix building based on no. of iterations < T
  b. Constraint V^u then update Z^u
  c. Constraint Z^u then update V^u
6 E-For
7 estimate indication low-dimensional data i.e. U^v=XZ^u
8 calculate final indication low-dimensional data $U^* = \frac{\sum_{u=1}^{n_u} U^u}{n_v}$

Output: final multi labeled cluster result

Convergence multi-dimensional data clustering using this approach.

4.3. Multi labeled Dimensional clustering. Our first cluster function is based on the similarity measure, and our average similarity weight measure is derived from documents in the same cluster. We then use this measure to create a multi-label clustering. Figure 4.2 depicts an architecture for exploring Multi labeled Dimensional clustering.

Multi-dimensional cluster results collection is shown in step-by-step detail in Figure 4.2. Weighted similarity was obtained using this method.

$$CS = \sum_{r=1}^{n} m_r \left[\frac{1}{m_t^2} \sum_{t_i, t_j S_r} S(t_i, t_j)\right] \tag{4.12}$$

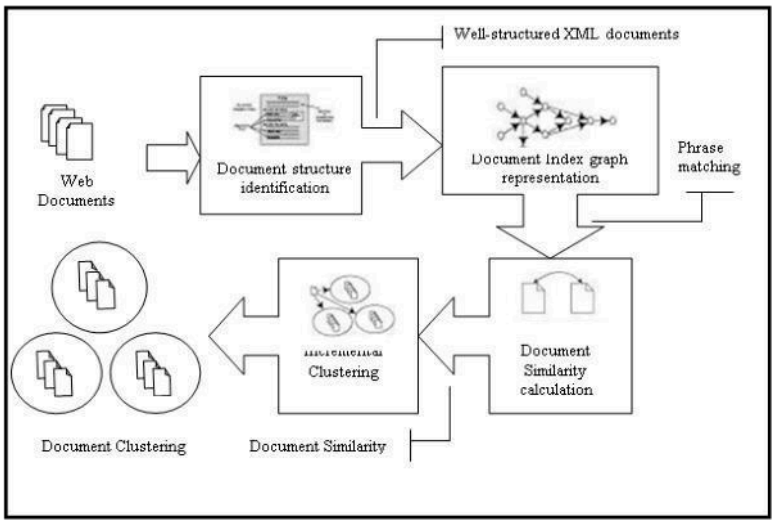


Fig. 4.2: Architecture proposed method

Class labels are n, and m is number of documents. D want to improve functionality so that we can locate functions that are comparable to

$$\begin{array}{l}\sum_{t_i,t_j\in S_r} S(t_i,t_j) = \sum_{t_i,t_j\in S_r} t_i^d, t_j - \frac{2m_r}{m-m_r}\sum_{t_i\in S_r} t_i^t \sum_{t_m\in S\setminus S_r} t_h + n_r^2 \\ = T_r^d T_r - \frac{2m_r}{m-m_r} T_r^d (T-T_r) + m_r^2 \\ \frac{m+m_r}{m-m_r}||T_r||^2 - \frac{2m_r}{m-m_r} T_r^d T + m_r^2\end{array} \tag{4.13}$$

Using weighted cluster functions to maximize the similarity of the document relations.

$$CF = \sum_{r=1} \frac{1}{m_r}\left[\frac{m+m_r}{m-m_r}||T_r||^2 - \left(\frac{m+m_r}{m-m_r} - 1\right) T_r^d T\right] \tag{4.14}$$

Comparing min-max functionality between words input documents to maximized weighted cluster functions, it assesses the depending on the maximum weight of the attribute. Incorporating cluster-related documents' functionality is represented as

$$\overline{CF} = \sum_{r=1}^{k} \frac{\lambda_r}{m_r}\left[\frac{m+m_r}{m-m_r}||T_r||^2 - \left(\frac{m+m_r}{m-m_r} - 1\right) T_r^d T\right] \tag{4.15}$$

Weighted functionality (Y) related document clustering rated most efficient.

$$Y = \sum_{r=1}^{k} \sum_{t_i\in S_r} \frac{1}{m-m_r} \sum_{t_h\in S\setminus S_r} S\left(t_i - t_h, \frac{C_r}{||C_r||} - t_h\right) \tag{4.16}$$

Discuss optimization multi-labeled clustering weighted cluster functions based similarity underlying cluster creation numerous multi-labeled capabilities.

$$MLC_{opt} = \sum_{r=1}^{k} I_r(m_r T_r) \tag{4.17}$$

Create optimum clusters outcome convergence clusters matrix relation measure depending number of iterations applied various input label data during multi-labeled clustering distinct sources. Because of the client base's

exponential development has been imported, it is important to identify the users who contribute the most useful data. Although the relevance of individual participants in the recommender systems could increase the recommender system's robustness & suggestion efficiency, there has n't been much study done along this line to find a more efficient approach. They suggest an approach with multi-label clustering to identify the core user as well as establish the idea of correlation among user & label cluster in order to address this issue. The following issue has been solved with this proposed methodology and results are giving proofs.

5. Experimental Evaluation of OCMHAMVC. A comparison of OCMHAMVC's performance with that of standard techniques is shown in this section. Experimental data, OCMHAMVC used measure different weighted cluster functions clustering multi-labeled document clusters, measure clustering functions work based on Euclidean similar distance, similar cosine and relative jacquard co-efficient similar measurement.

The clustering data assortment is an advanced version data analysis, many earlier models has been unable provided. The proposed method has been optimizing the infromation as wells providing deep extraction of data on scientific document. The understandability and transferability of optimized document analysis was providing deep information about scientific document and giving good performance measures.

The cluster prototypes are not that much efficient but proposed model has been getting many rules from algorithm and giving solution to padding problems. The cooperative solution and regularization have been called through clustering steps.

5.1. Input Clustering data. We employed real-time benchmark data before we used Reuter's 08-10 versions of k1b for clustering of documentation, as well as additional standard datasets from efficient & exhaustive data sources, in this experiment for multi-labeled document clustering. Clustering applications may be conducted in real time using downloaded data sets and cloud-related data sources that have measurable similarity. For the most part we utilize the BBC Series data set, Reuter's datasets, Series 3 sources, and MSRC dataset (http://mlg.ie/datasets/3sources.html, http://lig-membres.imag.fr/grimal/data.html, http://www-vision-cal tech-edu/Image Datasets /Caltech101.html, https://pgram-com/dataset/msrc-v1/) (entertainment, politics, and sports, medical and business related applications). It was found that the suggested technique outperformed other multi-dimensional clustering methods when evaluated on all of the datasets mentioned above. We have GMNMF (multi view non-negative matrix factorization) , CoNMF-P [3], and MVCC [4], all of which use non-negative comments as a basis for their non-negative matrix factorization results.

The accuracy, NF, lacquard coefficient, precision, recall, F1 measure, presentation computing cost and memory consumption are performance measures which are deciding the application stability and comparing earlier models. The proposed methodology attains more improvement and suitable for traditional document analysis algorithm.

5.2. Setting of Experiments. Samples from numerous data sources are randomly gathered, according to the authors' original publications, and the tagged data is eliminated. Search parameter weights with various notations are then applied to each sample in order to compare traditional techniques with the novel approach. With the use of multi-labeled dimensional clustering and Euclidean distance metrics, we conduct our research. Using these metrics, evaluate accuracy, NF, lacquard coefficient, precision, recall, F-score, presentation computing cost & memory consumption each cluster dataset proposed to approach. Document retrieval accuracy multiple labels proposed method shown in Figure 5.1 in comparison other well-established strategies.

The confusion matrix is used to define measures like accuracy, sensitivity, recall and precision. The measures can be decided by collection of true positive, true negative rate, false positive rate and false negative rate via statistical information. The clustering process is performed through many ways but in our research class based analysis were performed. The selected scientific document is applied to our designed application it can differentiated the information into multi class variance, by using following analysis document has been clustered.

On a sample of 100-500 html text documents, five alternative clustering algorithms are shown in Figure 5.1, multi-labeled data collection showing best accuracy. Clustering results displayed variety ways depending on the accuracy values provided in table 5.1, every dataset contains row, best value shown boldly & remaining values second best results various algorithms. Table 5.2 shows the recall values of many suggested techniques with consistent multi-labeled clustering results. This is a good sign. Figure 5.2 depicts recall performance various multi-labeled html text content variations (left to right).

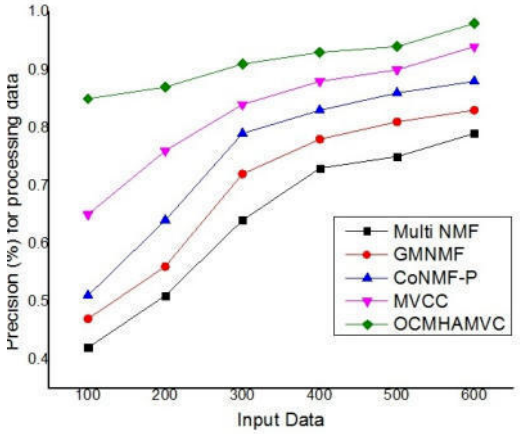


Fig. 5.1: Multi class label parameters performance

Table 5.1: MDC Result

Data input	Multi-NMF	GMNMF	Co-NMFP	MVCC	Proposed OCMHAMVC
100	00.43	00.480	052	0.66	0.86
200	0.52	0.57	0.65	0.77	0.88
300	0.65	0.73	0.80	0.86	0.92
400	0.74	0.79	0.84	0.90	0.94
500	0.76	0.82	0.87	0.91	0.95
600	0.80	0.84	0.89	0.95	0.99

Table 5.2: Recall MDC values

Input data	Multi-NMF	GMNMF	Co-NMFP	MVCC	Proposed OCMHAMVC
100	00.590	00.68	00.76	00.85	00.89
200	00.65	00.72	00.79	00.88	00.92
300	00.705	00.79	00.85	00.90	00.93
400	00.78	00.84	00.88	00.91	00.95
500	00.84	00.86	0.91	00.94	00.96
600	00.86	00.91	00.93	00.95	00.98

Multi-NMF, GMNMF, Co-NMFP, MVCC and OCMHAMVC methods are implemented on python 3.7.0 and it is identified that proposed OCMHAMVC was attains good improvement and suitable for future document summery applications.

It is demonstrated in figure 5.3 that OCMHAMVC delivers the best results when compared to conventional ways when the number of documents is increased; when number of attributes is increased, OCMHAMVC exhibits competent clustering results with multi attributes. The figure 5.3 clearly explains about MDC analysis using F1 score via data input. Here generating measures with multi labelled data gathering. The html text document is collected from normal text which is performed through proposed model. The generated statistical values are proven that proposed model is good at document summarisation applications.

With a strong preference for multi-labeled data gathering, five alternative clustering algorithms were performed to 100-500 html text documents in Figure 5.4.

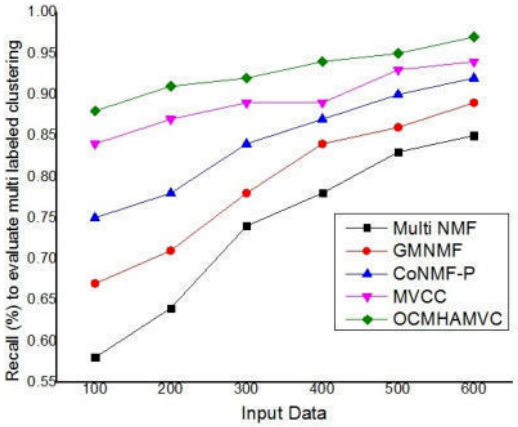


Fig. 5.2: MDC performance in recall

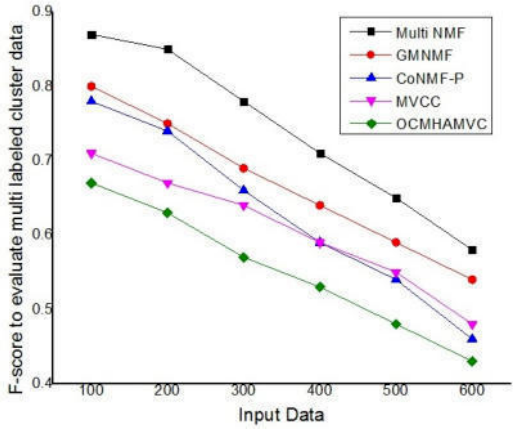


Fig. 5.3: MDC in F1- Score performance

Table 5.3: MDC's Values in F1- Score

Data input	Multi-NMF	GMNMF	Co-NMFP	MVCC	Proposed OCMHAMVC
100	00.88	00.81	00.75	00.72	00.68
200	00.86	00.76	00.76	00.68	00.64
300	00.79	00.70	00.67	00.64	00.58
400	00.72	00.65	00.60	00.60	00.54
500	00.66	00.60	00.55	00.56	00.49
600	00.60	00.55	00.47	00.50	00.44

For every dataset, the finest value is indicated in bold, while all of the other values are second-best findings from other methodologies.

OCMHAMVC shown figure 5.5 best time results when compared traditional approaches retrieving matched multi-dimensional documents html text documents. It takes less time to get multi-attribute relational documents from multiple domains using OCMHAMVC as the number of documents increases.

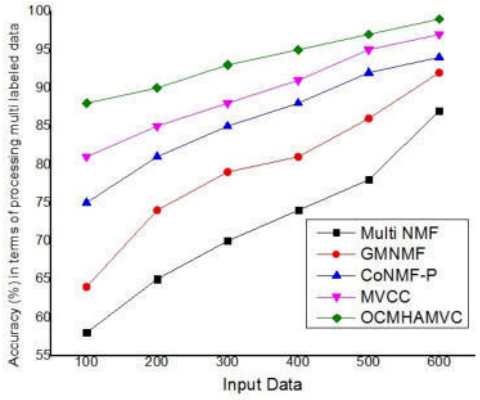


Fig. 5.4: Multi Dimension clustering Accuracy

Table 5.4: MDC's Accuracy

Input data	Multi-NMF	GMNMF	Co-NMFP	MVCC	Proposed OCMHAMVC
100	59	65	76	82	89
200	66	75	82	86	91
300	71	80	86	89	94
400	75	82	89	92	96
500	79	87	93	96	98
600	88	93	95	98	100

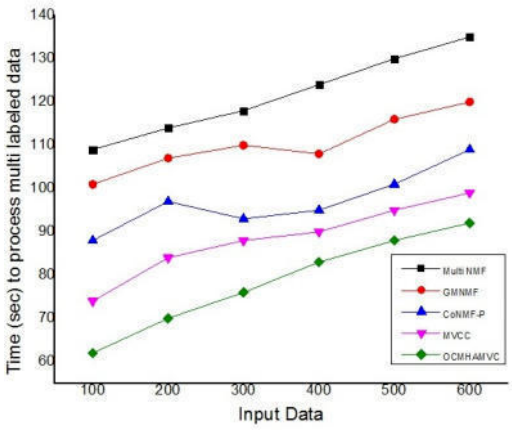


Fig. 5.5: MDC performance in time

Data set, best value indicated bold, other values second-best findings methodologies.

Data from 100-500 html text articles were used to test five alternative clustering algorithms, and the results shown in Figure 5.6 demonstrate that multi-labeled data collection is the most efficient in terms of computing cost when combined with the most desirable parameters.

For each data set in table 5.6, best value indicated bold and others second best findings from various methodologies.

Table 5.5: Multi Dimension Clustering in Time Values

Input data	Multi-NMF	GMNMF	Co-NMFP	MVCC	Proposed OCMHAMVC
100	110	102	89	75	63
200	115	108	98	85	71
300	119	112	94	89	77
400	125	109	96	92	84
500	131	115	102	96	90
600	136	122	110	100	93

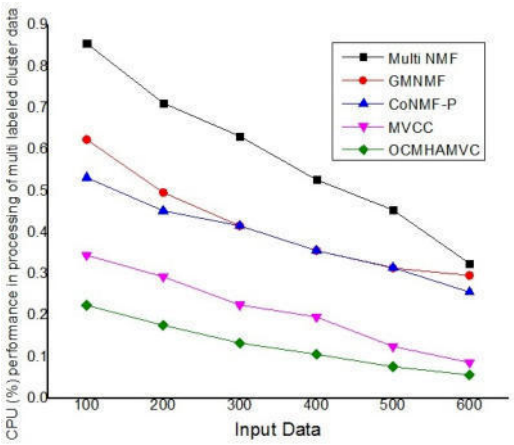


Fig. 5.6: Computational Cost Performance in CPU Processing

Table 5.6: Multi Dimension Clustering in Computational Cost Performance Values

Data input	Multi-NMF	GMNMF	Co-NMFP	MVCC	Proposed OCMHAMVC
100	00.857	00.625	00.533	00.346	00.226
200	00.713	00.497	00.454	00.294	00.177
300	00.633	00.417	00.417	00.226	00.134
400	00.528	00.357	00.357	00.197	00.107
500	00.454	00.315	00.316	00.127	00.077
600	00.326	00.297	00.258	00.088	00.058

Table 5.6 clearly explained about OCMHAMVC score and comparison has been performed with earlier models. In this context for various input like 100, 200.....600 at any instant proposed model got good improvement since 00.226 to 00.058.

OCMHAMVC has the best time results compared to traditional approaches when it comes to retrieving multi-attribute relational documents from html text documents, as shown in figure 5.7. When the number of documents is increased, OCMHAMVC provides efficient cluster results, which means less memory utilization. Shown table 5.7 depicts MDC values based on F-score enhance number of labelled text documents

Results from the figures and tables above are based on the experimental setup for the suggested technique and alternative approaches. Because of OCMHAMVC's fundamental convergence, it may be used on text-oriented documents several domains and yet satisfy text data connected to curves. OCMHAMVC additionally finds effective multi-dimensional clustering results decreasing iteration used on distinct text-oriented documents

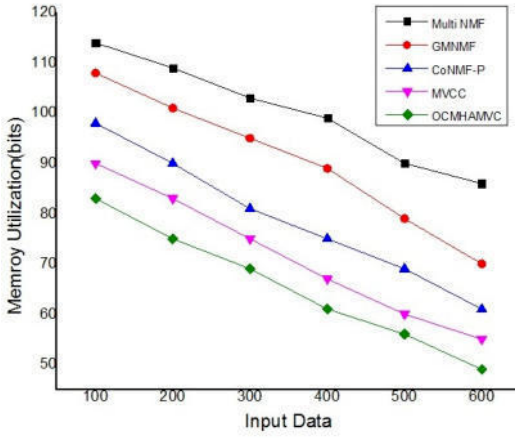


Fig. 5.7: Memory usage in procedure MLD performance

Table 5.7: Memory usage in MDC

Data input	Multi-NMF	GMNMF	Co-NMFP	MVCC	Proposed OCMHAMVC
100	115	109	99	91	84
200	110	102	92	84	76
300	104	97	82	76	70
400	100	90	76	68	62
500	92	80	70	62	57
600	8s7	72	62	56	50

while investigating documents using iterative functions. OCMHAMVC measures computational efficiency multi-dimensional document processing time, precision, recall, accuracy, memory use and CPU computational cost.

The OCMHAMVC model has getting computational efficiency, recall, accuracy, memory usage and CPU computational cost from confusion matrix and measures estimation on statistical data from scientific documents. The cluster analysis gives documents summarisation with easy analysis.

6. Conclusion. A new strategy representing data cluster of multiple categories using multi-labeled dimensional data is proposed in this study, namely OCMHAMVC. The suggested technique initially clusters the comparable label data that are important to similar cluster prototypes, given the fast evolution of labelled data. This cluster prototype has the same labels and attributes associated with the same classes as the original cluster. In order to gather comparable cluster prototypes, the suggested technique utilizes cooperative regularization with representative qualities to investigate associative aspects that are desirable from several perspectives. The OMF technique is used to assess low-dimensional data, and a prototype cluster of labelled sample data is formed by grouping related data into the OMF method's OMF evaluations. The proposed strategy is tested on a variety of real-time data sets, and it demonstrates how it compares well to currently available methods. Sampled class labelled cluster findings boundary relations compared multi-labeled dimensions to see how well they perform. Advanced machine learning multi-dimensional clustering would be a great addition to further enhance our suggested strategy.

REFERENCES

[1] Nguyen, D. T., Chen, L., & Chan, C. K. , *Clustering with multiviewpoint-based similarity measure*, IEEE transactions on knowledge and data engineering, 2011, 24 (6), pp. 988-1001.

[2] Liu, J., Wang, C., Gao, J., & Han, J. , *Multi-view clustering via joint nonnegative matrix factorization*, In Proceedings of the 2013 SIAM international conference on data mining Society for Industrial and Applied Mathematics, 2013, pp. 252-260.
[3] Wang, Z., Kong, X., Fu, H., Li, M., & Zhang, Y. , *Feature extraction via multi-view non-negative matrix factorization with local graph regularization*, In 2015 IEEE International conference on image processing (ICIP), 2015, pp. 3500-3504.
[4] He, X., Kan, M. Y., Xie, P., & Chen, X. , *Comment-based multi-view clustering of web 2.0 items*, In Proceedings of the 23rd international conference on World wide web, 2014, pp. 771-782.
[5] Wang, H., Yang, Y., & Li, T. , *Multi-view clustering via concept factorization with local manifold regularization*, In 2016 IEEE 16th International Conference on Data Mining (ICDM), 2016, pp. 1245-1250.
[6] Wei, S., Wang, J., Yu, G., Domeniconi, C., & Zhang, X. , *Multi-view multiple clusterings using deep matrix factorization*, In Proceedings of the AAAI conference on artificial intelligence, 2020, Vol. 34, No. 04, pp. 6348-6355.
[7] Li, Z., Tang, J., & Mei, T. , *Deep collaborative embedding for social image understanding*, IEEE transactions on pattern analysis and machine intelligence, 2018, 41(9), pp. 2070-2083.
[8] Liu, X., Zhu, X., Li, M., Wang, L., Zhu, E., Liu, T., ... & Gao, W. , *Multiple kernel k k-means with incomplete kernels*, IEEE transactions on pattern analysis and machine intelligence, 42 (5), pp.1191-1204.
[9] Luo, S., Zhang, C., Zhang, W., & Cao, X. , *Consistent and specific multi-view subspace clustering*, In Proceedings of the AAAI Conference on Artificial Intelligence, 2018, Vol. 32, No. 1.
[10] Mautz, D., Ye, W., Plant, C., & Böhm, C. , *Discovering non-redundant k-means clusterings in optimal subspaces*, In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, 2018, pp. 1973-1982.
[11] Nie, F., Cai, G., & Li, X. , *Multi-view clustering and semi-supervised classification with adaptive neighbours*, In Thirty-first AAAI conference on artificial intelligence, 2017.
[12] Tan, Q., Yu, G., Domeniconi, C., Wang, J., & Zhang, Z. , *Incomplete multi-view weak-label learning*, In Ijcai, 2018, pp. 2703-2709.
[13] Wang, X., Yu, G., Domeniconi, C., Wang, J., Yu, Z., & Zhang, Z. , *Multiple co-clusterings*, In 2018 IEEE International Conference on Data Mining (ICDM), 2018, pp. 1308-1313.
[14] Wang, X., Wang, J., Domeniconi, C., Yu, G., Xiao, G., & Guo, M. , *Multiple independent subspace clusterings*, In Proceedings of the AAAI conference on artificial intelligence, 2019, Vol. 33, No. 01, pp. 5353-5360.
[15] Yang, S., & Zhang, L. , *Non-redundant multiple clustering by nonnegative matrix factorization*, Machine Learning, 2017, 106(5), pp. 695-712.
[16] Yao, S., Yu, G., Wang, X., Wang, J., Domeniconi, C., & Guo, M. , *Discovering multiple co-clusterings in subspaces*, In Proceedings of the 2019 SIAM International Conference on Data Mining Society for Industrial and Applied Mathematics, 2019, pp. 423-431.
[17] Yao, S., Yu, G., Wang, J., Domeniconi, C., & Zhang, X., *Multi-view multiple clustering*, arXiv preprint arXiv:1905.05053. 2019.
[18] Zhao, H., Ding, Z., & Fu, Y., *Multi-view clustering via deep matrix factorization*, In Thirty-first AAAI conference on artificial intelligence, 2017.
[19] Zong, L., Zhang, X., Zhao, L., Yu, H., & Zhao, Q., *Multi-view clustering via multi-manifold regularized non-negative matrix factorization*, Neural Networks, 2017, 88, pp. 74-89.
[20] Li, S., Liu, Q., Dai, J., Wang, W., Gui, X., & Yi, Y., *Adaptive-weighted multiview deep basis matrix factorization for multimedia data analysis*, Wireless Communications and Mobile Computing, 2021.
[21] Ahn, J. H., Kim, S., Oh, J. H., & Choi, S., *Multiple nonnegative-matrix factorization of dynamic PET images*, In Proceedings of Asian Conference on Computer Vision, 2004, pp. 1009-1013.
[22] Trigeorgis, G., Bousmalis, K., Zafeiriou, S., & Schuller, B. W. , *A deep matrix factorization method for learning attribute representations*, IEEE transactions on pattern analysis and machine intelligence, 2016, 39(3), pp. 417-429.
[23] Zhao, Y., Wang, H., & Pei, J., *Deep non-negative matrix factorization architecture based on underlying basis images learning*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(6), pp. 1897-1913.
[24] Meng, Y., Shang, R., Shang, F., Jiao, L., Yang, S., & Stolkin, R. , *Semi-supervised graph regularized deep NMF with bi-orthogonal constraints for data representation*, IEEE transactions on neural networks and learning systems, 2019, 31(9), pp. 3245-3258.
[25] Tong, M., Chen, Y., Ma, L., Bai, H., & Yue, X. (2020) , *NMF with local constraint and Deep NMF with temporal dependencies constraint for action recognition*, Neural Computing and Applications, 2020, 32(9), pp. 4481-4505.
[26] J Li, J., Zhou, G., Qiu, Y., Wang, Y., Zhang, Y., & Xie, S. , *Deep graph regularized non-negative matrix factorization for multi-view clustering*, Neurocomputing, 2020, 390, pp. 108-116.
[27] Shu, Z. Q., Wu, X. J., Hu, C., You, C. Z., & Fan, H. H. , *Deep semi-nonnegative matrix factorization with elastic preserving for data representation*, Multimedia Tools and Applications, 2021, 80(2), pp. 1707-1724.
[28] Zhao, J., Xie, X., Xu, X., & Sun, S., *Multi-view learning overview: Recent progress and new challenges*, Information Fusion, 2017, 38, pp. 43-54.
[29] Li, Y., Yang, M., & Zhang, Z. , *A survey of multi-view representation learning*, IEEE transactions on knowledge and data engineering, 2018, 31(10), pp. 1863-1883.
[30] Hussain, T., Muhammad, K., Ding, W., Lloret, J., Baik, S. W., & de Albuquerque, V. H. C. (2021) , *A comprehensive survey of multi-view video summarization*, Pattern Recognition, 2021, 109, 107567.
[31] Xu, X., Yang, Y., Deng, C., & Nie, F. , *Adaptive graph weighting for multi-view dimensionality reduction*, Signal Processing, 2019, 165, pp. 186-196.
[32] Zhang, R., Nie, F., Li, X., & Wei, X. (2019) , *Feature selection with multi-view data: A survey*, Information Fusion, 50,

pp. 158-167.
[33] Luo, P., Peng, J., Guan, Z., & Fan, J. (2018) , *Dual regularized multi-view non-negative matrix factorization for clustering*, Neurocomputing, 2018, 294, pp. 1-11.
[34] Bai, L., Shao, Y. H., Wang, Z.,& Li, C. N., *Clustering by twin support vector machine and least square twin support vector classifier with uniform output coding*, Knowledge-Based Systems, 2019, 163, pp. 227-240.
[35] H Seung, H. S., & Lee, D. D., *The manifold ways of perception*, science, 2020, 290 (5500), pp. 2268-2269.
[36] Spielman, D. A. (2007, October) , *Spectral graph theory and its applications*, In 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), 2017, pp. 29-38.
[37] Hidru, D., & Goldenberg, A., *EquiNMF: Graph regularized multiview nonnegative matrix factorization*, arXiv preprint arXiv:1409.4018. 2014.
[38] Zong, L., Zhang, X., Zhao, L., Yu, H., & Zhao, Q., *Multi-view clustering via multi-manifold regularized non-negative matrix factorization*, Neural Networks, 2017, 88, pp.74-89.
[39] Guan, Z., Zhang, L., Peng, J., & Fan, J. , *Multi-view concept learning for data representation*, IEEE Transactions on Knowledge and Data Engineering, 2015, 27(11), pp. 3016-3028.
[40] Xu, C., Guan, Z., Zhao, W., Niu, Y., Wang, Q., & Wang, Z. , *Deep multi-view concept learning*, In IJCAI, 2018, pp. 2898-2904.
[41] Wang, J., Wang, X., Tian, F., Liu, C. H., Yu, H., & Liu, Y. , *Adaptive multi-view semi-supervised nonnegative matrix factorization*, In International conference on neural information processing. Springer, Cham, 2016, pp. 435-444.
[42] Kolli, S., & Sreedevi, M., *A novel index based procedure to explore similar attribute similarity in uncertain categorical data*, ARPN Journal of Engineering and Applied Sciences, 2019, 14(12), pp. 2266-2272.
[43] Kolli, S., & Sreedevi, M., *Adaptive clustering approach to handle multi similarity index for uncertain categorical data streams*, Jour of Adv Research in Dynamical & Control Systems, 2018, 10(04).
[44] Kolli, S. , *A novel granularity optimal feature selection based on multi-variant clustering for high dimensional data*, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021, 12(3), pp.5051-5062.
[45] Kolli, S., Krishna, P. R., & Reddy, P. B. , *A NOVEL NLP AND MACHINE LEARNING BASED TEXT EXTRACTION APPROACH FROM ONLINE NEWS FEED*, 2006.
[46] Gudadappanavar, S. S., & Mahapatra, S. , *Metaheuristic nature-based algorithm for optimal reactive power planning*,International Journal of System Assurance Engineering and Management, 2022, 13(3), pp.1453-1466.
[47] Saikumar, K., Rajesh, V., & Babu, B.S., *Heart disease detection based on feature fusion technique with augmented classification using deep learning technology*, Traitement du Signal, 2022, 39(1), pp.31-42.

Edited by: Vinoth Kumar
Received: Jun 30, 2022
Accepted: Dec 5, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 291–301, DOI 10.12694/scpe.v23i4.2031

UNSUPERVISED UNMIXING AND SEGMENTATION OF HYPER SPECTRAL IMAGES ACCOUNTING FOR SOIL FERTILITY

K LAVANYA*, R JAYA SUBALAKSHMI, T TAMIZHARASI, LYDIA JANE, AND AKILA VICTOR

Abstract. A crucial component of precision agriculture is the capability to assess the fertility of soil by looking at the precise distribution and composition of its different constituents. This study aims to investigate how different machine learning models may be used to assess soil fertility using hyperspectral pictures. The development of images using a random mixing of different soil components is the first phase, and the hyper spectral bands utilized to create the images are not used again during the analysis procedure. The resulting end members are then acquired by applying the NFINDR algorithm to the process of spectral unmixing this image. The comparison between these end members and the band values of the known elements is then quantified., i.e. it is represented as a graph of band values obtained through spectral unmixing. Finally we quantify the similarities between both graphs and proceed towards the classification of the hyper spectral image as fertile or infertile. In order to classify the hyper spectral image as fertile or infertile, we quantify the similarities between the two graphs. Clustering and picture segmentation algorithms have been devised to help with this process, and a comparison is then made to show which techniques are the most effective.

Key words: Hyper spectral imaging, Spectral unmixing, NFINDR Soil fertility, Machine Learning.

AMS subject classifications. 68T05

1. Introduction.

Precision Agriculture refers to a farm management concept involving responding, measuring and observing various field features in an attempt to optimize farming techniques.[1] Also known as satellite farming, the goal of precision agriculture is to maximize field productivity while minimizing utilization of resources such as fertilizers. Recently, precision agriculture has better adopted a spot under the limelight with the development of new technologies under the ambit of GPS, GPRS, satellite imaging and machine learning, however there is an inherent gap in its application in the Indian subcontinent for a number of reasons including but not limited to lack of data, motivation to optimize and implement modern farming techniques, and a lack of infrastructure to develop the same on a large scale. Hyper-Spectral Imaging (or HSI) refers to an emerging concept in satellite imaging focusing on analyzing the wider spectrum of light over just the typically analyzed RGB wavelengths. Hence, it involves breaking down each pixel into various spectral bands and provide more extensive data through the captured image. While it has previously been majorly used in the military sector, with the growing availability of hyper-spectral data it has found its way to agriculture although not as much in the Indian subcontinent. Thus, by combining the concepts of precision agriculture and hyper-spectral imaging, a new level of analysis for soil fertility may be approached. With optimized analysis it has immense potential in the Indian subcontinent to aid in implementing modern farming techniques to maximize yields and profits for farmers [2].

With the advent of technological advances in various sectors, it can be crucial to maintain the relevant standards of modernity in fields such as agriculture particularly in the Indian subcontinent due to the prevalence of the same. One of the primary problems faced by farmers can be efficiently analyzing large fields and lands for type of soil and fertility of the soil in order to decide on various factors such as the purchase of fertilizers, types of crops that may be planted, and increasing productivity from potentially fertile lands [3, 4]. Specifically, for larger farms this can be a challenge due to the diverse soil types and landscapes present in the Indian subcontinent and require technology to help overcome the same. Hence, there is a need for a comprehensive and accurate technological tool to help optimize and maximize both crop yields and profits for farmers.

*School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India (lavanya.k@vit.ac.in).

2. Key Concepts.

2.1. Hyperspectral Imaging. Hyperspectral images are those in which each pixel is measured in one continuous spectrum. The spectral resolution is depicted in wave-numbers or nanometers. Spectral resolution can be defined as the interval between the different wavelengths that are measured in a specific range of wavelengths. The more the bands (spectral channels) the higher the spectral resolution. Hyperspectral Imaging is a technique that analyzes a wide spectrum of light, instead of just assigning colors like red, green or blue to each pixel [5].

2.2. Spectral Unmixing. Spectral Unmixing is a technique that has been used to analyze the mixture of components in remotely-sensed images for over 25 years. The most widely used method employs the use of a single set of endmembers (3-4) on the entirety of the image and then using a constrained least squares method to perform a linear unmixing. However, the variety of spectral unmixing techniques continues to grow, with most techniques being specific for its field of application.

2.3. NFINDR Algorithm. NFINDR algorithm is a technique that has its basis on the fact that in X spectral dimensions, the X-dimensional volume that is formed by a simplex (generalized notion of a triangle) with its vertices specified by the purest pixels, will always be larger than those formed by any other combination of pixels [6].

2.4. Cosine Similarity. Cosine similarity is a metric that is often used to determine the similarity between two particular objects (typically documents or samples). From a mathematical viewpoint, it looks to measure the cosine of the angle that exists between two vectors that are projected in a multidimensional space. Larger the angle, lesser the similarity.

2.5. pH Index. pH Index refers to the acidity or basicity of the soil sample in the hyperspectral image. It was a measure developed specifically for image processing in agriculture using the RGB values of each pixel present in the image. Different ranges of pH indexes indicate different ranges of pH, normally ranging from 5.58 to 7.50 pH.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|\|A\|\| \, \|\|B\|\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

3. Data set description. With the continuously growing versatility in Indian farmlands, it became evident that attempting to create a data set by collecting physical soil samples and comparing their fertility with the results obtained from the proposed model based on hyperspectral images would prove to be difficult. Thus, we decided to create our own data set. We focused on initially producing images as a mixture of various soil components ensure a minimum of ten different nutrients at a time. This is a blind creation, wherein the bands used in creating the image aren't used further in the analysis process. These images are assembled together and left unclassified thus leaving a scope for unsupervised classification later.

The research conducted is based on two main objectives. The first is to create a data set by producing hyperspectral images as a mixture of various soil components. In order to achieve this, the initial phase of the research focused on understanding the breakdown of soil components, and the basic requirements required in order for a particular sample of soil to be determined/classified as fertile or unfertile. The soil components, their determining features, their natural forms and other important characteristics were obtained from the USGS Hyperspectral Library and the results of these research were tabulated extensively.

The second phase of research focused on determining various classification and segmentation techniques that could be used to classify a given hyperspectral image as fertile or unfertile. The development of models based on unsupervised classification was chosen for the ambit of this paper. The extent of their results, due comparisons drawn and other details are further elaborated upon in the next section of this paper.

4. Methodology.

4.1. Generation of Hyperspectral Images. The process flow is initiated with the creation of the hyperspectral images. With the limitations faced due to the unavailability of large public datasets of hyperspectral images (HSI), there arose a necessity to manually produce a hyperspectral image dataset, which can ideally

simulate the required satellite images. With traditional HSI rarely procured in a pure form, and in order to accurately simulate the satellite images, external noise was factored during the production of the images for the dataset as well [7].

Beginning with MATLAB as the software of choice, a feature was employed that uses Gaussian fields to produce hyperspectral images. The bands that are wished to be included as a part of the image are inputted, and the function creates it as a Gaussian Random Field. Noise is then added to the Gaussian random distribution to mimic the disturbances that exist in remotely sensed satellite based hyperspectral images.

4.2. Spectral Unmixing using NFINDR Algorithm. Once these images are produced, the subsequent dataset created imitates one that would have been otherwise physically accumulated. The next stage involved includes the execution of spectral unmixing by employing the NFINDR algorithm. Spectral unmixing is carried out to breakdown a spectrum of mixed pixels into a set of its constituent spectra (also known as endmembers), along with a corresponding set of abundances (fractions that indicate the proportion of endmembers).

The NFINDR algorithm is basically an automated technique that is used to find the purest pixels present in an image. The main objective of this algorithm is to duplicate the successful technique (non-automated) of pinpointing the extreme points of an n dimensional scatter plot. The convex nature of existing hyperspectral data allows the NFINDR technique to be performed in relatively quick and straightforward method.

In the proposed process flow, an inbuilt function of MATLAB has been employed:

$$\text{endmembers} = \text{nfindr (inputData, numEndmembers, Name, Value)}$$

This function extracts the endmember signatures from hyperspectral data, by using the NFINDR algorithm. numEndmembers represents the number of endmember signatures that are to be extracted using the NFINDR algorithm. This syntax is used when the options for the number of iterations along with dimensionality reduction is required. The endmembers obtained have certain wavelengths that corresponds to specific components of soil. These endmembers can thus be used to classify different sections of soil in the hyperspectral image as fertile or unfertile [8].

4.3. NFINDR Algorithm.

1. Compute principal component bands and reduce spectral dimensionality of input data. Set number of PC bands to be extracted equal to number of endmembers to be extracted.
2. Randomly choose n number of pixel spectra from the reduced data as an initial set of endmembers.
3. Begin iteration 1, denote initial set of endmembers as compute volume using

$$V(E^{(1)}) = |\det(E^{(1)})|$$

where

$$E^{(1)} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e_1^{(1)} & e_2^{(1)} & \cdots & e_1^{(1)} \end{bmatrix}$$

4. For the second iteration, select a new pixel spectra r, such that: $r \notin \{e_1^{(1)}, e_2^{(1)}, \ldots, e_p^{(1)}\}$
5. Replace each endmember in the set with r and then compute the volume of the resulting simplex V(E(2)).
6. Replace the ith endmember in the set with r, and if the computed volume V(E(2)) is greater than V(E(1)). Thus arises an updated set of endmembers.
7. For each following iteration, select a new pixel spectra r and repeat the 5th and 6th step. The iterations end when the total number of iterations has reached the specified value.

4.4. Cosine Similarity based Accuracy prediction. Once the results of the spectral unmixing have been obtained, they become the second vector that will be used when determining the cosine similarity. The first vector used will be the plot of the band values that were originally used to create the hyperspectral images. Cosine similarity is used in order to determine the accuracy of the spectral unmixing and to identify if the endmember wavelengths obtained can be matched up to the wavelengths used in the production of the hyperspectral images in the first place.

4.5. Unsupervised Classification as Fertile/Infertile. In order to classify a particular hyperspectral image of a soil sample as fertile or infertile, numerous classification and segmentation algorithms were employed. The results drawn from these algorithms were compared, and subsequently K-Means Clustering and Agglomerative Clustering amongst the clustering techniques were found to display the best results while the neural network algorithms of ENet and DeeplavV3 provided highest reliability for classification, and FMM and Felzenszwalb algorithms provided the best results for pH index [9, 10].

4.5.1. K-Means Clustering. The objective of K-Means takes similar data points, and are grouped together to find subsequent underlying patters. These collections of similar data points are referred to as a cluster.

A target number 'k' is defined, which indicates the number of centroids (real or imaginary location that represents the center of the cluster) that are needed in the dataset. Every data point is then assigned to each of the clusters by reducing the sum-of-squares (i.e. the algorithm identifies k number of centroid and then goes on to assign every data point to its nearest cluster, while ensuring to keep the centroids as small as accurately possible). During this process, each point refers to the results obtained from the NFINDR spectral unmixing process i.e. the spectral band values of a particular component of the original hyper spectral image. Using clustering with two centroids, the algorithm aims to separate the spectral bands as fertile or infertile, unsupervised.

The classification of the bands as fertile or infertile is then conducted by analyzing the centroid values, with the assumption that higher band values correspond to more fertile compounds, an assumption derived by analyzing the band values of various soil components. The fertility can hence be quantified based on comparing the aggregate presence of fertile components to the overall image while accuracy can be measured by comparing band values to their closest recognized material and the subsequent classification of such materials as fertile or infertile.

4.5.2. Fuzzy C-Means Clustering. The identity of each piece of data that corresponds to every center pixel is assigned by this technique proportion to the distance between the data point and the cluster center. The closer the data is from the cluster centre, the higher its cluster affinity. As a result, the sum of each data point must equal one, and the membership and cluster center's should be revised after each iteration . The advantages of this algorithm is that it gives the best result of overlapped data sets and as opposed to k-means clustering, where each data point should be exclusive to a single cluster center, membership here is allotted to each cluster center, so the data point can belong to more than one center [11].

$$\mu_{ij} = 1/\sum_{k=1}^{c}(d_{ij}/d_{ik})^{(2/m-1)}$$

$$v_j = \left(\sum_{i=1}^{n}(\mu_{ij})^m x_i\right) / \left(\sum_{i=1}^{n}(\mu_{ij})^m\right), \quad \forall j = 1, 2, \ldots, c$$

where n refers to the number of overall data points, v_j refers to the center of the jth cluster, m refers to the fuzziness index, c is the overall number of cluster centers, μ_{ij} is the affinity of the ith data point to the jth cluster center, d_{ij} is the distance between the ith data point and the jth cluster center.

The advantages of this algorithm is that it gives the best result of overlapped data sets and as opposed to k-means clustering, where each data point should be exclusive to a single cluster centre, membership here is allotted to each cluster center, so the data point can belong to more than one centre.

4.5.3. Agglomerative Clustering. Agglomerative clustering is essentially a strategy that is based on the concept of hierarchical clustering which is a type of cluster analysis that seeks to build a hierarchy of clusters. It relies on the core idea that objects are more related to nearby objects that objects which are further away.

Agglomerative clustering in specific is a bottom up approach where each observation starts in its own cluster, and pairs of clusters get merged as one moves up the particular hierarchy. The algorithm basically nests data points by building them from the bottom up, i.e. each data point acts as its own cluster and then they are combined together to create larger clusters.

Each data point refers to the spectral band values of a particular component obtained by executing spectral unmixing on the original hyper spectral image. By nesting these points from the bottom up, the algorithm uses unsupervised learning and aims to separate the spectral bands as either fertile or infertile,

The classification of the bands as fertile or infertile is then conducted by analyzing the resultant dendogram, while working on the consistent assumption that cluster with the higher band values correspond to more fertile compounds. Thus the fertility can be quantified by comparing the aggregate presence of fertile components to the overall image and accuracy can be measured by comparing band values to their closest recognized material and the subsequent classification of such materials as fertile or infertile [12].

4.6. Segmentation Algorithms.

4.6.1. Fast Marching Method. It is a numerical technique that was developed in early 1996 to solve the boundary value problems that arose in the Eikonal equation. The algorithm works just like Djikstra's algorithm but differs by how the node values are calculated. In the latter, a node's value is calculated by using a single neighboring node, whereas while solving the partial differential equation here, between 1 and n neighboring nodes are used.

During this process, the original noisy hyperspectral image is used as the input, while the algorithm aims to segment the image based on recognizable demarcations, boundaries and areas with familiar patterns. These familiarities are dependent on the FMM algorithm which creates two regions resulting in segments similar to that of the clustering results whenever possible. In cases of three or more segments, further grouping is conducted between the two largest regions and analysis is performed using these regions.

4.7. Algorithm Fast Marching Method.

1. Assume that the domain has been discretized into a mesh. Each node xi has a corresponding value Ui = U(xi) $\approx$ u(xi).
2. Label the nodes as far (those which have not yet been visited), considered (those visited and value tentatively assigned) and accepted (those who have been visited and have been assigned a value permanently)
   (a) Assign every node xi, the value Ui = + ∞ and label them as far, and for all nodes xi $\in \Omega$, set Ui=0 and label it xi as accepted.
   (b) For every far node, use the Eikonal update formula and calculate a new value for U' where U'<U and then set Ui = U'. Label xi as considered.
   (c) Allow x' to be the considered node with the small U value. Label x' as accepted.
   (d) Next, for every neighbor xi of x' that isn't accepted, calculate a tentative U'.
   (e) If U'<U then Ui=U'. Change label to considered if it is labelled as far.
   (f) If a considered node still exists, return to step (c), else terminate.

4.7.1. Efficient Neural Network. Efficient Neural Network (ENet) offers the capability to carry out real-time, pixel-by-pixel semantic segmentation. The method about 18 times faster, has 79 times less parameters, requires 75x less FLOPs and provides better accuracy to existing models.

The model architecture comprises of a 512 * 512 input image resolution. It can be broken down into the following steps:

(i.) Feature Map Resolution: Limited down sampling has been carried out which has a main pro. Filters that operate on such down sampled images tend to possess a larger receptive field, which permits them to procure more contexts. Such a feature can be crucial when trying to separate between various classes.

(ii.) Early down sampling: An important aspect for attaining great real-time functioning and performance is understanding that the cost of operating on bigger input frames can be costly. The ENet's first two blocks largely diminish the size of the input, and utilise only a tiny feature-maps set. Since visual information can be extremely spatially redundant, it can be abridged into a much more methodical representation.

(iii.) Factorizing filters: A succession of processes that are utilized as part of the bottleneck module may also be visualized as breaking up a sizeable convolutional layer into a set of much more uncomplicated operations. This factorization permits greater speedups, and decreases redundancy by heavily reducing the number of parameters.

Name	Type	Output size
initial		$16 \times 256 \times 256$
bottleneck1.0	downsampling	$64 \times 128 \times 128$
4× bottleneck1.x		$64 \times 128 \times 128$
bottleneck2.0	downsampling	$128 \times 64 \times 64$
bottleneck2.1		$128 \times 64 \times 64$
bottleneck2.2	dilated 2	$128 \times 64 \times 64$
bottleneck2.3	asymmetric 5	$128 \times 64 \times 64$
bottleneck2.4	dilated 4	$128 \times 64 \times 64$
bottleneck2.5		$128 \times 64 \times 64$
bottleneck2.6	dilated 8	$128 \times 64 \times 64$
bottleneck2.7	asymmetric 5	$128 \times 64 \times 64$
bottleneck2.8	dilated 16	$128 \times 64 \times 64$
Repeat section 2, without bottleneck2.0		
bottleneck4.0	upsampling	$64 \times 128 \times 128$
bottleneck4.1		$64 \times 128 \times 128$
bottleneck4.2		$64 \times 128 \times 128$
bottleneck5.0	upsampling	$16 \times 256 \times 256$
bottleneck5.1		$16 \times 256 \times 256$
fullconv		$C \times 512 \times 512$

Fig. 4.1: ENET network architecture

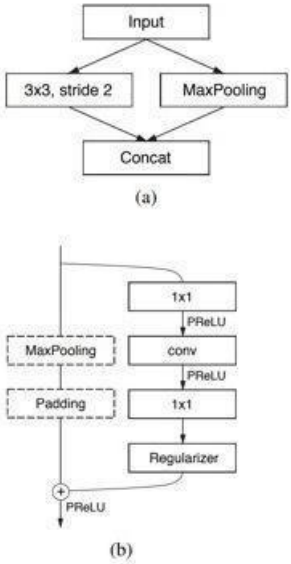


Fig. 4.2: (a) Graphical representation of the first block (b) Bottlenecks

(iv.) Regularization: While the concept of stochastic depth was attempted, in order to increase accuracy, it became apparent that dropping entire branches is actually a special case of applying Spatial Dropout. In this process either all of the channels, or none of them are ignored, and this spatial dropout is placed upon completion of the convolutional branches, just before the inclusion, which came out to work more efficiently compared to the traditional stochastic depth.

As visible in the network architecture, each bottleneck module is made up of the multiple components

which are detailed as follows:

- A 1x1 projection that minimises the number of dimensions.
- A main convolution layer (conv) (full, dilated, or regular convolution) (3x3).
- Expansion 1x1.
- PReLU and Batch Normalization are placed between all convolutional layers.
- Next, a max pooling layer is added to the main branch if down-sampling is the bottleneck. A 2x2 convolution with stride=2 is also used in place of the first 1x1 projection. This is followed by zero padding the activations to match the amount of feature maps, and occasionally using asymmetric (5 * 1 and 1 * convolutions).

The system architecture is composed of five stages. Stages 1, 2, and 3 (the encoder) each have five bottleneck blocks (with the exception of Stage 3 which does not down-sample). Stages 4, and 5 constitute the decoder and possess three and two bottlenecks respectively.

The last step is a fullconv, which produces a final output with the dimensions C * 512 * 512, where C is the number of filters.

4.7.2. Felzenszwalb Segmentation. An important concept to note before delving into the Felzenszwalb Segmentation is the Minimum Spanning Tree (MST) which refers to a cycle-free, graph's edges' minimum-weight subset which connects every node.

Felzenszwalb published an image segmentation approach platformed on Kruskal's MST algorithm in 2004 where analysis of edges is performed in sequence of increasing weightage, and the pixels of the endpoints are combined into a section if they do not generate a cycle in the graph and are 'similar' to the pixels of existing regions. Utilising the disjoint-set data structure, it can be possible to detect cycles in near- constant time. A comparison between the weight and a per-segment threshold is performed by heuristics in order to determine pixel similarity. The technique generates a forest of disjunct MSTs, each of which corresponds to a segment. As sorting edges in linear time is attainable using counting sort, the algorithm's complexity is quasi-linear.

Utilising a rapid, MST-based clustering on the image grid, this approach provides an over-segmentation of a multichannel (i.e. RGB) image. The parameter 'scale' is then used to determine the level of observation with less and larger parts generally being associated to a greater scale. Next, the diameter of a Gaussian kernel 'sigma', is used to smooth the image before segmentation. The only way to control the quantity of created segments as well as their size is by using the scale and the size of individual segments within a picture might vary dramatically depending on the local contrast. Similarly, the euclidean distance between pixels in colour space is used by the algorithm for RGB images.

For the analysis of soil, the colourised version of the hyperspectral is used as the input image with x scale resulting in 90-100 segments. While most of the resultant segments are of sizes less than 500 pixels, the larger segments can be utilised to analyse fertility of soil. The major drawback of this method is the over-segmentation of the image which in turn creates a bottleneck during result analysis.

4.7.3. Deeplabv3 with Pascal VOC model. One of the difficulties in utilising deep convolutional neural networks (DCNNs) to segment objects in images is that as the input feature map shrinks, the network traverses as a consequence of which, information about objects of a smaller scale can be lost.

DeepLab's contribution is the use of atrous convolutions, or dilated convolutions, to retrieve denser features with greater preservation of information from objects of a different scale [2– 3]. The atrous rate is a parameter in atrous convolutions that correlates to the stride at which the input signal is sampled. It's the same as putting 'r-1' zeros between two successive filter values along each spacial dimension in figure 4.2. Because 'r=2' in this situation, the number of zeros between each filter value is 1. The goal of this technology is to be able to change the filter's field-of-view and how dense the features are computed simply by altering r rather than learning additional parameters. The output stride, which is the ratio of the input picture resolution to the output image resolution, is tweaked.

DeepLabv3 adds an image-level functionality to the ASPP module, allowing it to capture longer-range data. It also has batch normalisation options to make training easier and uses atrous convolution to extract output features at different output strides during training and evaluation, allowing BN to be trained at output stride = 16 and evaluated at output stride = 8, resulting in great performance at output stride = 8.

Table 5.1: Spectral Unmixing Results.

Image	Image 1	Image 2	Image 3	Image 4	Image 5
Percentage Error	2.1112%	2.6917%	2.3810%	2.6557%	2.1118%
Identified Soil Components	alunite_cu91-217h.23978.asc	alun_na_mv00-11a.23894.asc	acid_mine_drain_assem2.23799.asc	alunite_cu91-217h.23978.asc	alunite_cu91-217h.23978.asc
	alunite_mix_mv2-ar3.25474.asc	alunite_cu91-217h.23978.asc	alunite_cu91-217g1.23924.asc	alunite_cu98-5c.24129.asc	alunite_nh4+jar_nmnh145596.24230.asc
	alunite_na_dickite_mv99-6-26b.23945.asc	alunite_cu98-5c.24129.asc	alunite_cu91-217h.23978.asc	alunite_mix_mv2-ar3.25474.asc	calcite_talc_pc99-1g.24536.asc
	beidellite_montmor_gds124.24340.asc	alunite_mix_mv2-ar3.25474.asc	alunite_na_dickite_mv99-6-26b.23945.asc	alunite_nh4+jar_nmnh145596.24230.asc	constructed_mixtures_calc_epid.24594.asc
	calcite_talc_pc99-1g.24536.asc	alunite_nh4+jar_nmnh145596.24230.asc	beidellite_montmor_gds124.24340.asc	calcite_talc_pc99-1g.24536.asc	constructed_mixtures_chl_epid_calc.24789.asc
	constructed_mixtures_calc_epid.24573.asc	constructed_mixtures_chl_epid_calc.24766.asc	constructed_mixtures_calc_epid.24594.asc	constructed_mixtures_chlor_calc.24833.asc	constructed_mixtures_chlor_calc.24833.asc
	constructed_mixtures_calc_epid.24594.asc	constructed_mixtures_chlor_calc.24833.asc	constructed_mixtures_chlor_calc.24833.asc	constructed_mixtures_chlor_epid.24870.asc	illite-high-Al_cu00-5b.25438.asc
	constructed_mixtures_chl_epid_calc.24811.asc	dolo_calc_talc_pc99-1e.25017.asc	goethite_phyllite_cu91_236a.25213.asc	goethite_phyllite_cu91_236a.25213.asc	kaolinite_wxl+other_cu00-19a.25536.asc
	goethite_phyllite_cu91_236a.25213.asc	goethite_phyllite_cu91_236a.25213.asc	halloysite_cu91-242d.25302.asc	hydrate_volcanic_glass_cu01-4a.25417.asc	kaolinite_wxl+other_cu91-200a.25557.asc
	hematite_tuff_cu91_223.25396.asc	gyp_jar_ill_brcm1.25237.asc	hematite_tuff_cu91_223.25396.asc	muscovite-medhi-Al_CU91-252d.26143.asc	muscovite-medlow-Al_cu91-250a.26166.asc
	kaolinite_wxl+other_cu00-19a.25536.asc	hydrate_volcanic_glass_cu01-4a.25417.asc	illite-high-Al_cu00-5b.25438.asc	muscovite-medlow-Al_cu91-250a.26166.asc	nontronite_cu00-13a.26189.asc
	musc_pyro_pyrp1.26100.asc	illite-high-Al_cu00-5b.25438.asc	limestone_cu02-11a.25827.asc	pyro_musc_pyrm1.26313.asc	opal_cu00-15e.26266.asc
	nontronite_cu00-13a.26189.asc	limestone_cu02-11a.25827.asc	muscovite-medhi-Al_CU91-252d.26143.asc	pyroxene.basalt.cu01_20a.26360.asc	stonewall_playa.26390.asc
	opal_cu00-15e.26266.asc	muscovite-lowAl_cu98-8h.26121.asc	stonewall_playa.26390.asc	talc_clinochl_hs327.26428.asc	
	pyro_musc_pyrm1.26313.asc				
	stonewall_playa.26390.asc				
	talc_clinochl_hs327.26428.asc				

Table 5.2: Unsupervised Clustering Results.

	Clustering Algorithms						
Image	K-Means Result	Percentage Fertility	Agglomerative Result	Percentage Fertility	Fuzzy C-Means Result	Percentage Fertility	Actual Result
Image 1	Infertile	25%	Fertile	65%	Infertile	45%	Infertile
Image 2	Infertile	40%	Infertile	5%	Infertile	40%	Infertile
Image 3	Infertile	45%	Infertile	25%	Fertile	55%	Infertile
Image 4	Infertile	40%	Infertile	10%	Infertile	45%	Infertile
Image 5	Fertile	85%	Fertile	90%	Fertile	55%	Fertile

The implementation in the current scenario uses the "PASCAL VOC" dataset which is a popular dataset of choice while building models that are being conditioned for image classification and segmentation.

5. Results and Discussions. Using the above-mentioned method, the considered area in each year from 2020 to 2022 is classified into fertile/unfertile sectors. The parameters that are mentioned below are:
Class: Each class identifies the global constant based on which the classification is done.
Fertility Percentage %: The percentage shows the percentage covered by the class compared to the other Classes.
Epochs: The number of iterations for which the image is processed to reduce error percentage

5.1. Spectral Unmixing. Using the NFINDR Spectral Unmixing algorithm in order to identify the component endmembers of a hyperspectral image, the results obtained were as in Table 5.1. The two properties obtained as results were the individual soil components present in the soil hyperspectral image (refer to table) and the percentage similarity to the endmembers making up the original image measured using cosine similarity.

As per the obtained results (Table 5.1), the NFINDR technique presents itself as a highly accurate and efficient method of spectral unmixing for the precision agriculture use case. With similarity values consistently lower than five percent, the endmembers obtained can be adequately identified as specific soil components through comparison of wavelength values. The results present an opportunity for the use of hyperspectral imaging and spectral imaging in the field of agriculture to accurately capture soil components over large areas with little physical overhead, a problem currently found in the sector. Additionally, decisions regarding fertility, regions of improvement, potential crop yields etc. can be taken with technological evidence [15].

5.2. Unsupervised Clustering. Unsupervised clustering algorithms were used for classifying component soil endmembers based on their fertility utilizing the results of the spectral unmixing. The algorithms used were K-Means Clustering, Agglomerative Clustering and Fuzzy C-Means Clustering. The results obtained were as in Table 5.2.

The properties obtained for each algorithm were the classification result i.e. whether the soil sample was adjudged to be majorly fertile or infertile, and the percentage fertility measured based on the size of the cluster and percentage presence of fertile or infertile components in the soil sample image (Table 5.2). The actual result was obtained using a combination of subjective analysis of the obtained soil endmembers and the pH index of the soil image as a whole.

Table 6.1: Segmentation results based on fertility

	Image Segmentation Algorithms										Actual Result	
Image	K-Means Segmentation Result	pH Value	FMM Segmentation Result	pH Value	Felzenszwalb Segmentation Result	pH Value	DeeplabV3 Result	pH Value	ENET Result	pH Value	Result	pH Value
Image 1	Infertile	0.0259	Infertile	0.0355	Infertile	0.0395	Infertile	0.0273	Infertile	0.0603	Infertile	0.0117
Image 2	Infertile	0.0329	Infertile	0.0458	Infertile	0.0405	Infertile	0.0563	Infertile	0.0544	Infertile	0.0123
Image 3	Fertile	0.0308	Infertile	0.0322	Infertile	0.0403	Infertile	0.0308	Infertile	0.0624	infertile	0.0098
Image 4	Infertile	0.0366	Infertile	0.0238	Infertile	0.0384	Infertile	0.0405	Infertile	0.0589	Infertile	0.0150
Image 5	Fertile	0.0409	Fertile	0.0507	Fertile	0.0419	Fertile	0.0304	Fertile	0.0398	Fertile	0.0221

6. Results and Discussions. The results provide crucial insight into the potential of clustering algorithms in the field of precision agriculture. Unsupervised clustering algorithms provide reasonable accuracy with respect to fertility classification and further insight through the fertility percentage measure. Apart from highlighting the usefulness of hyperspectral imaging in precision agriculture by directly determining fertility, it indicates the potential for further field-based research in the sector hence creating a platform for supervised algorithms and increasing the use-case for hyperspectral imaging in precision agriculture. Further, due to the high similarity between original soil components and spectral unmixing results, the prerequisites for the utilization of clustering algorithms are easily satisfied and provide promising results. Another notable advantage of utilizing clustering algorithms is the low overhead required for analysis in comparison to image segmentation. Unlike the latter, clustering directly uses spectral endmembers which require less physical memory to store and processing power required is significantly lower due to the lower complexity and requirements. This is particularly consequential as in agricultural fields, lower hardware and software requirements are preferred with lack of accessibility, resources and connectivity being features of rural India.

In comparing the three algorithms, K-Means clustering and Agglomerative clustering techniques have the highest accuracy in fertility classification with the latter preferred due to the lack of dependence on initial centroid values and subjectivity created in using the former. While the Fuzzy C-Means is adequately accurate, the results of the other tested techniques offer better results.

6.1. Image Segmentation. Image segmentation algorithms were used for classifying regions of the soil sample image based on their fertility using the colorized hyperspectral image. The algorithms used were K-Means Segmentation, FMM Segmentation, Felzenszwalb Segmentation, DeeplabV3 and ENET segmentation.

The properties obtained for each algorithm were the classification result i.e. whether the largest segment obtained was adjudged to be fertile or infertile, and the pH index of the largest segment obtained post-segmentation. The former is measured comparing the pH index of the largest and the next-largest segment obtained from the segmentation process. pH index was measured through the RGB index values of each individual pixel of the image (add reference). The actual result was obtained using a combination of subjective analysis of the obtained soil endmembers and the pH index of the soil image as a whole.

As per the obtained results, image segmentation is yet another potential avenue in precision agriculture using hyperspectral imaging. Working on the principle of classifying regions of fertility based on the soil component endmembers, the major function of such segmentation is the identification of fertile soil in large plots of land. This in turn can help agriculturists predict crop yields and prioritise such identified regions for the same. Additionally, by classifying whole regions (as captured by the image) as fertile or infertile it can help states and agriculturists during the purchase and transfer of land as well as during harvest, to identify regions of high yield and prioritising such regions. The pH index measure provides further insight into the properties of the soil hence giving quantifiable data to work with. For selection of comparison algorithms, we have taken five random images. Results are shown in table 6.2. While in other Algorithms, the variation in results according to noise level is high. But in PSNR it is comparatively low. As shown in table 6.1 PSNR is a good option to use for finding fertility of selected area in 2020 and 2022.

PSNR demonstrates that the picture is of acceptable quality.

$$PSNR = \log_{10} \frac{MAX_k^2}{MSE} \tag{6.1}$$

For SSIM calculation, reference image C2 is a constant added to avoid instability when other terms are

Table 6.2: Performance error metrics comparison

Image	Epoch	Felzenszwalb	Deeplabv3	Enet
		MSE	PSNR	SSIM
1	100	87.86	31.51	0.94
2	100	65.51	29.96	0.89
3	100	45.04	28.77	0.85
4	100	67.93	29.80	0.88
5	100	89.33	31.44	0.93

close to zero. Correlation factor is calculated by Eq. (6.2)

$$s(x,y) = \frac{2\sigma_{xy} + C_2}{\sigma_x \sigma_y + C_2} \tag{6.2}$$

where σ_{xy} in Eq. (6.2) is a covariance between the two images and C3 for avoiding instability. Finally, combining all the factors equations, namely Eq. (6.1), Eq. (6.2) we will get SSIM as shown in Eq. (6.3)

$$SSIM(x,y) = \frac{(2u_x u_y + C_1)(1\sigma_x + \sigma_y + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)(u_x^2 + u_y^2 + C_1)} \tag{6.3}$$

where σ_{xy} in Eq. (6.3) is a covariance between the two images and C3 for avoiding instability. It has indistinguishable units of estimation from the square of the amount being determined like variance, which is based on original data x, y, and is defined as in Eq. (6.3).

MSE is a full reference metric, and the qualities more like zero are better. The fluctuation of the estimator and its bias are both fused with the mean squared error.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \tag{6.4}$$

All algorithms tested, with the exception of the K-Means image segmentation, show high accuracy with respect to fertility classification while the more primitive methods of FMM and Felzenszwalb show higher accuracy for the pH index. However, the neural network algorithms of DeeplabV3 and ENET are preferred due to their higher reliability as seen through the higher difference between pH index values of each segmented region.

The drawback however stems from the problem of overhead where the latter algorithms require high computing power and prerequisite software in order to function, an issue overcome by using the former techniques. On the whole, image segmentation requires higher overhead for usage as the input is raw hyperspectral images which take up more storage and require higher software and hardware to run and maintain.

7. Conclusion. In general, the application of hyperspectral imaging in precision agriculture is a potentially lucrative research subject with significant application in the rural Indian areas. The solution's scope includes both short-term and long-term considerations. The former includes immediate analysis of soil fertility in fields, estimation of crop yields, maximization of profits, and uses in land transfer. The latter includes quantification of fertility using imaging rather than manual testing, analysis of the long-term fertility of land, and the effectiveness of various agricultural practices, among other things. When compared to the manual laboratory setting for testing soil, the results' high accuracy establishes the conditions for a contemporary, trustworthy data source for agriculture experts and farmers alike with fewer overhead and expenses.

Hyperspectral imaging is a prospective addition to the same, especially with the expansion of digital technology in rural areas and the introduction of technology. The findings and subsequent analysis also aid

in filling a research gap in the application of unsupervised classification models in the field as well as the dependability of produced hyperspectral image datasets. The latter is particularly important because it is less difficult, costs less to operate, and uses fewer digital resources than developing, testing, and deploying supervised algorithms.Even in the case of supervised algorithms, the numerous properties and features that can be derived from digitally made hyperspectral pictures serve as a foundation for training and testing such algorithms without the need for manual analysis and expensive research costs.

REFERENCES

[1] L. Liu, M. Ji, and M. Buchroithner *Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery* Sensors, vol. 18, no. 9, Sep. 2018, Art. no. 3169.

[2] A. K. Patel, J. K. Ghosh, S. Pande and S. U. Sayyad *Deep-Learning-Based Approach for Estimation of Fractional Abundance of Nitrogen in Soil From Hy- perspectral Data* in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 6495-6511, 2020, doi: 10.1109/JSTARS.2020.3039844.

[3] Li, Hongyang, Shengyao Jia, and Zichun Le *"Quantitative Analysis of Soil Total Nitrogen Using Hyperspectral Imaging Technology with Extreme Learning Machine.* Sensors (Basel, Switzerland) vol. 19,20 4355. 9 Oct. 2019, doi:10.3390/s19204355

[4] Chen, Yun, Yuqiang Li, Xuyang Wang, Jinliang Wang, Xiangwen Gong, Yayi Niu, and Jing Liu. *Estimating soil organic carbon density in Northern China's agro-pastoral ecotone using vis-NIR spectroscopy*Journal of Soils and Sediments. 2020, 20(10):3698-711.

[5] Anna Chlingaryan, Salah Sukkarieh, Brett Whelan*Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review* Computers and Electronics in Agriculture,Volume 151, Pages 61-69, ISSN 0168- 1699,2018, doi: 10.1016/j.compag.2018.05.012.

[6] L. Sun, F. Wu, T. Zhan, W. Liu, J. Wang and B. Jeon*"Weighted Nonlocal Low-Rank Tensor Decomposition Method for Sparse Unmixing of Hyperspectral Images* in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 1174-1188, 2020, doi: 10.1109/JSTARS.2020.2980576.

[7] Somenath Bera and Vimal K. Shrivastava *Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification* International Journal of Remote Sensing, 41:7, 2664-2683, 2019, doi: 10.1080/01431161.2019.1694725

[8] Chang Li, Yu Liu, Juan Cheng, Rencheng Song, Jiayi Ma, Chenhong Sui, Xun Chen *Sparse unmixing of hyperspectral data with bandwise model, Information Sciences* Information sciences, 512, pp.1424-1441,2020.

[9] Yuan, Yuan, Zihan Zhang, and Qi Wang*Improved Collaborative Non- Negative Matrix Factorization and Total Variation for Hyperspectral Unmixing* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. PP. 1-1,2020

[10] Song, Y.-Q.; Zhao, X.; Su, H.-Y.; Li, B.; Hu, Y.-M.; Cui, X.-S.*Predicting Spatial Variations in Soil Nutrients with Hyperspectral Remote Sensing at Regional Scale.* Sensors , 18,2018, doi: 10.3390/s18093086

[11] Haijun Qi, Tarin Paz-Kagan, Arnon Karnieli, Xiu Jin, Shaowen Li,Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data, Soil and Tillage Research,Volume 175,Pages 267-275, ISSN 0167 1987,2018

[12] Jia S, Li H, Wang Y, Tong R, Li Q. *Hyperspectral Imaging Analysis for the Classification of Soil Types and the Determination of Soil Total Nitrogen.* Sensors (Basel),(10),2017, doi:10.3390/s17102252

[13] Liguo Wang, San Wang, Xiuping Jia and Tianyi Bi *A Novel Hyperspectral Unmixing Method based on Least Squares Twin Support Vector Machines*, European Journal of Remote Sensing, 54:1, 72-85, DOI: 10.1080/22797254.2021.1877572

[14] Peng Y, Zhao L, Hu Y, Wang G, Wang L, Liu Z. *Prediction of Soil Nutrient Contents Using Visible and Near-Infrared Reflectance Spectroscopy.* ISPRS International Journal of Geo-Information, 8(10):437,2019, doi: 10.3390/ijgi8100437

[15] Xue Song, Yongxiang Gao, Zhiguang Liu, Min Zhang, Yongshan Wan, Xinyang Yu, Wenlong Liu and Lei Li *Development of a predictive tool for rapid assessment of soil total nitrogen in wheat-corn double cropping system with hyperspectral data*, Environmental Pollutants and Bioavailability, 31:1, 272-281,2019, DOI: 10.1080/26395940.2019.1679041

Edited by: Vinoth Kumar
Received: Jul 31, 2022
Accepted: Oct 19, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 303–319, DOI 10.12694/scpe.v23i4.2034

CLOUD BROKER RECOMMENDATION FRAMEWORK TO PROVIDE TRUSTWORTHY CLOUD SERVICES TO THE END USER

M. MARIMUTHU*, J. AKILANDESWARI †, VARASREE B‡, RAJESH KUMAR GUNUPUDI§, AND SOMULA RAMASUBBAREDDY¶

Abstract. In recent years, many cloud services have become available on the Website. Discovering suitable cloud services for the end user is incredibly complex and difficult. The cloud brokerage service is an application that aids in providing solutions for this problem. It recommends suitable cloud service providers to the end users depending on their relevant requirements. The Internet provides access to a wide variety of cloud brokers. As a result, choosing a cloud broker or service provider is both time-consuming and tedious. It is now becoming a necessity to choose a proper cloud brokerage service based on trust. Research works found in the literature address some of the issues and provide feasible solutions by proposing frameworks, optimizations and rule based algorithms. However, those works focus solely on delivering a trustworthy service to the end user through application of techniques and algorithms. There is no proper framework model in place to provide suitable and trustworthy recommended services to the users. This article provides a detailed description of the frameworks that are offered by the researchers, including issues and proposes a trustworthy recommendation framework (TRF) to provide trustworthy services to the end user. This article also presents a Trustworthy Recommended Weighted value (TRWv) approach for determining trustworthy services, and it is discovered that the proposed method achieves high accuracy (91.3%) when compared to similar works.

Key words: cloud broker, cloud service provider, end user, recommendation framework, logistic regression

AMS subject classifications. 62J05

1. Introduction.

In cloud computing, diverse computer system resources (networks, servers, storage, applications and services) are shared and configured automatically with less administrative effort as and when required by the end user. It is a new technology that provides end users with utility services as resources on demand. Computing resources is symbolized by on-demand self-service, extensive network connectivity, rapid elasticity, resource pooling and measured service. First and foremost, identify the most suitable cloud deployment model or architecture to implement cloud services [1]. Platform as a service (PaaS), Infrastructure as a service (IaaS), and Software as a service (SaaS) are the various types of cloud computing services. It is conceivable to contract IT resources such as networks, servers, storage, virtual machines and operating systems, from a cloud service provider (CSP) via IaaS. Platform as a service provides a reliable environment which enables the user to build, evaluate, distribute, and also to perform various administrative tasks with respect to an application. On demand, subscription-based access to software programmers via the Internet is referred to as SaaS.

The deployment paradigms are configured as public cloud, private cloud, hybrid cloud, and community cloud [2, 3]. Public clouds, typically disseminate computing resources such as servers and storage through the Internet, are initiated and operated by the trusted intermediate CSP. The cloud infrastructure is accessible to the general public. Private cloud is a cloud storage service, which is owned and monitored solely by a single organization via private network and it is physically hosted at an onsite premises. Some organizations pay for intermediate service providers to set-up their private clouds. A hybrid cloud is consists of two or more distinct clouds (private, public, and community) that are combined together which permits transfer of data and applications between them. A hybrid cloud increases the organization's flexibility by enabling the migration of

*Assistant Professor, Department of Computer Science and Engineering, Sona College of Technology (mari.btech@gmail.com).
†Professor, Department of Information Technology, Sona College of Technology (akilandeswari@sonatech.ac.in)
‡Information Technology, Institute of Aeronautical Engineering, Hyderabad, India.(b.varasree@iare.ac.in)
§Information Technology, VNRVJIET, Hyderabad, India.(gunupudirajesh@gmail.com)
¶Information Technology, VNRVJIET, Hyderabad,India.(svramasubbareddy1219@gmail.com)

data and applications between private and public clouds. Community cloud infrastructure is used by several organizations to achieve a common goal.

Cloud Service providers (CSPs) are vendors that provide Information Technology (IT) as a service via the Internet. Many CSPs are available online. Therefore finding a CSP is an exhausting task for end users [1, 4]. Majority of CSPs provide identical functionality, which creates the service selection problems for end users. Selecting the best service provider requirements is a challenging and time consuming job. To solve this issue, Cloud Broker (CB) plays a vital role in providing an optimal CSP to the end user. The CB has been considered as a major concern for emerging cloud technology. A CB is a firm that oversees the usage, efficiency, and distribution of cloud services, as well as negotiates contracts among CSPs and end customers. CB acts as a go-between for CSPs and end users. It provides three categories of services namely

Service Intermediation: Enables value-added services or enhances functionality, such as controlling cloud access.

Service Aggregation: A CB combines and integrates multiple services into one or more new services. It facilitates data integration and ensures data security while transporting data between a end user and several CSPs.

Service Arbitrage: In order to create new services, a CB combines various kinds of services. It is similar to aggregation, except aggregation is not fixed. It has the ability to choose the CSPs based on the data features.

CBs provide a variety of services, including application-to-application (A2A), business-to-business (B2B), and trade partner relationship management. The desired services are not provided by all CSPs to end users. The user's perspective is fully focused on selecting the precise CB, major concerns in selecting the precise cloud broker are trustworthiness and recommended services as trusted ones. [5] It is also an efficient method for assessing and establishing a relationship between a service provider and user [6, 7, 41].

Making a decision or recommending a cloud services in an inter-cloud environment is not an ordinary task. The unavailability of precise information like QoS features, TMP and trustworthy further added to the issue. In a cloud environment, the resource selection procedure is incredibly difficult and less reliable due to the immense complexity and constraints of existing methodologies [8]. In this context, manual approaches will be inefficient and time-consuming; therefore, automated recommendation systems are required to assist the consumer in choosing the best cloud services [9, 10].

An effective trust management framework provides optimal as well as trustworthy services to end users. However, in a cloud context, trust assessment is one of the most challenging tasks [11]. Most of the frameworks, algorithms, approaches and methodologies determine the trustworthiness of the cloud brokerage service or service provider to provide trusted services. The selection of cloud services falls into one of four categories: decision-making, prediction-based, approach-based, and optimization techniques [12]. Approach based methodologies are Quality of service (QoS) ranking algorithms or models based on subjective or objective assessment or integrating both. Decision making is the approach for identifying the best CSP depending on the end user's needs. Optimization is a technique used to identify the appropriate CSPs. Linear programming and predictive techniques are used to forecast events. If the cloud user does not have any idea about the service, the feedback derived from the historical data that depends on the service provider is used. By using this approach, the best service provider can be predicted [13, 14]. Machine learning algorithms, statistical algorithms and data mining algorithms are the few examples of prediction. There are other methodologies to identify the best CSP for the end user. First order logic, Computation techniques, Fuzzy set theory, Fuzzy ontology prioritized aggregation operator and ranking oriented prediction are some of them.

All the research articles only focused on trustworthy service using the above techniques, but none of the research articles concentrate much on the framework. The framework provides a complete idea of the trustworthy services to the end user and, by using this framework model, can incorporate any of the technologies listed above. This research article provides a detailed survey of the framework models available, and proposes a new Trustworthy Recommendation Framework (TRF) model for selection of services in the cloud environment using the Trustworthy Recommended Weighted value (TRWv) approach, which also provides trustworthy services to end users and is suitable for real-time environments.

The remaining sections of the article are organized as follows: Section 2 presents the overall perspective of relevant work on the framework model. The proposed trustworthy recommendation framework model to enable

a recommended trustworthy cloud services to the end user is outlined in Section 3. Section 4 elaborates the results that are compared with other framework models available in the literature. The conclusion and future developments of the research study are addressed in Section 5.

2. Related Work. Research has shown that the provision of trust services to end users remains a challenge because of their privacy, protection, reliability, availability and dynamically distributed environment. Most of the articles used service provider capabilities and end-user reviews to identify trustworthy cloud services. Similar services are identified for end users in a cloud environment based on the availability of such services in a dynamic nature or depending on the quality of services identified for the end user. In this section of the literature, a comprehensive description of the trust framework model available in the current scenario has been presented and, finally, feedback on the framework model has been given.

Somu et al. [15] proposed a TrustCom – a novel trust assessment framework for identifications of the trustworthy services. Selection of services is determined by the level of trust, the accuracy of trust value based on Trust Measure Parameters (TMPs) including availability, accountability, and cost. Hypergraph-based computation model was utilised to combine Rough set theory (RST) and existing features for feature selection when processing huge datasets because RST has constraints with respect to all subsets for picking the optimal feature subset. The cloud services are ranked based on the trust computation value. Noor et al. [16] presented a web service (WS)-based trust management in cloud environment. Many solutions have been proposed to identify the trust review/feedback, however establishing the legitimacy of user feedback has been neglected. Dynamic cloud environment is also very difficult to predict service availability. The author introduced an Adaptive credibility model (framework) that is used to figure out credible and malicious feedback. Once again, trust depends on the availability of services in the cloud environment.

Devi et al. [17] proposed a Linear Programming (LP) model which is used to order the cloud services dynamically. Due to the availability of many CSPs, it is difficult to choose the optimal cloud service. The quantitative and qualitative approaches are used to predict the service provider for end users based on QoS requirements. As per requirements given from the end user, the weightage is assigned. So the maximum weight is assigned to the highest order of the requirements to evaluate the rank based on objective method. The ranking score of the providers is used to determine the best CSP.

Qu et al. [18] proposed a CCCloud framework model, where service selection depends on subjective and objective assessment. Subjective assessments are obtained from end users and objective assessments are obtained from third parties about service providers' performance or capabilities. They conclude the trustworthy by comparing and aggregating both assessments of the overall performance of CSP services. The objective assessment is dynamically changed based on the similarity of the context and the weighted value assigned to the cloud services. It not only provides the credibility of a user i.e. legitimate or not, but it also prevents user collisions in cloud services. Jayapriya et al. [19] proposed the CorQosCloud framework architecture, where quality of service (QoS) being a significant factor for choosing the services in a cloud environment. Service recommendation is done by rating and ranking. In this article, along with active user, the correlated QoS ranking algorithm was also presented to get the exact feedback about the service to achieve accuracy of the service.

Kanwal et al. [20] proposed a Trust Evaluation Model (TEM) to calculate the trust scores. In cloud federation, CSPs are dynamically collaborating to share their Virtual Machine (VM) infrastructure facility, due to the demand of quality of service during in load balancing. In order to check the data privacy and security author propose the TEM, because lack of trust between CSPs. It establishes a trust relationship between CSPs among SLAs [21]. Liu et al. [22] proposed Trust-aware recommender systems to provide a reliable trust aware QoS cloud services to the end users in a cloud environment. It combines the clustering-based algorithm and trust-aware method to provide the active user with a more customized QoS forecasting and trustworthy cloud services are recommended. It also overcome the data sparsity problem, task similarity among a similar users may result in differing prediction outcomes [23].

Habib et al. [24] proposed Trust aided unified evaluation framework. In this article both trust and reputation are considered as they relate to CSPs. To measure the trustworthiness of the cloud services, leveraging the trust and reputation systems that relate to QoS cloud related parameters are identified. The primary goal of the research is to provide an estimation of the future behaviour of the service provider. Noor et al. [25]

Table 2.1: Summary of Trust framework model

S.No	Author	Framework	Observations
1.	Devi et al. 2020 [17]	Linear Programming (LP) model	In the detailed explanation of the SLA repository, it is not clear how the weight is assigned to the end user requirements. Because in a real cloud environment, the needs vary for each end user.
2.	Liu et al. 2019 [22]	Trust-aware recommender systems	Experiment setup has only 2 data sets to implement the approach to achieve a better performance. Identification of the task similarity and data credibility problem in a cloud environment is not clearly mentioned.
3.	Al-Faifi et al. 2018 [32] and Wang et al. 2019 [33]	MCDM algorithm	MCDM algorithm is used to find out the optimal services for end users using clustering techniques. Performance wise it provides optimal services to the end user but reputation and trust are not focused.
4.	Smithamol et al. 2018 [30]	Trust management mechanism (TMM)	To evaluate the trustworthiness of cloud services in a real-time cloud environment, such as a multi-cloud environment, performance must be monitored in a dynamic manner, i.e. the availability of cloud services vary in real time context.
5.	Somu et al. 2017 [15]	TrustCom – a novel trust assessment framework (RSHT)	TThe identification of the subset of trust measure parameters requires the least amount of time. The efficiency of trustworthy services is carried out at fixed intervals, and the performance of the framework is similar to existing feature selection techniques.
6.	Jayapriya et al. 2016 [19]	CorQosCloud framework	The accuracy of the framework model compares with a few other frameworks, but the accuracy is purely dependent on the availability of services. Time and location are also major components of the accuracy that is not described in detail.
7.	Qu et al. 2015 [18]	CCCloud framework	Subjective and objective assessments are not extracted from the real dataset prior to implementing the cloud service selection model; only 50 percent of it is extracted from the cloud services and the remaining 50 percent is partially generated. So the result cannot be accurate, and the performance is carried out without the assessment context.
8.	Noor et al. 2014 [25]	Generic analytical framework (TCSR)	Framework compares the trust management prototype based on the assessment criteria such as (security, privacy, personalization etc.). But the criteria are varied in a distributed cloud environment due to the availability of services.
9.	Kanwal et al. 2014 [20]	Trust Evaluation Model (TEM)	Qualities of Service (QoS) parameters are determined by security and privacy and it is extracted from SLAs. The trust score are exchanged between CSPs to enhance cloud federation.
10.	Zheng et al. 2013 [28]	QoS ranking prediction framework (KRCC)	Due to the availability of cloud services in different contexts, such as time and location, the accuracy of the ranking model should be enhanced. Cloud services are ranked based on past usage of consumers.

proposed d a trust management service in the cloud environment. Various trust management techniques are available, which are categorized into Policy, Recommendation, Repudiation and Prediction. Generic analytical frameworks, for each level a set of dimensions are identified to evaluate and analyse the trust. It also compares various trust management methods depending on the evaluation criteria.

Kumar et al. [26] talks about the cloud service selection in spite of the availability of the services across a wide range of cloud platforms. Many service providers have the same functionality, and therefore choosing an appropriate service provider is one of the most important tasks. A non-functional QoS requirement has been taken to the cloud service ranking in terms of Fuzzy logic. Noor et al. [27] addresses the credibility of the user feedback and also recommends to the end user a CSP that completely depends on trust. Zheng et al. [28] proposed a CloudRank QoS ranking prediction framework. The selection of cloud services are determined by a QoS ranking algorithm through a set of functionality. It is quite difficult to identify the cloud services in a

real time cloud environment. The consumer experience is included in this framework. Depending on the needs or requirements, the prediction framework provides an optimal cloud services to the end user. The framework not only provides optimal services, it also reduces the time and efforts consumed process of choosing optimal services for the end user.

Li et al. [29] discusses in the same issue, but the proposed T-broker serves as an intermediary between the user and the cloud environment. Depending on the trusted attributes, it provides the trustworthy cloud services to the end user. It has a lightweight feedback mechanism to improve the efficiency or accuracy of the cloud services that are offered by the end user. Smithamol et al. [30] presented a trust management mechanism for multi cloud environments. The trust management framework supports a multi-cloud environment with distributed cloud service availability in order to determine the CSP's trustworthiness. The framework mainly focuses on accuracy and efficiency [31].

Al-Faifi et al. and Wang et al. [32, 33] addresses the selection of cloud services depends on the hybrid multi criteria decision method. Using clustering techniques, similar features that are offered by service providers are identified. Ranking of the service providers is obtained after applying the multi-criteria decision making (MCDM) algorithm. The following Table 2.1 provides a brief overview of the trust framework model present in the literature.

As per the above mentioned literature, the selection of cloud services in a multi cloud environment is decision making/ranking/recommendation based. Due to the availability of services in a cloud context, the trust value could be measured [34]. But, measuring the trust value is not a simple task because of lack of subjective and objective assessment. Identifying or extracting real-time features from CSPs and end users to measure trust is extremely difficult. Even identifying the services for end user requirements is a difficult task; the requirements vary for each individual. In cloud federation [35], an enormous pool of services is constructed. Many of the pool's services are of the same nature, but they are accessed in different ways and have distinctive features. Consumers' must identify and pick the desired service from the available options. Due to the same functionality of services that are given differently, it is difficult for end users to choose the best option. This existing research is only suitable for consumers who wish to identify cloud services that are identical to the method they already knew or engage, but does not take into account those who look for the optimal services without background knowledge [36, 37, 38].

In order to solve these issues and assist end users in their cloud service selection, a systematic recommended framework model is needed. The framework proposed incorporates model creation and Trustworthy Recommended Weighted value (TRWv). The model creation will be done by using the ML algorithm (Logistic Regression) and TRWv done by service provider capabilities and end user feedback ratings. TRF is an intelligent automated framework that can tackle the issue of identifying the best cloud services among numerous services available in the Internet. By making recommendations for necessary services, it can be utilized to effectively deliver customized services to end users. When a end user seeks a recommendation for a cloud service, the system identifies the most relevant services based on the cloud consumer's past interactions. The proposed trustworthy framework and methods provide conceptual recommendations for creating and developing relevant recommended systems for end users.

3. Proposed Recommendation Framework. A strong design framework makes available all feasibility to the end users, offering accurate services also a good elucidation to the trustworthiness. The framework integrates the capabilities of the service provider and extracts the consumer feedback, reviews, ratings and recommendations from past historical data. A framework itself has various interfaces such as service provider capability, consumer realistic experience and automated tools (using a machine learning approach), incorporating the whole interface which is connected to a single model.

Machine Learning (ML) is a technique that offers the ability to learn itself, discover hidden insights, train and enhance from experience without explicit programming [39]. A machine learning approach has been used in this article to build a TRF model, where more number of features and a huge dataset is included in this study. The ML technique helps in predicting the services for new features requirements which are given by an end-user, because the target features are of categorical type. The target output has 10 entities namely CSP1, CSP2 ... CSP10, based on the end user requirements the output may fall into any of the categories. The overall proposed trustworthy recommendation framework Figure 3.1 provides trusted services to end users as

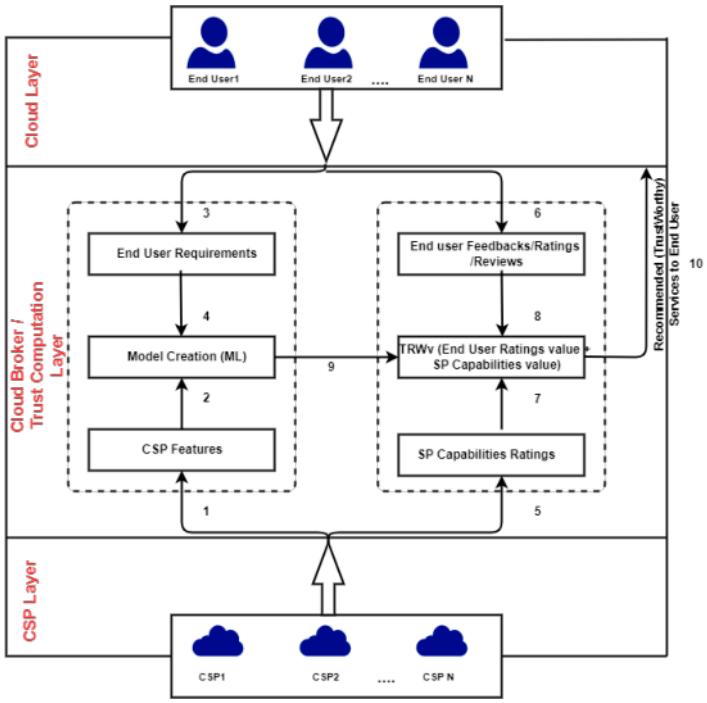


Fig. 3.1: Trustworthy Recommendation Framework

compared with other framework models.

Identifying CSP features or selecting features has been a difficult task. This article focuses only on the recommendation framework to provide optimal services to the end user. Machine learning model creation and trust value computation using Trustworthy Recommended Weighted value (TRWv) are the major parts of the recommendation framework. About 250 features have been identified to incorporate the ML model. It includes a range of services, such as computing, storage, networking, management, security and performance. These services are given by the CSPs. For the implementation of the TRWv approach, end-user ratings/reviews and service provider capabilities have been taken as trust measure parameters. It has a total of 76 features, 52 end-user rating features and 24 service provider features, such as SLAs (Service Level Agreement), Audits and Compliance, Measuring and Ratings, Security and Performance. The entire dataset or features are available in the GitHub repository [40].

The proposed framework has three layers such as 1.Cloud Layer 2. Cloud Broker/Trust Computation Layer 3. Cloud Service Provider Layer. Cloud Layer and CSP layer are interconnected with each other. In the cloud environment, all the detailed information about the end user is given by the cloud layer. To offer a trust service to the end user, the cloud layer is used to obtain the user requirements and service provider ratings and the information is provided to the trust computation layer. CSP layer provides a complete solution/idea about the service provider's availability on the Internet. To enable the services, all service providers have to register in the CSP registry. With the help of service provider (CSP-Registry) registry, the entire CSPs details are extracted from the Internet. The CSP layer was used to extract the service provider features and capabilities that are required for ML model creation and trust computation.

The Cloud broker/Trust computation layer acts as a middle layer in the proposed framework. Machine learning model creation and trust computation are done in this phase. The Trustworthy Recommended Weighted value (TRWv) is the primary component of the trust computation layer that determines whether the ML model returns the services which is trustworthy or not. The process is repeated until trusted services are obtained.

After identifying the features of each service, the ML model is built using a machine learning algorithm

(Logistic regression). The end user requirements are given to the ML model to obtain trustworthy services. After obtaining the end user requirements, the ML model will provide the services to the trust value computation approach TRWv. The TRWv combines SPVA and EVA. End-user ratings value average is referred to as EVA, while service provider capabilities value average is referred to as SPVA. The CSPs capabilities and end user feedback/reviews/ratings about each services provider available in the cloud layer and CSP registry are extracted, and a rating is given out of 10. The end user has to rate each service which are given by the service provider. Depending on the service provider's capabilities they can assign their ratings for their own services. The trust value was computed for the all the services with TRWv (Trustworthy Recommended Weighted value) by combining CSP capabilities and end user ratings. In order to measure the trust value, 70% of the weightage is given to the end user ratings and 30% weightage is given to the service provider capabilities. The trust value computation algorithm will check each service, whether the returned services trust threshold value is equal or greater than to 7. If the service value is greater than 7, the services are returned to end user through the cloud layer. If the service trust score is not satisfied by the threshold, the ML model accesses the next service. These services are known as recommended or trustworthy services. The results suggest that the trustworthy recommendation framework increases the accuracy of trust services and is more accurate in determining trustworthy services then other related frameworks that are discussed in literature section.

The trustworthy recommendation proposed framework contribution is outlined in the following steps:

1. Identify each service feature.
2. The model was created using a Logistic regression algorithm for each service along with service provider features.
3. Extract end user requirements from the cloud layer.
4. End user requirements have been given to the ML model.
5. Identify SP capabilities ratings in CSP layer.
6. Extract end users feedback/reviews/ratings.
7. SP capabilities values are given to TRWv.
8. End users values are given to TRWv. TRWv compute the trustworthiness of a service by integrating the value given by the CSP and user.
9. Depends on their end user requirements the model was given the services, which are returned to trust computation.
10. Identify whether the services are trustworthy or not using Trustworthy Recommended Weighted value (TRWv) approach/method that is given to end users. If the services are not trustworthy again these services are returned to ML model, until they get a trusted service.

The implementation of the trustworthy recommendation model is given as step by step pseudo code in Algorithm 1, which is represented as Trustworthy Recommended Weighted (TRW)_cloud broker. It begins with the creation of the model, progresses to the testing of the model, and concludes with the identification of trustworthy services for the end user.

The TRW_cloud broker algorithm has three parts namely,

1. Model creation
2. Testing the models
3. Identify the trustworthy services

It starts with reading a dataset from a service that is used to build a machine learning model. The dataset is represented in the form of a matrix with features (F) and rows (N), i.e., the number of samples. After identifying the independent (input) and dependent variables (output), need to perform exploratory data analysis (EDA) and data preprocessing to enhance the dataset. The better understanding of the data is necessary in order to remedy issues of missing values, spelling errors, and standardizing values such that they transform the data well enough for model creation. To apply the ML model, the data has to be divided into training and testing sets. The training set is used to build a predictive model, and the testing set is used to make a prediction. To implement a machine learning model, the LogisticRegression algorithm has been chosen because it is a classification algorithm used to allocate data to a discrete set of classes and is also suitable for multi-linear function classification problems. Problem formulation of LogisticRegression is given in equation (3.1) as

$$f(x_1, x_2) = b_0 + b_1 x_1 + \ + b_n x_n \tag{3.1}$$

Algorithm 1: Trustworthy Recommended Weighted (TRW)_cloud broker

Model Creation
Input: file directory of dataset (filepath)
Output: model
foreach *filepath in* **do**
  dataset ← $Dataframeof CSV$
  dataset ← $NULLvaluesreplacedas0$
  X ← $dataset.iloc[:, :-1].values$
  y ← $dataset.iloc[:, -1].values$
  X ← $Xislabelencoded$
  X_train,X_test,y_train,y_test ← $train\_test\_split(X, y, test\_size = 0.2, random\_state = 0)$
  classifier ← $LogisticRegression(random\_state = 0)$
  classifier.fit(x_train,y_train)
  model ← $y\_pred \leftarrow classifier.predict(X\_test)$
end
Testing the Model
Input: file directory user input
Output: services
foreach *filepath in* **do**
  dataset ← $Dataframeof CSV$
  dataset ← $NULLvaluesreplacedas0$
  initialize i to 0
  X ← $dataset.iloc[:, :-1].values$
  TRWv(services ← $model[i].predict(X)$)
  accuracy_score(y_test,y_pred)
end
Determine the Trustworthy Services
Input: EVA, SPVA
Output: Trustworthy Recommended Weighted value (TRWv)
Determine End-user feedback/ratings value average (EVA)
Determine Cloud service providers capabilities value average (SPVA)
ENTRIES (X): ← Enduser and CSPs ratings for each service
EVA = X.mean()
SPVA = X.mean()
TRWv = EVA [ ] *70% + SPVA [ ]*30%
Input: Trustworthy Recommended Weighted value (TRWv) for each service
Output: to determine Trustworthy service
initialize i to 0
Def *TRWv(predictedservices)*:
  if *TRWv[services][i] >=7.0* **then**
  return services
  else
    model[i].predict(services)
end

where b_0, $b_1 \ldots b_n$ are estimators or predicted weights and $x_1 \ldots x_n$ are number of inputs. Predicted probability of logistic regression function in equation (3.2)

$$p(x_1, x_2) = 1/(1 + exp(-f(x_1, x_2))) \tag{3.2}$$

It's usually close to 0 or 1. The expected probability that the outcome for a given x being 1 is typically

Table 3.1: Dataset for Proposed Model

Cloud Service Provider (CSP)	Model Creation		Trustworthy Features	
	Services(S)	Features Count	Trustworthy Features	Features Count
CSP1 CSP2 CSP3 CSP4 CSP5 CSP6 CSP7 CSP8 CSP9 CSP10	Services1 Services2 Services3 Services4 Services5 Services6	174	CSP Capability /Performance	22
			End users Ratings	54
10CSP		174F		76F

interpreted as the function $p(x)$. As a result, 1-$p(x)$ represents the probability that the output will be zero. To obtain the optimal weights, Maximize the log-likelihood function (LLF) for all occurrences $i = 1 \dots n$. Maximum likelihood estimation is the method, which is represented by the following equation (3.3).

$$LLF = \sum_{i} (ylog(p(x)) + (1 - y)log(1 - p(x))) \tag{3.3}$$

Logistic regression is used to determine the best predicted weights b_0, $b_1 ... b_n$, and the function $p(x)$ is as near as feasible to all real answers y_i , $i = 1...n$, where n is the number of occurrences. The LLF for the corresponding observation is equivalent to log(1- p(x_i)) when y_i=0. Whenever p(x_i) reaches $y_i = 0$, log(1- p(x_i)) becomes 0.

In our dataset, it has feature values of between 0 and 1. Although it has 60 services, 10 cloud service providers (CSP1, CSP2, CSP3 ... CSP10), each CSP offers 6 different services such as computing, storage,networking, management, security and performance which are denotes as Services1, Services2 and so on, 300 end users, and 250 features are taken to implement the problem. The entire dataset of the proposed framework model is shown in Tables 3.1. Even though it is a multi-linear function classification problem, each iteration acts as a single linear classification problem. Because with each iteration, it finds the most suitable service provider for the services based on the end user's requirements.

Once the model creation is successfully done, it will predict the result by checking the testing data. The model creation is done by 174 features, which includes the services features. Depending on the end user requirements, the TRWcloud broker algorithm predicts the outputs, which are shown in Figures 3.2 and 3.3. The end user has given the 174 features a value in the form of 0 or 1 while testing the model. The input samples are given in Figure 4.1 (ML Model Creation Features), which represents the end user input which varies from others in the value of 0 or 1. The predicted outputs are in the form of single array entities, which are shown in equation (3.4) for user Input1. It is plotted in the graph using a Python simulation, which is also presented in Table 3.3.

$$[CSP2, CSP5, CSP10, CSP4, CSP8, CSP1] \tag{3.4}$$

The proposed model output is interpreted as follows: The following CSPs are considered as trustworthy, recommended cloud services depending on the end user requirements in the form of features. These are the services that are returned by the TRW_cloud broker, which are represented in Figure 3.2 and Table 3.2 as, CSP2 provides Services1 (computing), CSP5 provides Services2 (storage), CSP10 provides Services3 (networking), CSP4 provides Services4 (management), CSP8 provides Services5 (security) and CSP1 provides Services6 (performance). Similarly, Figure 3.3 is predicted by the TRW_cloud broker algorithm for other user input2. The last part of the algorithm is to find out the trustworthiness of the services, which are returned by the service providers. Trustworthy Recommended Weighted value (TRWv) is used to find the trust value for each service. If the TRWv values for the services are greater than the threshold value, they are expected to be trusted as services. Otherwise, depending on the end user requirements, the ML model has to find the next

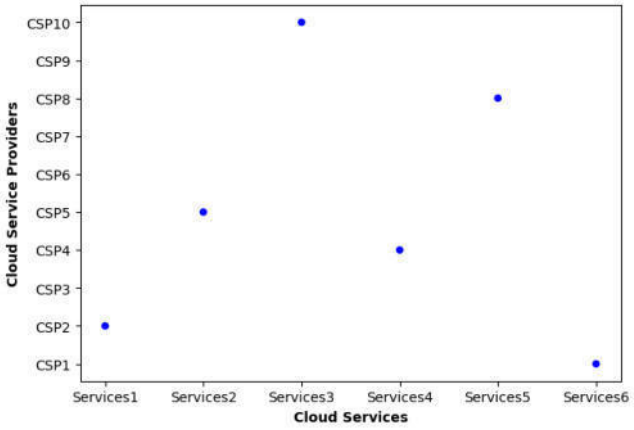


Fig. 3.2: Proposed Model Prediction for user Input1

Table 3.2: Predicted Output for user Input1

CSPs	Services1	Services2	Services3	Services4	Services5	Services6
CSP1						
CSP2						
CSP3						
CSP4						
CSP5						
CSP6						
CSP7						
CSP8						
CSP9						
CSP10						

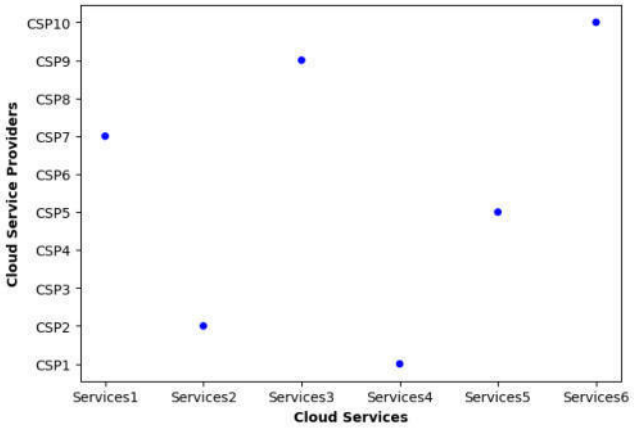


Fig. 3.3: Proposed Model Prediction for user Input2

cloud service until it gets a trusted service. The following sections elaborate on the features and accuracy of the results compared with other formworks that are discussed in the Section 2.

Fig. 4.1: Proposed framework features

4. Feature comparison with other frameworks. In the literature section, various frameworks proposed by the authors have already been discussed along with their advantages and observations. All the research articles that have shown performance/accuracy/efficiency are good compared to a few other frameworks. As a results of the availability of cloud services in a real multi-cloud environment with dynamic nature, context and location, time and trust measure parameters. However, in our proposed framework, identified trust measure parameters or features are evaluated by industry experts as well as a large number of features have been included during the research Figure 4.1. Table 4.2 presents a list of trustworthy features considered in our approach and excluded from previous research in the literature.

It includes ML model creation features, CSP capabilities features and end user feedback/reviews/ratings on parameters such as auto scaling, application logging patching, cost effective, automated OS, firewalls, elastic computing, advanced block storage and data repository. For SP capabilities performance, SLA, global infrastructure, pricing, security and audits and compliance etc. are identified. Similarly end user ratings have the following features such as cost efficiency, scale up, disaster recovery, on demand resources, full integrated and safety and security. All features are extracted in a real environment, and therefore the accuracy of the result is high compared to other frameworks. The framework integrated both the machine learning model and the Trustworthy Recommended Weighted value (TRWv) approach to provide trustworthy services to end-users. The following Table 4.1 highlights the features or trust parameters used in this article and compared them with other frameworks, to show how trustworthiness is improved in the proposed trustworthy recommendation framework.

Table 4.1: Comparison with other frameworks

Authors	Features	Techniques	Accuracy and efficiencey	Accuracy and No of User's involvement and Services	Weightage
Liu et al. 2019 [22]	response time	Trust-aware recommender systems	Mean Absolute Error	339 users 5825 web services	equal priority is given to CSP and users
Smithamol et al. 2018 [30]	response time, latency, and failure. throughput, CPU percentage, and network bandwidth.	Combining subjective and objective assessment	throughput and response time	142 users 4500 services	70% importance to objective trust, whereas 30% is subjective trust.
Somu et al. 2017 [15]	QoS attributes trust measure parameter (TMP) include 12 features such as availability, security , price etc.	Rough set: hypergraph-based trust measure parameter selection technique	To enhance the accuracy of the services, computation time play a vital role	10080 feedbacks that relates to QoS parameters. 7000 customers and 114 cloud services	Service selection techniques are verified in terms of size, ranking and time complexity.
Jayapriya et al. 2016 [19]	response time from web services and throughput from users	Correlated QoS Ranking Methodology	response time and availability of services Normalized Discounted Cumulative Gain (NDCG)	300 distributed users 500 web services	Priority is given to similar users.
Qu et al. 2015 [18]	3 subjective attributes such as privacy, response time etc. and 2 objective attributes response time and CPU performance	Creditability evaluation approach	Location and time	300 ordinary customers and 59 real cloud services	Priority is given to identifying the users' credibility to assess the cloud service truthfully
Kanwal et al. 2014 [20]	confidentiality, integrity, access control and authentication	Trust Evaluation Model (TEM)	Trust-score	2 web services 3 CSP (cloud node) 6 users 30 questionnaires	equal priority is given to CSP and users (Average of both scores has taken into the final trust value)
Zheng et al. 2013 [28]	response time, throughput, failure probability etc.	Kendall Rank Correlation Coefficient (KRCC)	QoS ranking prediction (performance)	300 distributed users 500 web services	Similar users are given weightage to rank services.
Proposed TRF	250 features to build a TRF (52 features from the client side and 24 features from the server side)	Machine learning model and TRWv approach	confusion matrix	300 users 60 services 10 CSPs	70 % priority to users and 30 % priority given to CSPs.

5. Results and Discussion. This section, elaborately discussed TRWv value and accuracy compared with other frameworks. To calculate or determine the TRWv, end-user ratings and service provider capabilities ratings are taken as an input. To measure EVA and SPVA, the end-user rating value average and service provider capabilities value average are considered. Using equations (5.1) and (5.2), each service's EVA and SPVA can be calculated. For example, CSP1 EVA and SPVA values for a service1 are 9.26 and 9.62 respectively. Using equation (5.3), the TRWv value for the CSP1 is computed as 9.368. Similarly, using the above equations, the CSPs for other services can be calculated.

Table 4.2: Trustworthy Features

	Trustworthy Features/Factors	Our Methodology	Other work in the literature
Trustworthy Features (Service Provider Capability and CC reviews)	Retention Time, Audits and Compliance (HIPAA,PCI,ISO,CSA), Global Infrastructure, Data Center Zones and Location, Disaster recovery, Features for enterprise, BI analytics support,Innovation features, Increased efficiency and flexibility,Comprehensive and reliable, Networking setup, Indispensable speed, Performance above expectation, Cloud managed stocks and inventory, Security and compliance experts, Infrastructure support, hosting, Migration support, Cloud platform for audits, VPN hosting, Cost saving and scalability, Cloud ecosystem, Native integration with common OS and Forecasting	Yes	No

Table 5.1: TRWv value for each service

CSPs	Services1	Services2	Services3	Services4	Services5	Services6
CSP1	9.368	7.602	9.252	9.174	8.74	9.064
CSP2	9.064	9.416	9.76	9.176	8.46	9.064
CSP3	9.062	9.064	7.94	9.3	6.46	9.7
CSP4	8.902	8.322	9.7	9	8.21	9.52
CSP5	9.934	9.4	9.856	9.52	7.514	9.188
CSP6	9.23	9.096	9.398	9.188	8.862	9.062
CSP7	9.062	9.3	7.94	9.3	8.46	7.628
CSP8	6.828	8.342	8.608	8.714	7.628	9.332
CSP9	9.332	8.96	6.392	7.98	7.92	9.062
CSP10	9.368	8	8.622	9.2	8.826	6.94

The TRWv for a services are computed for all the CSP and given in Table 5.1, which represents the TRWv value or score for all the services that are available in the dataset (60 cloud services, 10 CSP). From Table 5.1, the proposed framework provides trusted services to the end user because the threshold values are greater than 7.0 for most services, except a few. The untrusted services are provided by CSP3 (services5), CSP8 (services1), CSP9 (services3), and CSP10 (services6), and the proposed framework repeats the process until the end user receives a trustworthy service.

$$EVA = (Sum\ of end\ user\ ratings)/(No\ of\ Features) \tag{5.1}$$

$$SPVA = (CSPs\ Capabilities\ ratings)/(No\ of\ features) \tag{5.2}$$

$$TRW_v\ score = EVA\ * 70\% + SPVA * 30\% \tag{5.3}$$

A confusion matrix (CM) is used to determine the accuracy of the suggested proposed model. The outcome of the CM is represented as Table 5.2. From the above results, true negative is 120, true positive is 148, false positive is 22 and false negative is 16. Using the below equations (5.4, 5.5, 5.6 and 5.7), accuracy and F1 score of the proposed model is computed.

$$Accuracy = (TP + TN)/N = 0.913 \tag{5.4}$$

$$F1score = 2 * (recall * precision)/(recall + precision) \tag{5.5}$$

Table 5.2: Confusion Matrix

	Predicted (False)	Predicted (True)	Total
Actual (False)	TN = 120	FP=22	142
Actual (True)	FN= 16	TP=148	158
	136	164	

Table 5.3: Accuracy comparison with other framework

S.No	Author	Framework	Accuracy
1.	Liu et al. 2019 [22]	Trust-aware recommender systems	0.908
2.	Smithamol et al. 2018 [30]	TMM	0.90
3.	Somu et al. 2017 [15]	RSHT	0.88
4.	Jayapriya et al. 2016 [19]	CorQoSCloudRank	0.90
5.	Qu et al. 2015 [18]	CCCloud	0.904
6.	Noor et al. 2014 [25]	TCSR	0.88
7.	Zheng et al. 2013 [28]	KRCC	0.894
8.	Proposed - Trustworthy Recommendation Framework(TRF)	LogisticRegression and TRWv approach	0.913

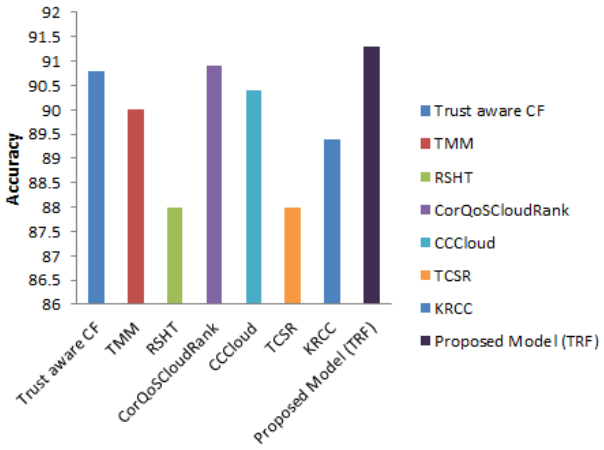


Fig. 5.1: Accuracy comparisons

$$Recall = TP/(TP + FN) = 0.923 \tag{5.6}$$

$$Precision = TP/(TP + FP) = 0.904 \tag{5.7}$$

$$F1\ score = 0.9134 \tag{5.8}$$

The proposed framework model's Accuracy and F1 score are respectively 91.3 and 91.34. The following Table 5.3 presents the accuracy comparison with other framework models and Figure 5.1 depicts it graphically.

The confusion matrix is used to evaluate the classification accuracy in order to validate the model. If the accuracy is 90% or more, there is a good confidence in the results that the model gives. Otherwise, parameter

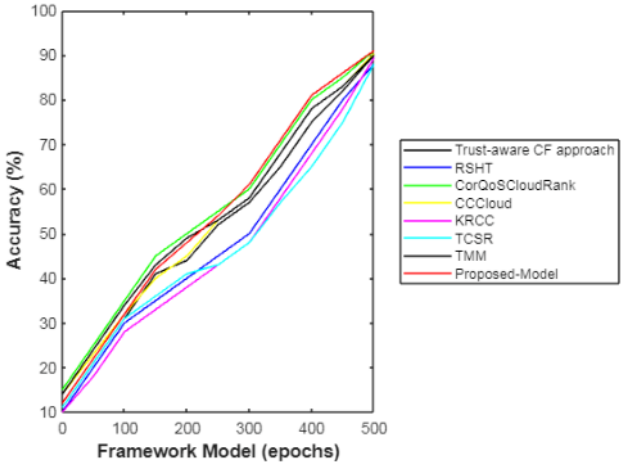


Fig. 5.2: Proposed framework Accuracy comparisons

tuning should be performed to increase accuracy of the results in the testing environment. The default iteration for the testing model is 100, but in our cases the maximum number of iterations taken. Because the accuracy of results automatically increases as the number of iterations or epochs increases, this is also depicted in Figure 5.2. After 450 training epochs, the proposed algorithm and the model's accuracy utilizing a variety of epochs start to converge. Depending on the end user requirements, the proposed model will provide an optimal, trustworthy service to the end user. If the model predicts the same output for more than one service, the priority is given to the first service, or based on the past history of services, the output service will be predicted. The entire approach has been implemented using LogisticRegression and TRWv approach. Since it is an automated ML model, features and end users can be increased at any time. The proposed trustworthy framework model has been implemented using Python 3.7 environment and the whole dataset has been submitted to the Github repository.

This section also compares the performance of the proposed methodology with other models described in the literature review section. the proposed framework model offers an optimal and reliable service to the end user, as the model has trained almost 250 features. The model not only provides the optimal services to the end users , it also provides trustworthy services, because 52 features are identified from end users and 24 features from service providers to enable the recommendation framework to provide trusted services. In addition, the proposed trustworthy recommendation framework model outperforms the other framework model Figure 5.2. Since the proposed framework is intended for real time-environment and provides the best and most reliable services to its end users, also the model would reduce estimation time for an expanding number of features and provide accurate cloud services to the end users, so financial constraints are irrelevant in this case. When compared to other frameworks discussed in the related work, the proposed trustworthy recommendation framework has a good performance or efficiency.

6. Conclusion and Future enhancement. It is a tedious task to determine a trustworthiness of cloud services depending on their CSPs capabilities along with end user requirements. The researchers also worked hard to develop a framework model for selection of cloud services as a significant solution for end users/cloud consumer. Cloud service provider can also have many cloud services with similar features. For end users, it is necessary to ensure their optimal services are trustworthy. Our proposed Trustworthy recommendation framework (TRF) model provides the optimal solution for the end user to identify the suitable trust services in a cloud environment. To check the trustworthiness of the cloud services, Trustworthy recommended weighted value (TRWv) approach is implemented to measure the TRWv. In order to assess the performance of TRF, accuracy and F1 score are calculated using confusion matrix. The best framework cannot be inferred, due to dynamic, distributed, and non-transparent nature of the service provider's availability in a federated cloud

environment. Furthermore, the framework model can incorporate maximum no of CSPs features to enhance the dynamic behaviour of the cloud environment.

REFERENCES

[1] M. Marimuthu, J. Akilandeswari and P. R Chelliah, *Identification of trustworthy cloud services: solution approaches and research directions to build an automated cloud broker*, Computing., 104 (2022), pp. 43–72.

[2] B. Alouffi, M. Hasnain, A. Alharbi, W. Alosaimi, H. Alyami and M. Ayaz , *A Systematic Literature Review on Cloud Computing Security: Threats and Mitigation Strategies*, IEEE Access., 9 (2021), pp. 57792–57807.

[3] Somula. Rama Subbareddy and R. Sasikala,*A survey on mobile cloud computing: mobile computing+ cloud computing (MCC= MC+ CC)*, Scalable Computing: Practice and Experience., 19(4) (2018), pp. 309–337.

[4] R. Regaieg, M. Koubaa, Z. Ales and T. Aguili, *Multi-objective optimization for VM placement in homogeneous and heterogeneous cloud service provider data centers*, Computing., (2021), pp. 1–25.

[5] R. Hentschel and S. Strahringer, *A broker-based framework for the recommendation of cloud services: a research proposal*, In Conference on e-Business, e-Services and e-Society, Springer, Cham., (2020), pp. 409–415.

[6] M. Yang , V. S. Jacob and S. Raghunathan, *Cloud Service Model's Role in Provider and User Security Investment Incentives* , Production and Operations Management., 30(2) (2021), pp. 419–437.

[7] Somula.Ramasubbareddy, Anilkumar. Chunduru , Venkatesh. B ,Karrothu. Aravind , Kumar.Pavan and Sasikala. R,*Cloudlet services for healthcare applications in mobile cloud computing*, proceedings of the 2nd international conference on data engineering and communication technology.,(2019), pp. 535–543.

[8] A. Hussain, J. Chun and M. Khan, *A novel framework towards viable cloud service selection as a service (cssaas) under a fuzzy environment*, Future Generation Computer Systems., 104 (2020), pp. 74–91.

[9] R. Djiroun, M. A. Guessoum, K. Boukhalfa and E. Benkhelifa, *A novel cloud services recommendation system based on automatic learning techniques*,International conference on new trends in computing sciences (ICTCS) IEEE., (2017), pp. 42–49.

[10] H. Mezni and T. Abdeljaoued, *A cloud services recommendation system based on Fuzzy formal concept analysis*, Data & Knowledge Engineering., 116 (2018), pp. 100–123.

[11] C. Mao, R. Lin, C. Xu and Q. He, *Towards a trust prediction framework for cloud services based on PSO-driven neural network*, IEEE Access., 5 (2017), pp. 2187–2199.

[12] P. S Pawar, M. Rajarajan, T. Dimitrakos and A. Zisman, *Trust assessment using cloud broker*, In IFIP International Conference on Trust Management. Springer, Berlin, Heidelberg., (2014), pp. 237–244.

[13] M. Usha, J. Akilandeswari and A. S Fiaz, *An efficient qos framework for cloud brokerage services*, International Symposium on Cloud and Services Computing, IEEE., (2012), pp. 76–79.

[14] M. Tang, X. Dai, J. Liu and J. Chen, *Towards a trust evaluation middleware for cloud service selection*, Future Generation Computer Systems., 74 (2017), pp. 303–312.

[15] N. Somu, K. Kirthivasan and V. S. Sriram, *A rough set-based hypergraph trust measure parameter selection technique for cloud service selection*, The Journal of Supercomputing., 73(10) (2017), pp. 4535–4559.

[16] T. H. Noor and Q. Z. Sheng, *Web service-based trust management in cloud environments*, In Advanced Web Services. Springer., (2014), pp. 101–120.

[17] R. Devi and R. Shanmugalakshmi, *Cloud providers ranking and selection using quantitative and qualitative approach*, Computer Communications., 154 (2020), pp. 370–379.

[18] L. Qu, Y. Wang, M. A. Orgun, L. Liu., H. Liu and A. Bouguettaya, *CCCloud: Context-aware and credible cloud service selection based on subjective assessment and objective assessment*, IEEE Transactions on Services Computing., 8 (3) (2015), pp. 369–383.

[19] K. Jayapriya, N. A. B. Mary and R. S. Rajesh, *Cloud service recommendation based on a correlated QoS ranking prediction*, Journal of Network and Systems Management., 24(4) (2016), pp. 916–943.

[20] A. Kanwal, R. Masood and M. A. Shibli, *Evaluation and establishment of trust in cloud federation*, In Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication., (2014), pp. 1–8.

[21] Somula. Rama Subbareddy and R. Sasikala,*RTTSMCE: a response time aware task scheduling in multi-cloudlet environment*, International Journal of Computers and Applications., 43(7) (2021), pp. 691–696.

[22] J. Liu and Y. Chen, *A personalized clustering-based and reliable trust-aware QoS prediction approach for cloud service recommendation in cloud manufacturing*, Knowledge-Based Systems., 174 (2019), pp. 43–56.

[23] Ramasubbareddy. Somula , Swetha. Evakattu , Luhach. Ashish Kumar and Srinivas T. Aditya Sai,*A Multi-Objective Genetic Algorithm-Based Resource Scheduling in Mobile Cloud Computing*, International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)., 15(3) (2021), pp. 58–73.

[24] S. M. Habib, S. Hauke, S. Ries and M. Muhlhauser, *Trust as a facilitator in cloud computing: a survey. Journal of Cloud Computing: Advances*, Systems and Applications., 1(1) (2012), pp. 1–18.

[25] T. H. Noor, Q. Z. Sheng, S. Zeadally and J. Yu, *Trust management of services in cloud environments: Obstacles and solutions*, ACM Computing Surveys., 46(1) (2013), pp. 1–30.

[26] R. R. Kumar, S. Mishra and C. Kumar, *Prioritizing the solution of cloud service selection using integrated MCDM methods under Fuzzy environment*, The Journal of Supercomputing., 73(11) (2017), pp. 4652–4682.

[27] T. H. Noor, Q. Z. Sheng, A. H. Ngu, A. Alfazi and J. Law, *Cloud armor: a platform for credibility-based trust management of cloud services*, ACM International Conference on Information Knowledge Management., (2013), pp. 2509–2512.

[28] Z. Zheng, X. Wu, Y. Zhang, M. R. Lyu and J. Wang, *QoS ranking prediction for cloud services*, IEEE transactions on

parallel and distributed systems., 24(6) (2012), pp. 1213–1222.
[29] X. Li, H. Ma, F. Zhou and W. Yao, *T-broker: A trust-aware service brokering scheme for multiple cloud collaborative services*, IEEE Transactions on Information Forensics and Security., 10(7) (2015), pp. 1402–1415.
[30] M. B. Smithamol and S. Rajeswari, *TMM: Trust Management Middleware for Cloud Service Selection by Prioritization*, Journal of Network Systems Management., 27(1) (2019), pp. 66–92.
[31] Somula. Rama Subbareddy and R. Sasikala,*A load and distance aware cloudlet selection strategy in multi-cloudlet environment*, International Journal of Grid and High Performance Computing (IJGHPC)., 11(2) (2019), pp. 85–102.
[32] A. Al-Faifi, B. Song, M. Hassan, A. Alamri and A. Gumaei, *A hybrid multi criteria decision method for cloud service selection from Smart data*, Future Generation Computer Systems., 93 (2019), pp. 43–57.
[33] Y. Wang, J. Wen, W. Zhou, B. Tao, Q. Wu and Z. Tao, *A cloud service selection method based on trust and user preference clustering*, IEEE Access., 7 (2019), pp. 110279–110292.
[34] F. Nawaz, M. R. Asadabadi, N. K. Janjua, O. K. Hussain, E. Chang and M. Saberi , *An MCDM method for cloud service selection using a Markov chain and the best-worst method*, Knowledge-Based Systems., 159 (2018), pp. 120–131.
[35] S. Chauhan, E. S. Pilli and R. C. Joshi, *BSS: a brokering model for service selection using integrated weighting approach in cloud environment*, Journal of Cloud Computing., 10(1) (2021), pp. 1–14.
[36] Somula. Rama Subbareddy and R. Sasikala,*Round robin with load degree: an algorithm for optimal cloudlet discovery in mobile cloud computing*, Scalable Computing: Practice and Experience., 19(1) (2018), pp. 39–52.
[37] Somula. Rama Subbareddy and R. Sasikala,*A honey bee inspired cloudlet selection for resource allocation*, smart intelligent computing and applications., (2019), pp. 335–343.
[38] Ramasubbareddy. Somula, Ramasamy. Sasikala , Sahoo. Kshira Sagar , Kumar R. Lakshmana , Pham. Quoc-Viet and Dao. Nhu-Ngoc,*Cavms: application-aware cloudlet adaption and vm selection framework for multicloudlet environment*, IEEE Systems Journal., 15(4) (2020), pp. 5098–5106.
[39] A. B. Saxena, D. Sharma and D. Aggarwal, *Trust Prediction Tree: Use of Decision Tree in Cloud Scenarios*, International Journal of Mechanical Engineering., 7(5) (2022), pp. 271–278.
[40] Dataset uploaded, https://github.com, [Accessed: 01-July-2022]
[41] Cloud broker, *https://en.wikipedia.org/wiki/Cloud-broker*, [Accessed: 01-July-2022]

Edited by: Vinoth Kumar
Received: Jul 19, 2022
Accepted: Nov 1, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 321–338, DOI 10.12694/scpe.v23i4.2037

FAULT TOLERANT LOAD BALANCING WITH QUADRUPLE OSMOTIC HYBRID CLASSIFIER AND WHALE OPTIMIZATION FOR CLOUD COMPUTING

SOUNDARARAJAN ANURADHA *AND P. KANMANI †

Abstract. Cloud Computing (CC) environment is developing as a recently discovered caliber for computing applications over the network. Fault tolerance is one of the foremost issues in CC environment. Since the negligence in resource have a profound effect on job execution, throughput, response time and performance of the entire network. In this work, in order to address the issue, Quadruple Osmotic Hybrid Classification and Whale Optimization (QOHC-WO) is introduced to fault-tolerance under the requirement of different user request tasks. Initially, Quadruple Fault Tolerance Level is applied to allocate the fault tolerance level. Followed by, Hybrid Vector Classifier is used to categorize the user request tasks (task) and cloud server nodes (node). Then, the Osmotic function is employed for performing the migration among virtual machines with lesser response time. This helps to solve the maximum level of fault issue. Finally, Improved Whale Optimization Algorithm is applied to find the optimal allocation of tasks with the corresponding node. In addition, the Bandit function and Whale optimization are used to address the trade-off between exploitation and exploration. Experimental setup of the proposed QOHC-WO method and existing methods are carried out with different factors such as task response time, the number of VM migrations, and percentage of fault detected rate with respect to a number of tasks. The analyzed results validate that the proposed QOHC-WO method achieves a higher fault detection rate with minimum response time as well as task migration than the state-of-the-art methods.

Key words: Cloud Computing, Quadruple, Quadruple Fault Tolerance Degree, Osmotic, Hybrid Vector Classifier, Bandit, Whale Optimization.

AMS subject classifications. 68M14

1. Introduction.

With the increasing number of cloud providers, the scheduling issue has resulted in a hotspot in the CC environment. The scheduler must arrange the scheduling mechanism in an effective manner to use the available resources and allocate them to the requested tasks. Hence, fault tolerance in load balancing is one of the principal threats in CC, that is necessary to allocate the user request task uniformly beyond all the cloud server nodes, detect the fault and eliminate fault from the network and share user request tasks to all the cloud server nodes to improve the cloud network performance.

In [1], Crystal, a server-centric and recursively constructed structure was designed. It was made from the building blocks via vertex configuration pattern of Archimedean solids. It was also found to be intrinsically fault-tolerant due to the incorporation of multi-homing mechanism. Scaling was also doubled in an exponential manner due to its design of architecture in a recursively manner. Besides, a two routing algorithms were also designed to improve routing efficiency and administer different types of failures in the network.

Simulation results revealed that Crystal was found to be highly fault-tolerant with high throughput and also provided low latency. However, the capabilities of the server were not taken into consideration, therefore compromising the overall completion time. Objective of this work is to introduce and evaluate the proposed user request task scheduling and cloud server node load balancing by introducing Osmotic Hybrid Vector Classifier model considering the capabilities of each virtual machine (VM), therefore improving the overall completion time.

An energy-efficient fault-tolerant replica management policy was designed in [2] considering the deadline and budget constraints. Simulation results revealed that the dynamic replica placement algorithm efficiently minimized the mean job time minimized the utilization of the network bandwidth and enhanced the storage

*Department of Computer Science and Applications, Shri Sakthikailassh Women's College, Salem, Tamilnadu, India (anusuriyaa@gmail.com)

†Department of Computer Science, Tiruvalluvar Government Arts College, Rasipuram, Tamilnadu, India (pushpakanmani@gmail.com)

space usage. With the consideration of energy effectiveness, a cluster scaling mechanism considering the energy effectiveness was presented to minimize the system energy consumption by means of sleeping and waking up the data nodes based on load state. In addition, to prevent access failures and data loss due to the node failures, a node failure recovery model according to the availability metrics was utilized for node failures.

Simulation results also revealed that the proposed method was found to be better in terms of both energy efficiency and fault tolerance. However, with a trade-off between exploitation and exploration, while choosing between replicas. To address this issue, in this work, Improved Whale Optimization algorithm is presented that addresses the issues of exploitation by means of Bandit function and exploration by means of Whale optimization, therefore enhancing the percentage of fault detection.

The main contributions of the paper are summarized as follows:

1. The main contribution of the proposed QOHC-WO method is to consider the capabilities of each VM and the task length to assign the tasks into the most appropriate VMs.
2. The novelty of Quadruple Fault Tolerance Level is introduced for assigning the fault tolerance level.
3. A novel Osmotic Hybrid Vector Classifier (OHVC) model is proposed where the classification of node and task is made. The new idea of the Osmotic function is used to perform a migration between virtual machines for reducing the task response time.
4. An Improved Whale Optimization algorithm is introduced for discovering the factor regarding the right assignment of the task to the node in an optimal manner. The novelty of the Bandit function and Whale Optimization is applied to solve the exploitation issue. This helps to enhance the fault tolerance rate.
5. Finally, a series of experiments were conducted to measure the performance analysis of the proposed QOHC-WO method along with conventional methods based on various performance metrics.

The rest of this paper is structured as follows. The related work is summarized in Section 2. Section 3 introduces the problem description, system mode, and then the novel Quadruple Osmotic Hybrid Classification and Whale Optimization (QOHC-WO) method for cloud computing is presented. Section 4 conducts extensive experiments to evaluate the performance of our algorithm followed by a discussion in Section 5. Section 6 concludes the paper.

2. Literature Survey. Customizable software systems comprises of critical, non-critical and interdependent configurations. Besides, two important factors, like, reliability and performance is heavily said to be influenced on completion of communication or interactions between the configurations. Besides, failure of critical configurations results in severe influence on both reliability and performance.

In [3], first critical configurations were first identified next a fault tolerant candidate was identified, therefore, resulting in the improvement of successful interactions. A Dynamic Fault Tolerant VM Migration (DFTM) method was designed in [4] for providing infrastructure reliability via advanced recovery mechanism. Here, Linear Programming model was used for route analysis and Alternative Switch Identification Algorithm was used to identify the new VM migration. Fault tolerant mechanism specifically, for storage and access was presented in [5] using priority-queue based scheduling, therefore resulting in the improvement of data access.

However, the high performance computing system results in higher failure rates due to huge involvement of servers filled with heavy workloads, resulting in the unavailability of systems for computation. An innovative perspective on applying a fault tolerant mechanism was presented in [6] that in turn eliminated the possibility of network congestion with migration technique to adaptively handle fault occurrences.

Yet another fault tolerant framework called Fault tolerance Algorithm Using Selective Mirrored Tasks Method(FAUSIT) was presented in [7]. Here, a balance between parallelism and the topology was first introduced, next a selective mirrored task method was utilized to enhance the fault tolerance, therefore both minimizing the make span and computing cost incurred. In[8], both the physical resource usage rate and fault tolerant requirement of the user were considered and guaranteed user service in large probability.

A survey on fault tolerance, fault tolerance methods, load balancing techniques and schemes in a CC environment was presented in [9]. Nature inspired meta-heuristic algorithm was used in [10] to solve the issues related to cloud load balancing. To achieve optimal container resource allocation, anew hybridized algorithm, called, Whale Random update assisted Lion Algorithm (WR-LA) was proposed in [11] to prove its superiority in terms of computational complexity. A novel Dynamical Load Balanced Scheduling (DLBS) approach was

presented in [12] for both improving the network throughput and balancing workload in a dynamical manner. Besides, efficient heuristic scheduling algorithms were also designed that in turn balanced data flows according to different time slots.

With several overloaded controllers in the network and only one controller with maximum load is usually said to be addressed by means of a single load-balancing operation, therefore improving the load-balancing efficiency. To address this issue, in [13], a load balancing strategy with multiple controllers based on response time was designed. By using the correct response time threshold and with multiple controllers in a simultaneous manner, load-balancing issue was said to be addressed. Dynamic reconfiguration of resources based on the real time requirements was designed in [14]. However, larger number of migration was said to take place. To address this issue, a game based consolidation model with energy and load constraints was designed in [15].

The cloud schedulers utilized considered only single resource (RAM) for workload coordination that in turn resulted in violation due to non-optimal VM placement. To address this issue, a nova scheduler was proposed in [16]that utilized multi-resource based VM placement, enhancing the application performance in terms of execution time. Yet another workflow job scheduling algorithm was designed in [17] on the basis of load balancing, therefore minimizing the system response time issue. A global league championship algorithm was used in [18] for scheduling of the task globally in cloud environment, therefore significantly minimizing the task execution time. However, the user's quality of service requirement was not satisfied. An improved ant colony algorithm was proposed in [19] to improve make span, throughput and average turnaround time for job scheduling. Performance comparison of heuristic algorithm and analysis of task scheduling were made in [20].

Structured Allocation-based Consistent Hashing (SACH) was introduced in [23] to enhance the load balancing of cloud infrastructure. But, the memory was not minimized. An elastic pull-based Dynamic Fault Tolerant (E-DFT) scheduling mechanism was introduced in [24] to reduce the response time. However, it failed to enhance efficiency. Energy and cost optimization mechanism was developed in [25] for increasing the fault tolerance and obtaining reliability. The designed mechanism failed to reduce the response time.

Based on the literature review presented above it is recognized that task response time and the number of VM migrations at the deployment of VM. Our proposed work is different from all aforementioned VM placement schemes as we have proposed Osmotic Hybrid Vector Classifier with Quadruple Fault Tolerance level. This in turn performs migration of VM based on an Osmotic function on occurrence of fault and therefore reduced the task response time. We empowered Improved Whale Optimization Algorithm to consider the optimization of task allocation to the corresponding node in an optimal manner to minimize VM migration.

3. Methodology. In this section, the proposed method of allocation of the task (i.e. user request task) to the corresponding node (i.e. cloud server node) with appropriate fault tolerance in distribution process in the CC environment is designed. Fig. 3.1 shows the sample cloud server fault tolerance in load balancing.

As depicted in the above figure, with n tasks submitted to the CC environment, with the n node being ready for allocating the required resources, a fault tolerance mechanism is introduced in the load balancer. With this fault tolerance mechanism smooth flow between the task and node is said to be accomplished in CC environment. With the above said basic design, a method, called, Quadruple Osmotic Hybrid Classification and Whale Optimization (QOHC-WO) is proposed with the objective of improving the task response time, number of VM migrations and the percentage of fault detected rate. The WOHC-WO is split into following steps, namely, user task request, transformation of fault tolerance levels, user request task and cloud server node classification, finally optimization. Figure 3.2 given below shows the architecture of the QOHC-WO method.

Figure 3.2 demonstrates the Block diagram of Quadruple Osmotic Hybrid Classification and Whale Optimization. Initially, the user task request users submit resource requests are considered in the CC environment. Next, according to the Quadruple Fault Tolerance Level as presented above, the fault tolerance levels of user request tasks are characterized and a correlation between the task and node is established that can regulate certain fault tolerance levels of user request tasks. The third step is the appropriate classification of user request task and cloud server. In our work, the user request tasks are classified based on the dependent task (DT) and the in dependent task (IT) and cloud server nodes are classified based on the nodes either overload (OL), under load (UL) or balanced (B) respectively by means of Osmotic Hybrid Vector Classifier. According to the specification of the tasks and load in the nodes, load balancing is said to be attained at the cloud server with higher level of fault tolerance. This is performed via Osmotic function that migrates between the VM

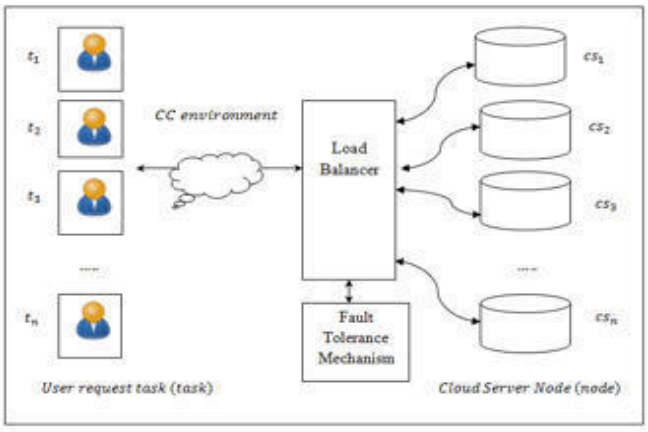


Fig. 3.1: Cloud server fault tolerance in load balancer

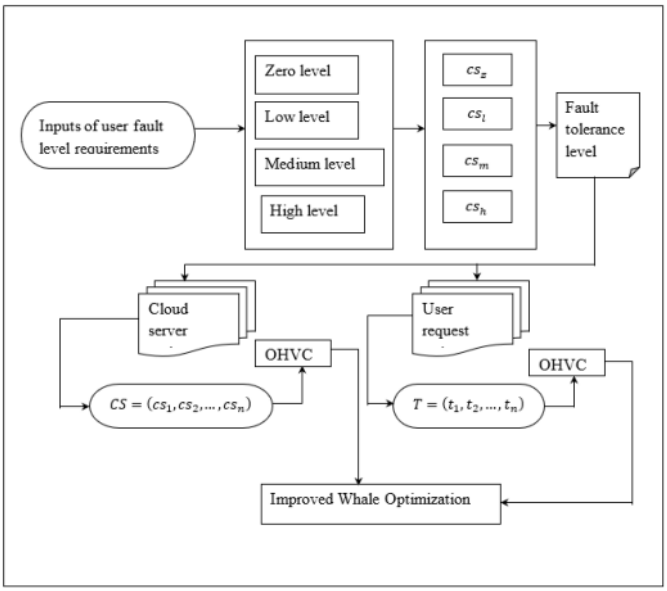


Fig. 3.2: Block diagram of Quadruple Osmotic Hybrid Classification and Whale Optimization

machine in case of fault and therefore handling higher level of faults. Finally, optimization of task assignment to node is determined using the Improved Whale Optimization algorithm, addressing exploitation using Bandit function and exploration using Whale function and hence called improved algorithm. To start with the problem description, followed by the system model and then the elaborate description of the proposed method is presented.

According to the specification of the tasks and load in the nodes, load balancing is said to be attained at the cloud server with higher level of fault tolerance. This is performed via Osmotic function that migrates between the VM machine in case of fault and therefore handling higher level of faults. Finally, optimization of task assignment to node is determined using the Improved Whale Optimization algorithm, addressing exploitation using Bandit function and exploration using Whale function and hence called improved algorithm. To start

with the problem description, followed by the system model and then the elaborate description of the proposed method is presented.

3.1. Problem description. The features of the requirements of several application programs in VMs to physical resources decide specific needs of the physical resources. Besides, the VM resources are specifically split into four types, namely, CPU, memory, I/O, and network resources. The significance of each resource differs with several applications settings of the user's. Hence, the physical machines are dispended to the VMs based on the discrete user demands that in turn utilize physical resource in an appropriate manner and enhance the user's service experience.

But, most of the current research work take into consideration only the user's individual requirements (i.e. user request task) but the question of the states of the node (i.e. cloud server nodes) and which task to be assigned to which node in an optimal manner with differences of individual demands of fault tolerance is said to be ignored. This results in the imperfection of appropriate fault tolerance in distribution process of virtual machine. Therefore, the distribution issue of virtual machines based on the states of the cloud server nodes (i.e., under load/overload/balanced) and the task dependencies (i.e. dependent task/ independent task) needs to be addressed imperatively.

3.2. System model. Let the user request task be represented as '$T = (t_1,\ t_2,\ \ldots, t_n)$' and the available cloud server nodes in the Cloud Computing environment denoted as '$CS =\ cs_1,\ cs_2,\ \ldots, cs_n$' with the maximal number of cloud server nodes represented as 'n'. Besides, the set of virtual machines utilized in the cloud server node 'cs_j' is represented as '$V_s = \{v_{j1},\ v_{j2},\ \ldots, v_{jn}\}$'.

3.3. Quadruple Fault Tolerance Degree. Fault tolerance is a main part of distributed system to guarantee the continuity and functionality of a system at a point where there's a fault or failure. In this work, the level of fault tolerance is divided in to four types. They are Zero Level Fault Tolerance, Low Level Fault Tolerance, Medium LevelFault Tolerance and High Level Fault Tolerance respectively and is represented as given below in the form of a quadruple

$$U = \{U_z,\ U_l,\ U_m,\ U_h\} \tag{3.1}$$

From the above equation (3.1), U_z relates to the zero levelfault tolerance, U_l relates to the low levelfault tolerance, U_m represents the medium level fault tolerance and U_h represents the high levelfault tolerance respectively. The user request task can select one levelin a random manner from the quadruple. However, specific fault-tolerance are not said to be operated directly. Therefore, in this work, the user's request task fault tolerance levelscorrelate to the states of the cloud server nodes (i.e. under load/overload/balanced). The fault tolerance of the cloud server nodes in this work is represented in the form of quadruple and is written as given below.

$$CS = \{cs_z,\ cs_l,\ cs_m,\ cs_h\} \tag{3.2}$$

Then the level of quadruple of the nodes 'CS' and the quadruple of the task corresponds with each other. For example, a task assigns the zero level fault tolerance as the requirement, then this degree corresponds to 'cs_z' in the nodes. In a similar manner, if a task assigns the primary level fault tolerance as the requirement, then this degree corresponds to 'cs_p' in the nodes and so on. The cases scenarios for the quadruples are given below.

Case 1: If the value of 'cs_h' is three, we need four nodes to provide service for the tasks. If any three nodes have failure and the CC environment can however furnish the service that meets the requirement of tasks, this condition correlates with high level fault tolerance of tasks.

Case 2: If the value of 'cs_m' is two, we require three nodes to provide service for the tasks. If any two nodes have downtime at the same time and the tasks can still be arrived at, this condition correlates with the medium level fault tolerance of tasks.

Case 3: If the value of 'cs_l' is one, we require two nodes to provide service for the tasks and this condition correlates with the low degree fault tolerance of tasks.

Case 4: If the value of 'cs_z' is zero, we require only one node to provide service for the tasks and this condition correlates with the zero level fault tolerance of tasks.

Therefore, we can measure both the quantity of the required nodes and the tasks. The quantity of node and the task is represented as 'NCS' and 'NT' respectively. Hence, the constraint association is represented as given below.

$$NCS \ \geq CS + 1 \tag{3.3}$$

$$NT \ \geq T + 1 \tag{3.4}$$

The level of fault tolerance submitted by the tasks is zero level fault tolerance. With the above said fault tolerance degree, the proposed method predict the status of node and task based on the response mechanism that integrates fault tolerance need and task constraints for solving the fault tolerance of VM.

3.4. Osmotic Hybrid Vector Classifier. In this section, the classification of task and the node is performed by means of Osmotic Support Vector Machine. As both the task and node are classified, this model is said to be a hybrid classifier model, called, Osmotic Hybrid Vector Classifier (OHVC). Support Vector Machine (SVM)[21] being a supervised machine learning technique classifies the given data. In this work, the classification of the tasks into dependent task and independent task and classification of nodes into under load, over load or balanced node is done.

Here, a kernel type of technique is followed that modifies the data and based on these modification, an optimal boundary between the probable outputs are said to be made. The main objective here remains in identifying the optimal separating hyperplane that maximizes the margin of the data (i.e., task and nodes). The hypothesized hyperplane equation is given below:

$$w^T x + b = 0, \ x \ \in CS, \ T \tag{3.5}$$

From the above equation (3.5), 'w' and 'x' symbolizes the vectors, with '$w^T x$' representing the dot products of the two different vectors. In this work, besides the vector 'x', includes both the cloud server nodes 'CS' and the task 'T' respectively. Let 'H_0' represent a hyperplane separating the data (i.e. cloud server node and the task) and satisfying the following equation.

$$wx + b = 0 \tag{3.6}$$

Let us select two other hyperplanes 'H_1' and 'H_2' and have the following equations

$$wx + b = 1 \tag{3.7}$$

$$wx + b = -1 \tag{3.8}$$

Then with the introduction of a constraint (y_i) always having the value of '-1', maximizes the margin, resulting in optimal hyperplane as given below.

$$Min \ in \ (w, b) \ \frac{w^2}{2} \tag{3.9}$$

$$Subject \ to \ y_i \left(wx + b\right), \ for \ i = 1, 2, .., n \tag{3.10}$$

From the result of the above value, the optimal hyperplane 'OH' is obtained. With the obtained result, Osmotic concept is used to address the issues related to migration. Osmosis refers to the uncontrolled variation of molecules from a higher to a lower water concentration. The objective behind the application of osmotic function in our work is to migrate between the virtual machines in case of the occurrence of fault. The osmotic function is represented as given below:

$$\pi = OFs = CFData_{inflow} \tag{3.11}$$

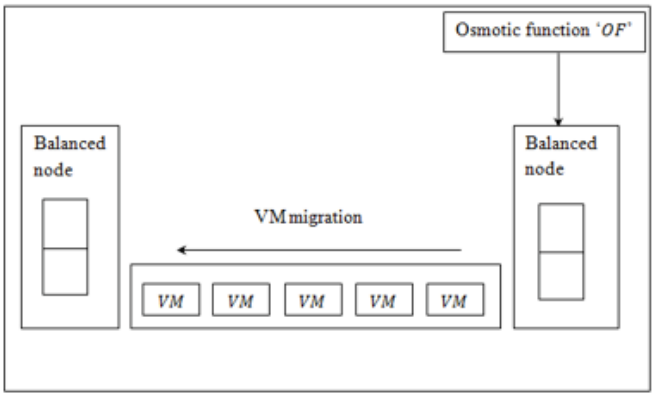


Fig. 3.3: The migration operation between VMs using Osmotic function

From the above equation (3.11), 'π' refers to the osmosis pressure law, with 'CF' representing the correction factor (i.e., migration between VMs) and '$Data_{inflow}$' refers to the inflow of data (i.e., either the task or the availability of nodes). Fig. 3.3 illustrates the migration operation between VMs using Osmotic function.

In CC environment each node (CS) has a different number of virtual machines (VMs). The set of all 'CS' in data center is '$CS = \ cs_1, \ cs_2, \ \ldots, cs_n$', where '$n$' refers to the number of 'CS' and the set of all VMs in datacenter is '$VM = VM_1, \ VM_2, \ \ldots, VM_n$' In Osmotic Hybrid Support Vector Machine, the hyperplane identified through the SVM measures the standard deviation 'σ' for each 'CS' to identify the under load 'UL', over load 'OL' and balanced 'B' nodes. This is said to be hybrid because of the classification of both node and the task performed via Osmotic SVM. This necessitates to identify the load of each 'CS' that depends on the load of 'VM' deployed into it. Then, the average load of the '$jth\ VM$' in '$ith\ CS$' (VM_{ij}) is written as follows:

$$VM_{ij} = LU_{CPU_j} + LU_{MEM_j} + LU_{BW_j} \tag{3.12}$$

From the above equation (3.12), 'LU_{CPU_j}', 'LU_{MEM_j}' and 'LU_{BW_j}' refers to the load utilization of the cpu, memory and the bandwidth of the 'jth' virtual machine respectively. The average load of '$CS_i\,(Avg_i)$' and the standard deviation for '$CS_i\,(\sigma_i)$' is written as follows:

$$AVG_i = \frac{\sum_{j=1}^{m} VM_j}{m} \tag{3.13}$$

$$CS_i\sigma_i = \sqrt{\frac{1}{n}(AvgLoad - AVG_i)^2} \tag{3.14}$$

From the above equation (3.14), 'n' refers to the number of all 'CS' and '$AvgLoad$' is the average load of all 'CS':

$$AvgLoad = \frac{1}{n}\sum_{i=1}^{n} AVGL_i \tag{3.15}$$

In a similar manner, the average task of the '$jth\ VM$' in '$ith\ CS$' (VM_{ij}) is written as follows:

$$VM_{ij} = \ Q_{IT} + Q_{DT} \tag{3.16}$$

From the above equation [16], 'Q_{IT}' and 'Q_{DT}' refers to the independent and dependent task queue of the 'jth' virtual machine respectively. In a similar manner, the standard deviation for '$T_i\,(\sigma_i)$' is written as follows:

$$T_i\sigma_i = \sqrt{\frac{1}{n}(AvgQueue - AVG_i)^2} \tag{3.17}$$

$$AvgQueue = \frac{1}{n}\sum_{i=1}^{n} AVGQ_i \tag{3.18}$$

From the above equation (3.18), 'n' refers to the number of all 'T' and '$AvgQueue$' is the average queue of all 'T'. The pseudo code representation of Osmotic Hybrid Vector Classifier is given below.

Algorithm 1: Osmotic Hybrid Vector Classification

Input: Task '$T = (t_1, t_2, \ldots, t_n)$', Cloud Server nodes '$CS = cs_1, cs_2, \ldots, cs_n$',

Output: Time-efficient hybrid classification

1: **Begin**
2: **Initialize** fault tolerance level, load 'L', queue 'Q'
3: **For** each task 'T' with cloud server nodes 'CS'
4: Obtain fault tolerance level using (1)
5: Obtain hypothesized hyperplane using (5)
6: Select three hyperplanes using (6), (7) and (8)
7: Obtain optimal hyperplane 'OH' using (9) and (10)
8: Measure osmotic function using (11)
//Node classifier
9: Measure average load of virtual machine using (12)
10: Measure average load and standard deviation of cloud server node using (13) and (14)
11: **If** '$\sigma_i < OH$' then 'CS_i' is under load **end if** call 'OF'
12: **If** '$\sigma_i > OH$' then 'CS_i' is over load **end if** call 'OF'
13: **If** '$\sigma_i = OH$' then 'CS_i' is balanced **end if**
//Task classifier
14: Measure average task of virtual machine using (16)
15: Measure standard deviation for task using (17)
16: **If** 'σ_i' has influence on queue 'Q' then 'T' is dependent task **end if** call 'OF'
17: **If** 'σ_i' does not has influence on queue 'Q' then 'T' is independent task **end if**
18: **End for**
19: **End**

As given in the above algorithm, to start with, the number of tasks (i.e. 500) to be assigned to the node, along with the queue, load and fault tolerance level is initialized. In this work, a quadruple fault tolerance level is used. Next, the average task and the average load in the node are analyzed. Followed by which the osmotic pressure is applied upon the occurrence of fault so that migration between the tasks and the node is said to take place, in order to achieve more cloud balanced system. Also as given in the above algorithm, the osmotic function 'OF' is applied at three different places. The first place, when overload is said to occur at the node, the second place, when under load is said to occur and the third place when the task is said to be dependent task. This is because of the reason that fault occurs during these three cases. In this way by applying osmotic

function virtual machines are said to be migrated across CC environment, with minimum completion time or task response time.

3.5. Improved Whale Optimization Algorithm. Finally, a scheduling algorithm is carried out to distribute the task and the available actual resources (i.e. file) from the node and deployment rules are applied to optimize data center. In this work, an Improved Whale Optimization (IWO) Algorithm is used with the objective of improving the percentage of fault being detected by means of exploitation-exploration tradeoff. The algorithm is called as improved because a Bandit function is used to the Whale Optimization so that the problem related to choosing between options is said to be reduced and therefore improving the fault detection rate. In other words, by combining the exploitation of Bandit with the exploration of Whale, promising candidate solution is said to be arrived at. The pseudo code representation of Improved Whale Optimization Algorithm is given below.

Algorithm 2: Improved Whale Optimization Algorithm

Input: Coefficient vector represented by 'E', random vectors 'rv_1' and 'rv_2'
Output: Optimized task-node allocation
1: **Initialize** current iteration 'CI', current task position vector 'V_p', position vector of the node in queue 'V'
2: **Begin**
3: **For** each Coefficient vector
4: Measure prey encircling (i.e., initialization of task and node in queue) using (19) and (20)
5: Evaluate positive vector using (21) and (22)
6: Perform searching of prey (i.e., appropriate task with respect to node) using (23) and (24)
7: Evaluate bandit function using (25)
8: **Return (optimal task-node allocation)**
9: **End for**
10: **End**

Figure 3.4 shows the flow process of Quadruple Osmotic Hybrid Classification and Whale Optimization. Initially, the number of user tasks and nodes as input is taken from the dataset. Next, the Quadruple Fault Tolerance Level is employed to find the fault tolerance level of data. Followed by, the Osmotic Hybrid Vector Classifier is utilized to measure the average load and standard deviation ' ' for each 'CS'.In addition, the node and task are classified by using this algorithm. Lastly, Improved Whale Optimization is employed to perform searching of prey depending on the bandit function for finding the optimal task-node allocation.

Whale optimization algorithm (WOA) is a new biological Meta-heuristic algorithm. The significant concept of meta-heuristic algorithms such as particle swarm optimization (PSO), genetic algorithm (GA) is used to achieve optimal solutions. But, it failed to find the local optimal solution with slow convergence speed and low convergence accuracy. In order to overcome the issue, the advantage of the Whale Optimization algorithm is easy to implement the process and robust. Whale Optimization is a population-based algorithm that simulates the social behavior of humpback whales. The population-based algorithm is used for avoiding the local optima and getting a globally optimal solution to handle the dissimilar constrained or unconstrained optimization issues. Whale Optimization is a population based algorithm that simulates the social behavior of humpback whales. The Whale Optimization in this work is used to optimally allocate the task to the corresponding node. This is performed by applying two rules to enhance the position of candidate solutions in each step that are encircling prey, search for prey and optimal prey allocation. They are described as follows by the following equations starting with prey encircling (i.e. initialization of task and node in queue).

$$D = [F.V_p(CI) - V(CI)] \tag{3.19}$$

$$V(CI+1) = V_p(CI) - E.D \tag{3.20}$$

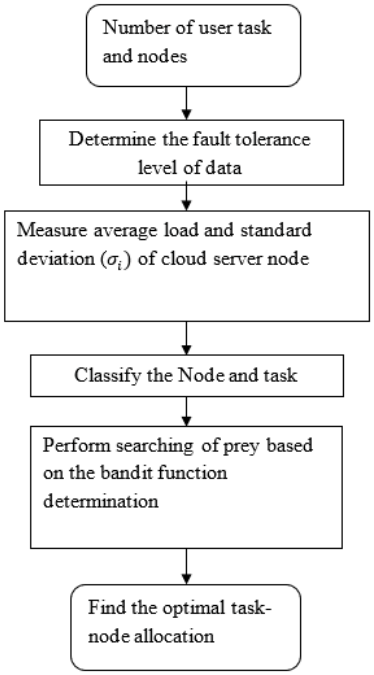


Fig. 3.4: Flow Process of Quadruple Osmotic Hybrid Classification and Whale Optimization

From the above equations (3.19) and (3.20) 'CI' refer to the current iteration with coefficient vector represented by 'E', 'F',current task position vector being 'V_p' and 'V' referring to the position vector of the node in queue 'Q'. The vectors 'E' and 'F' are measured as given below.

$$E = 2e.rv_1 - e \tag{3.21}$$

$$F = 2.rv_2 \tag{3.22}$$

From the above equations (3.21) and (3.22), 'rv_1' and 'rv_2' refers to the random vectors with elements 'e' decreased from '2' to '0' over certain iteration period. Next searching for prey (i.e. appropriate task with respect to node) is written as follows.

$$D = (f * t_{rand} - t) \tag{3.23}$$

$$V\,(CI + 1) = t_{rand} - E.D \tag{3.24}$$

From the above equations (3.23) and (3.24), 't_{rand}' refers to the random task position vector, therefore converging towards exploration. Finally optimal prey allocation with the solution towards exploitation is performed by applying the Bandit function. Here, with the utilization of Bandit function, exploitation is identified as a set of distributions '$CS = \{cs_1,\ cs_2,\ \ldots, cs_n\}$', each distribution being linked with the rewards delivered by one of the levers. Let '$\mu = \{\mu_1,\ \mu_2,\ \ldots, \mu_n\}$' represents the mean values linked with these reward distributions corresponding to each node. The node iteratively plays one lever per round and looks at the corresponding reward. The objective here remains in maximizing the sum of the collected rewards by the node. In a similar manner, the regret after rounds is said to be the anticipated difference between the reward sum

associated with an optimal strategy and the sum of the collected rewards. It is mathematically expressed as given below:

$$\rho = T\mu^2 - \sum_{CI=1}^{T} R_{CI}\,[CS] \tag{3.25}$$

From the above equation (3.25), the cloud server nodes 'CS' reward 'R' in round 'CI' is analyzed with the maximal reward mean being 'μ^2', ensuring fair exploitation of the cloud server node. In this manner, a trade-off between exploitation-exploration is said to be obtained resulting in maximal percentage of fault detected.

4. Experimental setup. In this section, the performance evaluation of the proposed method, Quadruple Osmotic Hybrid Classification and Whale Optimization (QOHC-WO), and three other existing methods, Crystal [1], energy-efficient fault-tolerant replica management policy [2], and Proactive fault tolerance framework (PFTF) algorithm [6] is implemented with CloudSim simulation using JAVA platform. Experiments are conducted with the aid of Personal Cloud datasets via Active Personal Cloud Measurement extracted from http://cloudspaces.eu/results/datasets. The number of 17 attributes such as row id, account id, file size (i.e. task size), operation_time_start, operation_time_end, time zone, operation_id, operation type, bandwidth trace, node_ip, node_name, quoto_start, quoto_end, quoto_total (storage capacity), capped, failed and failure info. These are utilized for performing efficient task scheduling to the multiple virtual machines in the cloud. The implementation is conducted with the hardware specification of Windows 10 Operating system, core i3-4130 3.40GHZ Processor, 4GB RAM, 1TB (1000 GB) Hard disk, ASUSTek P5G41C-M Motherboard, Internet Protocol. For accomplishing the experimental evaluation, the QOHC-WO considers a number of task in the range of 50-500 from the Personal Cloud datasets. The effectiveness of the method is evaluated in terms of task response time, the number of VM migrations, and the percentage of fault detected rate. The results of four different techniques are discussed with the aid of tables and graphical representation.

4.1. Performance evaluation of task response time. The response time for a task is the time between when it becomes active (e.g. an external event or occurrence of fault) and the time it completes. Several factors can cause the response time of a task to be longer than its execution time. In our work, the response time of a task is evaluated based on the standard deviation of server and the task respectively. It is measured as given below:

$$TRes_{time} = \sum_{i=1}^{n} T_i * Time\ \ [CS_i\sigma_i + \ T_i\sigma_i] \tag{4.1}$$

From the above equation (4.1), the task response time '$TRes_{time}$' is measured based on the number of tasks 'T_i', and the time consumed in obtaining the average node '$CS_i\sigma_i$' and obtaining the average task '$T_i\sigma_i$' respectively. It is measured in terms of milliseconds (ms). The sample calculation of task response time using the proposed QOHC-WO, existing Crystal [1] and energy-efficient fault-tolerant replica management policy [2] and PFTF algorithm [6]. is given below.

Sample calculations for task response time:

Proposed QOHC-WO: With '50' numbers of tasks considered for experimentation, '$0.035ms$' time consumed in calculating average nodes and '$0.012ms$' time consumed in calculating average tasks, the overall task response time is measured as given below.

$$TRes_{time} = 50 * [0.035ms + 0.012ms] = 2.35ms$$

Existing Crystal [1]: With '50' numbers of tasks considered for experimentation, '$0.043ms$' time consumed in calculating average nodes and '$0.017ms$' time consumed in calculating average tasks, the overall task response time is measured as given below.

$$TRes_{time} = 50 * [0.043ms + 0.017ms] = \ 3ms$$

Table 4.1: Comparison of task response time for different tasks

Number of tasks	Task response time (ms)			
	QOHC-WO	Crystal	Energy-efficient fault-tolerant replica management policy	PFTF algorithm
50	2.35	3	3.5	4.12
100	2.85	3.35	5	5.67
150	3.15	3.55	5.35	6. 23
200	3.55	3.95	5.55	7.05
250	4.35	4.95	6.15	7.56
300	5.25	5.35	7.35	8.02
350	5.55	6.16	8.25	8.89
400	5.95	6.35	9.15	9.35
450	6.25	6.85	10.25	10.21
500	6.66	7.15	11.45	11.33

Existing Energy-efficient fault-tolerant replica management policy [2]: With '50' numbers of tasks considered for experimentation, '0.048ms' time consumed in calculating average nodes and '0.022ms' time consumed in calculating average tasks, the overall task response time is measured as given below.

$$TRes_{time} = 50 * [0.048ms + 0.022ms] = 3.5ms$$

Existing Proactive fault tolerance framework (PFTF) algorithm [6]: With '50' numbers of tasks considered for experimentation, '0.0412ms' time consumed in calculating average nodes and '0.0412ms' time consumed in calculating average tasks, the overall task response time is measured as given below.

$$TRes_{time} = 50 * [0.0412ms + 0.0412ms] = 4.12ms$$

Table 4.1 summarizes the comparison of task response time for 500 different tasks presented in the nodes in CC environment. Task response time of QOHC-WO is compared with Crystal [1] and Energy-efficient fault-tolerant replica management policy [2] in table 1. In order to perform a fair comparison similar tasks were provided for three different methods.

Fig. 4.1 given above illustrates the task response time. The x axis in the above figure denotes the tasks considered for experimentation in the range of 50 to 500. The y axis in the above figure denotes the corresponding task response time observed using the proposed QOHC-WO and the existing [1] and [2] methods. The task here considered obtain from the active personal cloud measurement, varies with different files of different sizes.

Figure 4.1 given above illustrates the task response time. The x axis in the above figure denotes the tasks considered for experimentation in the range of 50 to 500. The y axis in the above figure denotes the corresponding task response time observed using the proposed QOHC-WO and the existing [1], [2] and [6] methods. The task here considered obtain from the active personal cloud measurement, varies with different files of different sizes. Hence, the task response time is not linearly increased. Besides, from the simulation conducted using CloudSim, for 50 numbers of tasks, the task response time using the proposed method was found to be 2.35 ms, 3 ms using [1], 3.5 ms using [2] and 4.12ms using [6]. From this it is inferred that the task response time is said to be comparatively lesser using the QOHC-WO method when compared to [1], [2]

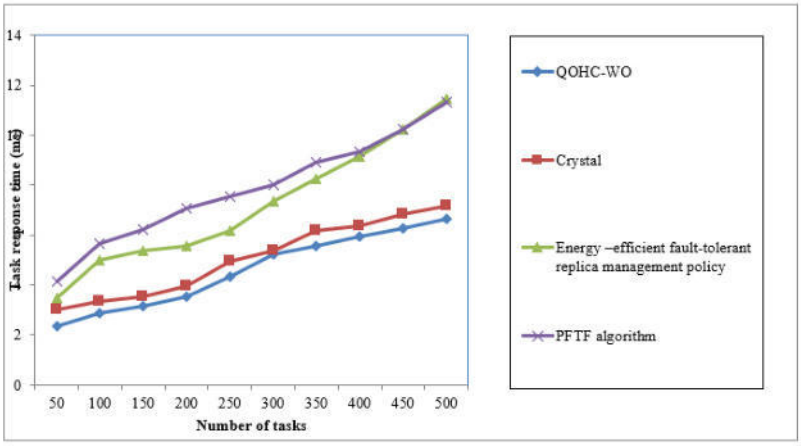


Fig. 4.1: Task response time

and [6]. This is because of the application of Osmotic Hybrid Vector Classification algorithm. By applying this algorithm, two main things are said to happen. First, the classification of node is node and then the classification of task is performed separately using the using the hybrid vector classification algorithm. With this, imperative classification is made via optimal hyperplane identification. Next, with the application of Osmotic function upon the occurrence of fault, efficient migration between the virtual machines is said to take place. Therefore, the task response time using the QOHC-WO method is reduced by 10% compared to [1], 36% compared to [2] and 42% compared to [6] respectively.

4.2. Performance evaluation of task migration. The task migration is obtained using the net migration rate [22]. The net migration rate refers to the difference between the number of immigrants (i.e., task entering into the node for request) and the number of emigrants (i.e., task leaving the node in search of another node) throughout the year (i.e. for an average task size).

$$T_{mig} = \frac{(T \rightarrow CS - T \leftarrow CS)}{T} * 1000 \tag{4.2}$$

From the above equation (4.2), the task migration 'T_{mig}' is obtained based on the difference between the task entering into the node for request '$T \rightarrow CS$' and the task leaving the node '$T \leftarrow CS$' to the overall task considered for experimentation 'T'. The sample calculation for virtual machine migration using the three methods, QOHC-WO, existing Crystal [1] and energy-efficient fault-tolerant replica management policy [2] is given below.

Sample calculation for VM migration

Proposed QOHC-WO: With '50' numbers of tasks considered for experimentation, '7' numbers of task entering the node and '4' numbers of task leaving the node, the overall task migration rate is given below.

$$T_{mig} = \frac{3+2}{50} * 100 = 10$$

Existing Crystal [1]: With '50' numbers of tasks considered for experimentation, '9' numbers of task entering the node and '7' numbers of task leaving the node, the overall task migration rate is given below.

$$T_{mig} = \frac{5+4}{50} * 100 = 18$$

Table 4.2: Comparison of task migration for different tasks

Number of tasks	Task migration			
	QOHC-WO	Crystal	Energy-efficient fault-tolerant replica management policy	PFTF algorithm
50	10	18	20	26
100	25	40	45	60
150	65	75	80	90
200	95	120	125	135
250	105	130	140	150
300	125	145	160	180
350	130	160	170	190
400	135	175	180	205
450	150	180	185	220
500	180	185	200	230

Existing energy-efficient fault-tolerant replica management policy [2]: With '50' numbers of tasks considered for experimentation, '9' numbers of task entering the node and '7' numbers of task leaving the node, the overall task migration rate is given below.

$$T_{mig} = \frac{6+4}{50} * 100 = 20$$

Existing PFTF algorithm [6]: With '50' numbers of tasks considered for experimentation, '8' numbers of task entering the node and '5' numbers of task leaving the node, the overall task migration rate is given below.

$$T_{mig} = \frac{8+5}{50} * 100 = 26$$

Table 4.2 summarizes the comparison of task migration rate for 500 different tasks in CC environment. Task migration of QOHC-WO is compared with Crystal [1] and Energy-efficient fault-tolerant replica management policy [2] and PFTF algorithm [6] in table 4.2.

Fig. 4.2 given above illustrates the task migration for 500 different numbers of tasks. As the upload and download operations alternatively changes since the provider or the node only needed to handle per account at a time, the task migration is not proportionately same. Besides, from the figure it is also inferred, that with the increase in the number of tasks, also the migration rate increases.

Figure 4.2 given above illustrates the task migration for 500 different numbers of tasks. As the upload and download operations alternatively changes since the provider or the node only needed to handle per account at a time, the task migration is not proportionately same. Besides, from the figure it is also inferred, that with the increase in the number of tasks, also the migration rate increases. From the sample calculation, it is observed that with '50' numbers of task, the task migration rate using QOHC-WO method was observed to be '10', '18' using [1] and '20' using [2] respectively. From this analysis it is inferred that the task migration rate is lesser using QOHC-WO method when compared to [1], [2] and [6]. Followed by, various performance results are observed for each method. For each method, ten different results are observed. The performance of the proposed QOHC-WO method is compared to other existing methods. The task migrations are minimal in

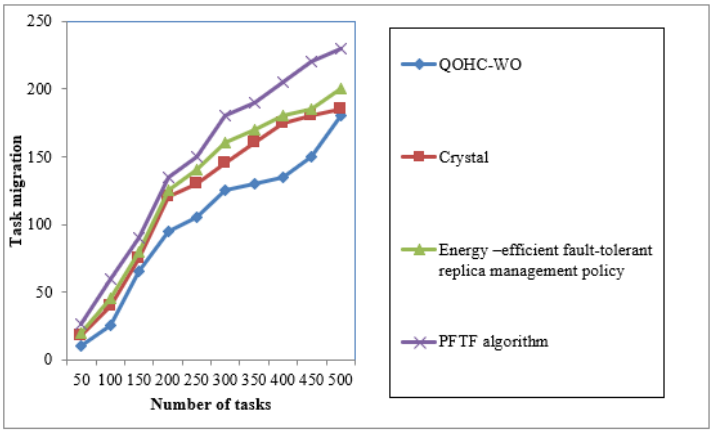


Fig. 4.2: Task migration

the QOHC-WO method due to extensive Osmotic Hybrid Vector Classifier algorithm in identifying the most appropriate VM to each of the tasks which is illustrated in Figure 5. So, the task request was said to be completed in the shortest time. But in the case of [1] and [2] algorithms with vertex configuration pattern [1] and considering the deadline and budget constraints [2] has not taken the task lengths. Instead, it considers only the resource capabilities of the node. Therefore, the task migration is said to be reduced using QOHC-WO method. Finally, the average of ten comparison results designates that the task migration of the proposed method is minimized by 21% when compared to [1], 26% when compared to [2], and 35% compared to [6].

4.3. Performance of fault occurrence rate. Finally, the fault occurrence rate is measured. The fault occurrence rate refers to the occurrences of fault during load balancing in CC environment. Lower the fault occurrence more efficient the method is said to be and vice versa. The fault occurrence rate is expressed as given below.

$$FOR = \frac{F_{occur}}{T} * 100 \tag{4.3}$$

From the above equation (4.3), the fault occurrence rate 'FOR' is measured based on the number of tasks provided to the node 'T' and the fault occurred 'F_{occur}' during that particular session. It is expressed in terms of percentage (%). The sample calculations for fault occurrence rate using the three methods, QOHC-WO, existing Crystal [1] and energy-efficient fault-tolerant replica management policy [2]and PFTF algorithm [6] is given below.

4.4. Sample calculation for fault occurrence rate.

Proposed QOHC-WO: With '50' numbers of tasks considered for experimentation and '4' number of faults occurred the overall fault occurrence rate is measured as given below.

$$FOR = \frac{4}{50} * 100 = 8\%$$

Existing Crystal [1]: With '50' numbers of tasks considered for experimentation and '6' number of faults occurred the overall fault occurrence rate is measured as given below.

$$FOR = \frac{6}{50} * 100 = 12\%$$

Existing Energy-efficient fault-tolerant replica management policy [2]: With '50' numbers of tasks considered for experimentation and '7' number of faults occurred the overall fault occurrence rate is

Table 4.3: Comparison of fault occurrence rate for different tasks

Number of tasks	Fault occurrence rate (%)			
	QOHC-WO	Crystal	Energy-efficient fault-tolerant replica management policy	PFTF algorithm
50	8	12	14	16
100	11	13	16	19
150	12	14	17	20
200	14	16	18	21
250	10	12	15	18
300	12	14	16	19
350	14	15	18	21
400	15	16	17	20
450	12	14	18	21
500	13	15	20	22

measured as given below.

$$FOR = \frac{7}{50} * 100 = 14\%$$

Existing PFTF algorithm [6]: With '50' numbers of tasks considered for experimentation and '8' number of faults occurred the overall fault occurrence rate is measured as given below.

$$FOR = \frac{8}{50} * 100 = 16\%$$

Table 4.3 summarizes the comparison of fault occurrence rate for 500 different tasks. Fault occurrence rate of QOHC-WO is compared with Crystal [1] and Energy-efficient fault-tolerant replica management policy [2] in table 4.3.

Figure 4.3 given above shows the fault occurrence rate. From the figure, it is illustrative that the fault occurrence rate is neither in the increasing trend nor in the decreasing trend. As shown in the graphical chart, there are four various colors of lines such as blue, brown, green and violet indicates the fault occurrence rate of four techniques namely QOHC-WO, Crystal [1] , energy-efficient fault-tolerant replica management policy [2] and PFTF algorithm [6] respectively. Among the four methods, the proposed QOHC-WO has the ability for reducing the fault occurrence rate. This is because of the reason that each task requested with the node differs in terms of resources (i.e. file size). Besides, for experimentation, the average of similar tasks is used for each simulation run and hence fair comparison is said to be ensured for all the three different methods. The fault occurrence rate is found to be reduced using the QOHC-WO method when compared to [1], [2] and [6]. This is because of the application of Improved Whale Optimization Algorithm. By applying this algorithm, first while choosing between options (i.e. between nodes for the corresponding task), an exploitation-exploration tradeoff is said to occur. To have a balanced between this tradeoff, with Bandit exploitation and Whale exploration, the VM migration is said to be minimized and hence the fault occurrence rate is also said to be reduced. By applying the QOHC-WO method, the fault occurrence rate was reduced by 15% when compared to [1], 29% when compared to [2] and 39% when compared to [6] respectively.

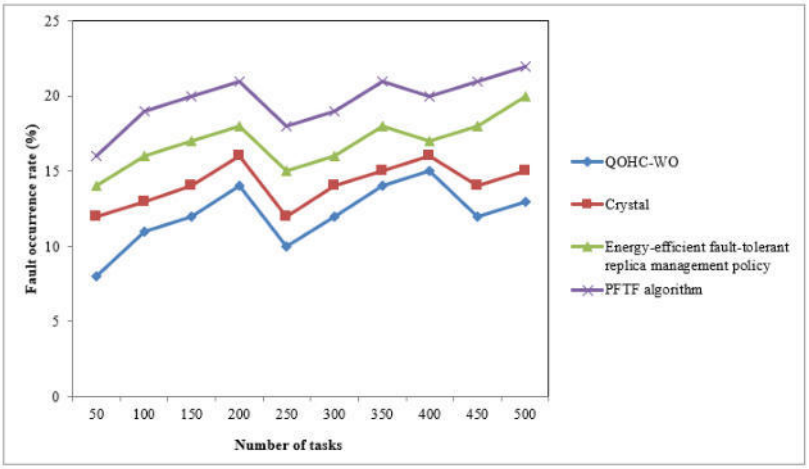


Fig. 4.3: Fault occurrence rate

The fault occurrence rate is found to be reduced using the QOHC-WO method when compared to [1] and [2]. This is because of the application of Improved Whale Optimization Algorithm. By applying this algorithm, first while choosing between options (i.e. between nodes for the corresponding task), an exploitation-exploration tradeoff is said to occur. To have a balanced between this tradeoff, with Bandit exploitation and Whale exploration, the VM migration is said to be minimized and hence the fault occurrence rate is also said to be reduced. By applying the QOHC-WO method, the fault occurrence rate was reduced by 15% when compared to [1] and 29% when compared to [2] respectively.

5. Conclusion. Several optimization methods have been available in the literature whose performances rely on various aspects. In this paper, the proposed Quadruple Osmotic Hybrid Classification and Whale Optimization (QOHC-WO) method is designed with classification and optimization to handle the three different scenarios in the CC environment. The osmotic hybrid vector classifier is used to consider the migration of VM in case of occurrences of faults with lesser task response time and the number of task migration. Depending on the classifier outcomes, the quadruple fault tolerance level, and osmotic hybrid vector classifier is employed that distributes the tasks with the nodes. Finally, Improved Whale Optimization is applied to handle the tradeoff between exploitation and exploration with aid of Bandit function and Whale optimizer. This helps to minimize the fault occurrence rate. The comprehensive experimental evaluation is conducted and the obtained result indicates that the proposed QOHC-WO method provides improved performance than the existing classification and optimization approaches, having a higher fault detection rate, and lesser job response time, as well as the number of task migration.

REFERENCES

[1] Nasirian S, Farhad F, *Crystal: A scalable and fault-tolerant Archimedean-based server-centric cloud data center network architecture*, Computer Communications, Elsevier, Sep 2019 [Crystal]

[2] Li L, Ping Wang Y, Chen Y, Luo Y, *Energy-efficient fault-tolerant replica management policy with deadline and budget constraints in edge-cloud environment*, Journal of Network and Computer Applications, Elsevier, Oct 2019 [energy-efficient fault-tolerant replica management policy]

[3] Chinnaiah M R, Nalini N,*Fault tolerant software systems using software configurations for cloud computing*, Journal of Cloud Computing: Advances, Systems and Applications, SpringerOpen, Jul 2018

[4] Sivagami V M, Easwarakumar K S, *An improved dynamic fault tolerant management algorithm during VM migration in Cloud Data Center*, Future Generation Computer Systems, Elsevier, Nov 2018

[5] Qin Y, Yang W, Xiao Ai, Chen L,*Fault tolerant storage and data access optimization in data center networks*, Journal of Network and Computer Applications, Elsevier, Apr 2018

[6] Tamilvizhi T, Parvathavarthini B, *A novel method for adaptive fault tolerance during load balancing in cloud computing*, Cluster Computing, Springer, Jul 2017
[7] Wu H, Jin Q, Zhang C, Guo H, *A Selective Mirrored Task Based Fault Tolerance Mechanism for Big Data Application Using Cloud*, Wireless Communications and Mobile Computing, Wiley, Feb 2019
[8] Liu S, Jia W, Pan W,*Fault-tolerant feedback virtual machine deployment based on user-personalized requirements*, Frontiers of Computer Science, Springer, Jul 2017
[9] Kumar S, Rana D S, Dimri S C,*Fault Tolerance and Load Balancing algorithm in Cloud Computing: A survey*, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 7, July 2015
[10] Milan S T, Rajabion L, Ranjbar H, Navimipour N J*, Nature inspired meta-heuristic algorithms for solving the load-balancing problem in cloud environments*, Computers and Operations Research, Elsevier, May 2019
[11] Kapil N. Vhatkar, Girish P. Bhole,*Optimal container resource allocation in cloud architecture: A new hybrid model*, Journal of King Saud University –Computer and Information Sciences, Elsevier, Oct 2019
[12] Tang F, Laurence T. Y, Tang C, Li J, Guo M,*A Dynamical and Load-Balanced FlowScheduling Approach for Big Data Centers inClouds*, IEEE Transaction on Cloud Computing, IEEE Transactions on Cloud Computing Volume: 6, Issue: 4, Oct.-Dec. 1 2018
[13] Cui J, Lu Q, Zhong H, Tian M, and Liu L, *A Load-balancing Mechanism for Distributed SDN Control Plane Using Response Time*, IEEE Transactions on Cloud Computing Volume: 6, Issue: 4, Oct.-Dec. 1 2018
[14] Sotiriadis S, Bessis N, Amza C, Buyya R,*Vertical and horizontal elasticity for dynamic virtual machine reconfiguration*,IEEE Transactions on Services Computing Volume: 12, Issue: 2, March-April 1 2019
[15] Guo L, Hu G, Dong Y, Luo Y, Zhu Y,*A Game Based Consolidation Method of Virtual Machines in Cloud Data Centers With Energy and Load Constraints*, IEEE Access, Dec 2017
[16] Liaqat M, Naveed A, Alim R L, Shuja J,Man Ko K,*Characterizing Dynamic Load Balancing in Cloud Environments Using Virtual Machine Deployment Models*, IEEE Access, Sep 2019
[17] Li C, Tang J, Ma T, Yang X, Luo Y,*Load balance based workflow job scheduling algorithm in distributed cloud*, Journal of Network and Computer Applications, Elsevier, Dec 2019
[18] Abdulhamid S M, Shafie M Latiff A,Abdul-Salaam G, Hussain Madni S M, *Secure Scientific Applications Scheduling Technique for Cloud Computing Environment Using Global League Championship Algorithm*, PLOS ONE |, 2016.
[19] Idris H, Ezugwu A E, Sahalu B , Junaidu, Aderemi O. A, *An improved ant colony optimization algorithm with fault tolerance for job scheduling in grid computing systems*, PLOS ONE, https://doi.org/10.1371/journal.pone.0177567 May 17, 2017.
[20] Hussain Madni S H, Abd Latiff M S, Abdullahi M,Abdulhamid S M, Usman M J,*Performance comparison of heuristic algorithms for task scheduling in IaaS cloud computing environment*, PLOS ONE | https://doi.org/10.1371/journal.pone.0176321 May 3, 2017.
[21] Nalepa J, Kawulok M, *Selecting training sets for support vector machines: a review*, Artificial Intelligence Review, Springer, Jan 2018.
[22] Net Migration Rate: Definition, Formula & Statistics - Video & Lesson Transcript |Study.com. *Study.com*. Retrieved 2017-03-29.
[23] Yuichi Nakatani, "Structured Allocation-Based Consistent Hashing With Improved Balancing for Cloud Infrastructure", IEEE Transactions on Parallel and Distributed Systems (Volume: 32, Issue: 9, Sept. 1 2021) DOI: 10.1109/TPDS.2021.3058963.
[24] PushpanjaliGupta, Prasan KumarSahoo, Bharadwaj Veeravalli, "Dynamic fault tolerant scheduling with response time minimization for multiple failures in cloud", Journal of Parallel and Distributed Computing, Elsevier, Volume 158, December 2021, Pages 80-93 https://doi.org/10.1016/j.jpdc.2021.07.019.
[25] Sudha Danthuluri , Sanjay Chitnis, "Energy and cost optimization mechanism for workflow scheduling in the cloud", Materials Today: Proceedings, Elsevier,2021. https://doi.org/10.1016/j.matpr.2021.07.168.

Edited by: Vinoth Kumar
Received: Jul 27, 2022
Accepted: Oct 26, 2022


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 339–349, DOI 10.12694/scpe.v23i4.2040

SMART HYBRIDIZED ROUTING PROTOCOL FOR ANIMAL MONITORING AND TRACKING APPLICATIONS

TANVEER BAIG Z *AND CHANDRASEKAR SHASTRY †

Abstract. Wireless sensor networks (WSN) have been exploited for countless application domains, most notably the surveillance of environments and habitats, which has already become a critical mission. As a result, WSNs have been implemented to monitor animal care and track their health status. However, excessive energy utilization and communication traffic on packet transmissions lead to system deterioration, especially whenever perceived information captured in the monitoring area is transferred to the access point over multiple dynamic sinks. Further to manage the energy and data transmission issue, the energy consumption and location aware routing protocol has been architected on the wireless Nano sensor nodes. In this article, a novel hybrid energy and location aware routing protocol to cloud enabled IoT based Wireless Sensor Network towards animal health monitoring and tracking has been proposed. However proposed data routing protocol incorporates the trace file for path selection for data transmission to base station using sink node. Trace file has been obtained on processing the cluster heads established in the network. Therefore, clustering of node in the network has to be achieved using LEACH protocol which enhances the network scalability and network lifetime by clustering the nodes with Metaheuristics constraints like location or node density comparability. The objective of the proposed model is to enhance the network scalability and energy consumption by establishing the multiple node clusters with high density cluster head through Metaheuristics Node Clustering optimization techniques. Metaheuristics based node clustering is been obtained using Improved Particle Swarm Optimization. Further it is employed to compute the optimal path for sensed data transmission to base station. Node clustering provides high energy consumption among the sensing nodes and to establish the high energy clusters towards sensed information dissemination to base station on dynamically reforming the nodes clusters with respect to Node density and node location. Simulation analysis of the proposed energy efficient routing protocol provides high performance in energy utilization, packet delivery ratio, packet loss and Average delay compared against the conventional protocols on propagation of the data through sink node to base station.

Key words: Wireless Nano Sensor Network, Cluster based data Routing, Animal Habitat monitoring, Particle Swarm optimization, Routing protocol.

AMS subject classifications. 68M14

1. Introduction.

Wireless Nano Sensor Network is employed as monitoring infrastructure with network topology on fixed deployment and capable of dynamically reconfiguring on basis of the data acquiring and transmission. WNSM can be projected to numerous applications for military purposes, natural disaster management, medical data acquiring and management etc. Implication of the WNSM is enhances exponentially towards supporting the animal habitat monitoring with Nano sensor to avoid multiple complication of the animals. Due to periodic transmission of sensed information utilizes the network resources such as node power, buffer space and network bandwidth [6].

The vital process for the WNSN is to Cluster nodes which collect the sensor information from the node and transmit the sensed information to base station with sink node on ensuring the efficient energy consumption and high transmission rate within defined energy [7]. Data Routing plays another crucial role for resolving the data transmission delay. Key requirement of the incorporating the Metaheuristics based clustering protocol is to increase the energy consumption on the heterogeneous sensor nodes towards data dissemination of sensed information of the base station. Especially sensed data routing technique will avoid the utilization of multiple mobile sink for data gathering as cluster head performs the specified operations for data transmission. Furthermore, path selection and coordination of the sensor nodes has been achieved on employing the clustering technique to cluster the nodes on specified strategies with respect to node location and node density [8].

*Dept Of Ece, Jain (Deemed to- be University) Bangalore, Karnataka India (tanveerbaigbe@gmail.com)
†2Dean PG Studies, Faculty of Engineering, Jain (Deemed to be University), Karnataka, India (bs.chandrasekar@jainuniversity.ac.in)

Node clustering can be carried out using particle swarm optimization on various fitness constraints on the nodes to establish the effective data transmission in the network to base station. Traditionally, path-based node clustering is employed to group the nodes on basis of the sink point location and sensor node location which establishes the data communication with reduced transmission delay. Henceforth, nodes will be dynamically updating the clusters and Cluster head according to the node energy and node location [9]. Correspondingly to produce the optimal performance against network degradation, sink trajectory has to be computed for sensed data routing of the wireless sensor network has been included in this article.

The proposed a new hybrid energy and location aware routing technique addresses the sensor network towards animal monitoring with on data collection by cluster head to disseminate the data to the base station through sink nodes. In addition, the node energy usage can be minimized on efficient sink node trajectory selection mechanism. The sink trajectory point in the network has been computed on the sink trajectory will minimize the transmission delay on the sensor information. Data communication order is computed by the inter-distance between the nodes and node radius and finally data transmission constraints have been employed for habitat monitoring.

The proposed a new hybrid energy and location aware routing technique addresses the sensor network towards animal monitoring with on data collection by cluster head to disseminate the data to the base station through sink nodes. In addition, the node energy usage can be minimized on efficient sink node trajectory selection mechanism. The sink trajectory point in the network has been computed on the sink trajectory will minimize the transmission delay on the sensor information. Data communication order is computed by the inter-distance between the nodes and node radius and finally data transmission constraints have been employed for habitat monitoring.

The remaining article is segmented as follows, detailed review of conventional approaches on energy enabled cluster-based routing protocols has been analysed in the section 2. In depth specification of proposed hybrid energy and location aware routing protocol has been designed using Metaheuristics improved particle swarm optimization technique has been represented in the Section 3. The simulation results of the approaches for data transmission to base station along its performance validation against various measures have been highlighted in Section 4. Finally, article has been concluded with final summary and future research directions in section 5.

2. Related Work. There are conventional approaches to the address the issues of the energy consumption and location aware routing Wireless Sensor Network with respect to cluster-based routing architectures. The architectures are as follows.

2.1. Low Energy Adaptive Clustering Hierarchy for Wireless Sensor Network. In this model, the number of low power and low cost sensor has been deployment as self-adaptive and traffic dependent network protocol on the traffic of the network. The node data transmissions of the sensed data are adaptively changes to the traffic pattern. Power changes occur based on traffic [9]. The node will be time- synchronized for path negotiation and data contention on basis of node density. Path Allocation model of the protocol enhances the transmission capabilities on the less utilized nodes to prevent network degradation. Further linear programming architecture has been employed to Dynamic node hopping sequence. The routing architecture provides optimal stability among the node transmission time with respect to node availability and energy consumed on the effective path within specified delay along throughput constraints to solve energy whole problem.

2.2. Metaheuristics model for Energy Optimized Clustering. In this architecture, the number of node is deployed in the environment to provide a scalable and adaptive data transmission path on specified time against sensed data of the network. It eliminates the energy whole problem in the network by eliminating the nodes with low node density for data transmission to base station. It is effectively used to maximize the lifetime of the network by employing the met heuristics principles on the nodes in the specified cluster. The idea of clustering is based on energy density among neighbouring nodes for data transmission. Optimal path allocation is subject to deliver the sensed data on available node density probability and throughput constraints of the transmitting nodes in the cluster. Further protocol manages the nods and node data transmission dynamically on its available energy density of the sensor nodes. The Balamurugan et al proposed the meta-heuristic methods for Cluster Head selection for improving data transmission using Improved Particle Swarm Optimization based Optimal Cluster Routing (MPO-IPSO-OCR) [1]. R.Zhang et al [2] designed model to

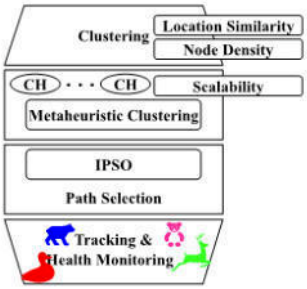


Fig. 3.1: Architecture of Proposed Model

improve intrusion detection system to identify accurate intrusion attack in a network Experimental results show that the proposed IPSO-SVM model can accurately identify the intrusion behaviour, and the detection accuracy is more than 82%. Y. Zaied et al [3] discussed an energy-efficient routing protocol for Cognitive Radio Wireless Sensor Network (CRWSN) this enhances the network energy consumption and network lifetime using MATLAB. Yanheng et al [4] brought a model to improve the efficiency of rechargeable Wireless rechargeable sensor networks (WRSNs) nodes using scheduling optimization. M.Khalid et al [5] paper focused on three Location based Routing protocols Such as Vector Based Forwarding (VBF), Hop by Hop-Vector Based Forwarding (HH-VBF) and Focused Beam Routing (FBR) for network life time and node alive consideration.

3. Proposed Model. In this section, a novel hybrid energy and location aware routing protocol to cloud enabled IOT based Wireless Sensor Network towards animal health monitoring and tracking has been designed on composition of optimal node clustering strategies, trajectory point of the sink node and energy defined path prediction. Sensor node deployed to acquire, transform and transmit the animal habitat information to the base station through sink nodes.

3.1. Trace File Generation using Channel Sensing.

3.2. Network Architecture –WSN. Wireless Sensor Network is established with provision to sense and transmitting the data sensed through sink node to base station. In these cases, sensing node is employed to sense the habitat of animals to monitor their behaviour from distinct location. Each sensor node in a network transmits the data among the nodes on the deployed network with node changing topology employing routing technique. Data transmission range of the node is similar among the network and its node link established in bi directional manner.

Architecture of WSN is depicted as undirected graph. The routing technique of WSN are modelled as Cluster based data routing to periodically control the node information which is essential to establish the path to route sensed data. Cluster head is established to eliminate frequent data loses and network failure. WSN is considered as equation (3.1) represents Network as Unidirected graph G.

$$UG = (P, I) \tag{3.1}$$

where P - collection of Sensor nodes & I - collection of the Sensor node edges.

3.3. Node Clustering. According the topology of WSN, clustering of the nodes has to be carry out in order to reduce the energy whole problem. Clustering of the node represents a cluster containing cluster head which acquires the continuous monitoring information of specific node in the node cluster along node energy and sensor information. Each node's energy and sensed information has been updated to its cluster head to alleviate the network overhead and energy degradation on path selection. On data clustering, the cluster topology changes inside a cluster on basis of the nodes density but it will not influence the network as effective queue management strategies have been modelled to cluster head.

The data clustering of nodes minimizes energy depletion of the sensor node on providing stable data transmission performance to base station with minimized hop and queue length. Furthermore, cluster head

Table 3.1: Clustering parameter

Clustering Parameter	**Description**
N	No. of sensor nodes on dynamic topology
R	Node Radius
N_d	Node Density
N_H	Node Hop distance

determination among the collection of sensor node is computed in the specified location of the network with particular network density and node distance. Clustering of node is established along the location similarity and node density similarity constraints towards habitat monitoring.

3.4. Clustering of Node on basis of Location Similarity. Collection of sensor node has been grouped as cluster on basis of its location uniqueness and its nearest location with least distance. The clustering establishing constraints of node forms the cluster which acts as quickest neighbour with optimal area radius to the selected node considered as cluster head. Table 3.1 provides the wireless sensor network parameter for cluster generation.

Clustering of the node is projected on basis of node scheduling and sensed data routing on basis of Wireless Sensor network parameters. The partitions of the sensor node are Cluster Head, transmitting node and intermediate node. It is established to increase the network transmission rate. In this node clustering on specified location, determination of the energy for data transmission is computed on basis of following equation (4.1).

$$\text{Clustering of Node C } = \lambda \int_{k=0}^{n} (\mathrm{r} * \mathrm{d})^2 L^R \tag{3.2}$$

where R – Network Area with radius between nodes, L – Node length between nodes, ND – Node density of each node

3.5. Clustering of Node Collection on basis of Node density. Collections of nodes have been clustered with respect to its node density. Clustering constraints on the nodes to determine the cluster head among the collection of nodes is based on node density and node radius with one hop neighbour. Node clustering regularly gathers the node energy as transmitting cost of the node respectively. Cluster head is computed using Algorithm 1.

Figure 3.2 depicts the cluster head formation architecture on the available sensor node for monitoring and transmission.

3.6. Trace File generation using Channel Sensing. The Node information is collected as Trace file to form path selection information using cluster head of the network. Further trace information is shared among the generated cluster heads. Trace files contains the node information gathered from the clustered nodes on parameter such as node energy density, node queue length, node bandwidth etc. Trace file considered as log file of each node and it is processed for log aggregator in the cluster heads. Trace File support for effective data transmission to base station.

Metaheuristics Clustering

Objective of the Metaheuristics clustering is to enhance the node scalability on establishing the multiple cluster head. Each Cluster heads has been considered as decision making nodes which transmit the sensed information to the specified sink trajectory point designating to the base station. Particular Network has segmented into various paths for effective data communication with respect to node queue length and node energy density. Clustering which facilitates the path selection towards data transmission on basis of node density and node location. Further sink trajectory for data collection has been computed on basis of the node queue length.

The Metaheuristics based node and data clustering is established using spatial information of nodes to enhance the optimal transmission performance. Effective path composing the nodes computed using clustering

Algorithm 1: Cluster and Cluster Head Formation

```
Ni Represents the collection of the sensor node
Calculate Node Density Nd
For (Node i=0 to N)
Fix Nd Maximum = Threshold
     If ( Nid  >= Nd Maximum  )
         Choose Nid as Cluster head among the nodes
     Else
        Identify the another node with high node density
Compute Node Degree Nr
For (Node i=0 to N)
Fix Nr Maximum = Threshold
      If ( Nir  >= Nr Maximum)
           Choose  Nir as Node degree of the nodes
      Else
          Compute other nodes as node degree
 Compute Cluster head
 If (Node has large energy density)
        Establish particular nodes as Cluster Head
 Else
   Compute the another nodes on basis of energy condition of the node
```

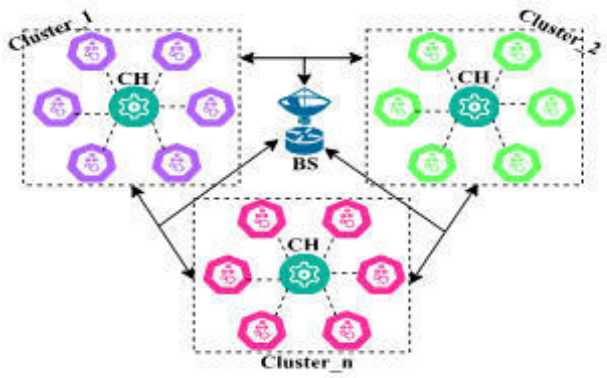


Fig. 3.2: Cluster head Formation

mechanism such as energy based node clustering and location based node clustering to reduce the usage of high energy on the nodes for data transmission.

Cluster head manages trace data of the nodes in the clusters for path allocation. For Path allocation for data transmission, Cluster head determine the node degree of the specified cluster node for data communication with other clusters node. Data communication between the two clusters is as represented as

Node degree changes between the clusters can be computed for intra cluster data communication as equation (3.3)

$$\Delta_v \;=\; |\, d_v \;-\; \delta \,| \tag{3.3}$$

Summation of node distance among the node neighbours is calculated as equation (3.4)

$$D_v = \sum_{v' \in N(v)} \left\{ dist \left(R, v' \right) \right\} \tag{3.4}$$

Sensor node's energy and location will act as primary constraint for Metaheuristics clustering. During large data transmission Traffic in the network, employment of Metaheuristics algorithms reduces the congestion and network degradation for data transmission to base station, it is given by equation (3.5)

$$PathPrediction = \frac{n_c}{\sum_i (x_i - \mu)^2} \tag{3.5}$$

where n_c - Total cluster heads in the WSN, x_i - Total Energy of the particular cluster in WSN, μ - Average number of neighbor's nodes in the cluster head.

The periodic node up gradation has been acquired by the cluster heads in order to minimize the data transmission time, cost and delay as possible.

Path Selection -Improved Particle Swarm Optimization

Transmission path selection for sensed information is to identify the nodes from different cluster head which capable for effective data dissemination to base station using Improved Particle Swarm Optimization (IPSO). The fitness criteria for selecting the transmitting node among different cluster are computed using the total available energy of the nodes in cluster head. The sensed information incorporates IPSO approach to generate dynamic strategies for effective path for data dissemination to base station through sink nodes.

Due to the variations in received signal strengths in various locations, effective path radius for data dissemination can be represented as equation (3.6)

$$PR_{i,j} = \Gamma_{i,j} N_j \tag{3.6}$$

where N_j depicts the radius between the transmitting nodes, $\Gamma_{i,j}$ - Node energy efficiency factor.

During multiple data communication to base station, the total energy consumed for day a communication on particular path can be computed as equation (3.7)

$$MR_{i,j} = NE_{i,j} \Gamma_{i,j} N_j / (f_j + 1) \tag{3.7}$$

where $NE_{i,j}$ denotes node energy loss Factor, f_j denotes fitness function.

Path selection strategy for data dissemination computed using IPSO for node indexing in the cluster and among the cluster for data transmission. The multiple strategy of the node in the cluster provides the route outcome to transmit the sensor information acquired in specified location or specified time period. Strategies can be energy efficiency factors, energy utilization constraints and location specification with respect to fitness function. The improved PSO optimization computes better path for effective data dissemination of the base station through sink on the particular location. In this sink trajectory computed using fitness function.

Path selection is given by velocity factors equation (3.8)

$$P = Nd = P + V1 * random * (pBest - Nd) + V2 * random * (gBest - Nd) \tag{3.8}$$

where P is the data acquired on the Path for data dissemination, P is the particle represents the acquired node information on the respective cluster, $V1$ is the total node information in the particular cluster, $V2$ is the total node information in the entire network, $pBest$ is the appropriate Node location in particular cluster, $gBest$ is the ppropriate Node position in entire cluster head, $random$ is a random variable.

Optimal data communicating path O(p)

As an outcome, optimal path for data transmission has been selected efficiently. It is represented to be efficient for data transmission to base station.

Table 4.1: WSN Simulation Parameters

Simulation Parameter	**Value**
Network Simulator	MATLAB 2018b
Network Size	500m *500m
Node deployed	50
Node Power	0.5mW/Hz
Data Traffic	CBR
Data Packet size	1028 bytes
Buffer size	150 packets
Simulation period	45 minutes

Energy Consumption = Energy consumed during designing- Energy consumed during data transmission

Energy consumed during data transmission= Initial energy of the node before transmission - energy of the node after

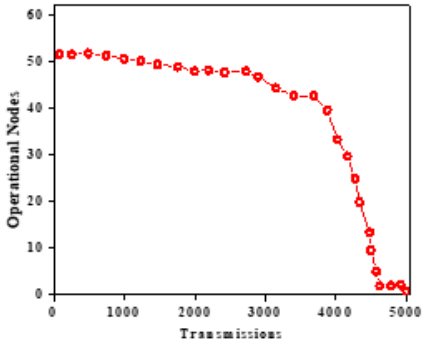


Fig. 4.1: Transmission capability of operation nodes

4. Simulation Analysis. In this part, hybrid energy and location aware routing protocol to cloud enabled IoT based Wireless Sensor Network towards animal health monitoring and tracking model has stimulated by employing MATLAB Simulator. On experimental validation of the approach, the WSN properties and its network performance measures has been validated with respect to node density, energy utilization, Packet delivery ratio, Average delay and packet loss. Proposed Metaheuristics defined clustering based routing protocol performs better compared to the conventional cluster-based routing protocols such as LEACH protocol and PSO based node clustering techniques using fitness constraints. The proposed clustering protocol employed to enhance the performance of the network through node density coordination among various clusters. Simulation set up of the proposed network architecture has been described in the table 4.1.

In detailed simulation of the node containing sensed information has been transmitted has using network properties represented in the table 4.1. Further it is node transmits the sensed information cluster head. Cluster heads determines the appropriate path for data dissemination to base station through sink. The desired path is computed by cluster head on optimal selection of energy density and location of the nodes with reference to intermediate node on the cluster head. Figure 4.1 represents the no of transmission capabilities of the operational nodes.

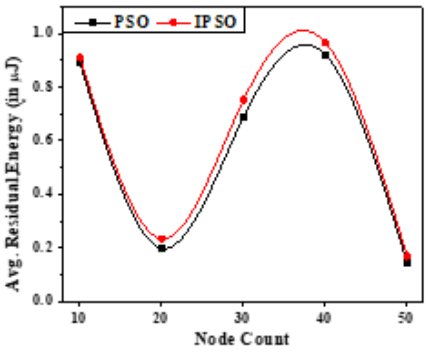


Fig. 4.2: Performance Analysis of Energy Consumption

Energy computation is computed using following representation.

1. Energy Consumption = Energy consumed during designing- Energy consumed during data transmission.
2. Energy consumed during data transmission= Initial energy of the node before transmission - energy of the node after the evaluation of the model on the node energy consumption is depicted in the figure 4.1.

Throughput of the network considered as total energy density of the clustered nodes in the network. Network energy depletion can be eliminated on employing the dynamic strategies for path selection.

Cluster based data routing protocol is been synchronized in term of time and location. The proposed IPSO algorithm determines the optimal data transfer path for data dissemination to base station on communicating with multiple cluster head in the total network. IPSO algorithm calculates the shortest data transmission path for sensed data routing to base station through sink node [20]. Total node energy consumed at several intervals has been evaluated and represented in the figure 4.2.

The optimal path for sensed information dissemination to base station against node failure is computed and its energy consumption has been calculated. Those computations prove that proposed model operates with less energy consumption and less data losses. Data transmission on the data traffic is been controlled using Metaheuristics based node clustering mechanism. The Metaheuristics clustering technique is capable of establishing the cluster head with effective data transmission rate and reduced energy utilization on the computed paths.

Energy computation is computed using following representation.

The evaluation of the model on the node energy consumption is depicted in the figure 4.2. Throughput of the network considered as total energy density of the clustered nodes in the network. Network energy depletion can be eliminated on employing the dynamic strategies for path selection.

Average delay of the network is computed as ration of number of the sensor information to the total number of sensor information transmitted to the base station in specified time interval. In this IPSO strategies has been employed to identify the optimal path for data dissemination on basis of location and energy based fitness constraints, $f()$ as represented in equation (4.1).

$$f() = \sum_{m-k}^{n-1} Sim\left(\delta_i^j(x_k) - \delta_i^j(y_{m+1})\right) + F(y_n) \tag{4.1}$$

The average delay evaluation of the cluster based routing architectures is represented in the figure 4.3.

$$Averagedelay = 100\% * (nd - ne)/nd \tag{4.2}$$

The average delay is computed as nodes in the network against the sensed data transmission to base station on the specified time and specified location to fixed data sizes. Table 4.2 provides the performance comparison of the different cluster-based routing technique for WSN for data transmission and energy consumption.

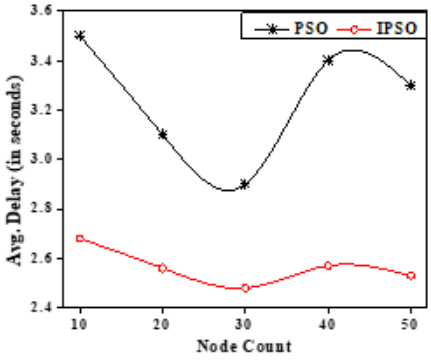


Fig. 4.3: Performance Analysis of average delay

Table 4.2: Performance Comparison of Cluster based data Routing mechanism for WSN

Technique	Average delay in mbps	Energy Consumption in joules	Packet Delivery Ratio	Network life time in ms
Data Routing and path selection using PSO- Existing	75.58	10.23	95.78	**20**
Data Routing and path selection using Improved PSO Proposed	79.26	12.59	99.85	**48**

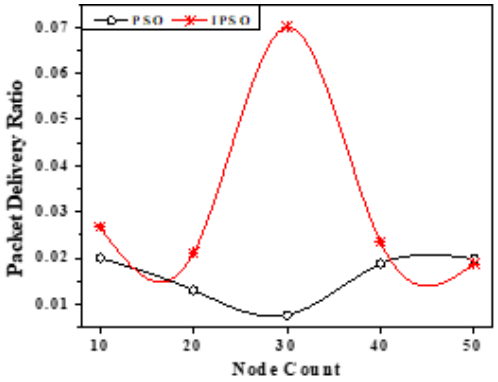


Fig. 4.4: Performance Analysis of Packet Delivery Ratio

Packet delivery ratio is determined as ratio containing the number of sensed data information in the particular time to the particular node's transmitting data to base station in the particular time in specified path and energy conditions. Figure 4.4 describes the performance evaluation of the packer deliver ratio.

The packet loss is computed on basis of energy loss of the network in parallel lead to data transmission loss. In this work, an energy management strategy of the node has been controlled using IPSO. The energy controlling reduces the node failure while transmitting data to base station via sink node.

In addition, cluster-based data routing mechanism is capable of handling the network complication effectively. The figure 4.5 depicts the performance evaluation of the packet loss against energy and location of the nodes. Network lifetime of the nodes depends on the scheduling of the nodes in the cluster for data transmission

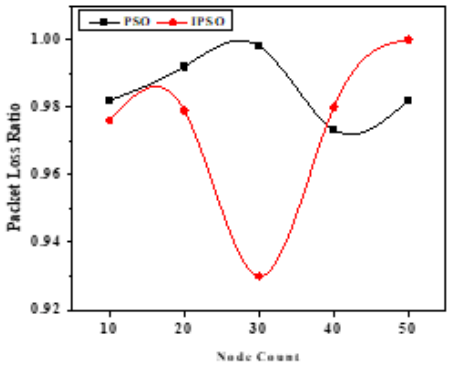


Fig. 4.5: Performance Analysis of Packet loss

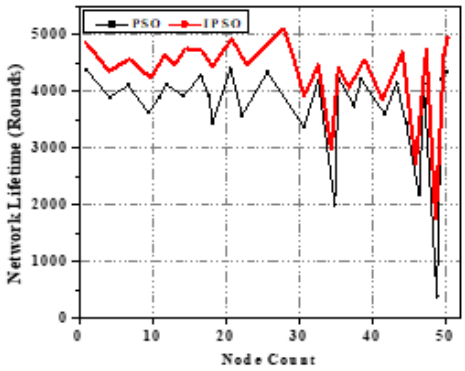


Fig. 4.6: Performance Analysis of Proposed Framework against Existing Technique through Network lifetime

to base station with assistance of the sink nodes.

In this work, network life time will be increased on basis of operating the nodes on the fixed threshold to avoid the node failure and degradation of the network performance. Figure 4.6 represents the performance analysis of the network lifetime against the node deployed.

In this analysis it has been proven that energy conservation of the node on the specified location maximizes the number of data transmission optimal path selection for animal habitat monitoring and tracking using improved PSO technique. It has been evaluated against conventional cluster based data routing protocol evaluate the amount on the energy utilized.

5. Conclusion. We design and simulated a novel hybrid energy and location aware routing protocol to cloud enabled IOT based Wireless Sensor Network towards animal health monitoring and tracking has considered as effective path selection solution towards information dissemination to base station by sensor node. Further it reduces the data losses, average delay and energy consumption to more extent. Metaheuristics based clustering approach provides optimal condition during node failure with reduced energy consumption. The Proposed data routing model increases the energy consumption operational nodes by imposing effective strategies of the improved PSO.

REFERENCES

[1] Balamurugan A. Balamurugan, S. Janakiraman, M. D. Priya and A. C. J. Malar, "Hybrid Marine predators optimization

and improved particle swarm optimization-based optimal cluster routing in wireless sensor networks (WSNs)," in China Communications, vol. 19, no. 6, pp. 219-247, June 2022.
[2] R. Zhang, Y. Song and X. Wang, "Network Intrusion Detection Scheme Based on IPSO-SVM Algorithm," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2022, pp. 1011-1014.
[3] Y. Zaied, W. Saad and M. Shokair, "Energy Efficient Of Grid Clustering Based TEEN Protocol for Cognitive Radio Wireless Sensor Networks," 2021 International Conference on ElectronicEngineering (ICEEM), 2021, pp. 1-6.
[4] Yanheng Liu et al "Scheduling Optimization of Charging UAV in Wireless Rechargeable Sensor Networks"2021 IEEE Symposium on Computers and Communications (ISCC) | 978-1-6654-2744-9/21/$31.00 ©2021 IEEE |.
[5] M. Khalid, Z. Ullah, N. Ahmad, H. Khan, H. S. Cruickshank and O. U. Khan, "A comparative simulation-based analysis of location based routing protocols in underwater wireless sensor networks," 2017 2nd Workshop on Recent Trends in Telecommunications Research (RTTR), 2017, pp. 1-5.
[6] L. Gao, and X. Wang, "A game approach for multi-channel allocation in multi-hop wireless networks," in Proc. 9th ACM Int. Symp. Mobile Ad Hoc Netw. Comput., 2008, pp. 303–312.
[7] Yin S. and Lin, X.: 'MALB: MANET Adaptive Load Balancing', IEEE 60th Vehicular Technology Conference VTC 2004-fall, 2004, 4, pp. 2843-2847.
[8] M. Felegyhazi, M. Cagalj, S. Bidokhti, and J.-P. Hubaux, "Noncooperative multi-radio channel allocation in wireless networks," in Proc. IEEE 26th Conf. Comput. Commun. May. 2007, pp. 1442–1450.
[9] I.Katzela, and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," IEEE Pers. Commun., vol. 3, no. 3, pp. 10–31, Jun. 1996.
[10] Yen, J. Li, and C. Lin, "Stability and fairness of AP selection games in IEEE 802.11 access networks," IEEE Trans. Veh. Technol., vol. 60, no. 4, pp. 1150–1160, Mar. 2011.
[11] Zhu, D. Niyato, and P. Wang, "Network selection in heterogeneous wireless networks: evolution with incomplete information," in Proc. 2010 IEEE WCNC.
[12] D. Niyato and E. Hossain, "Dynamics of network selection in heterogeneous wireless networks: an evolutionary game approach," IEEE Trans. Veh. Technol., vol. 58, no. 4, pp. 2008–2017, May 2009.
[13] . M. Transier, H. Fubler, J. Widmer, M. Mauve, and W. Effelsberg, "A Hierarchical Approach to Position-Based Multicast for Mobile Ad-Hoc Networks," Wireless Networks, vol. 13, no. 4, pp. 447-460, Aug. 2007.
[14] P. Namboothiri and K. Sivalingam, "Capacity analysis of multihop wireless sensor networks using multiple transmission channels: A case study using ieee 802.15.4 based networks," in Proc.
[15] Michael J. Neely, "Dynamic Data Compression for Wireless Transmission over a Fading channel", University of Southern California,http://www-rcf.usc.edu/_mjneely.
[16] P. Kyasanur and N. Vaidya, "Routing and Link-Layer Protocols for Multichannel Multi-Interface Ad Hoc Wireless Networks," ACM SIGMOBILE Mobile Computing and Comm. Rev., vol. 10, pp. 31-43, 2006.
[17] X. Fang, C. Zhu, and P. Fan, "Greedy-based dynamic channel assignment strategy for cellular mobile networks," IEEE Commun. Lett., vol. 4, no. 7, pp. 215–217, July 2000.
[18] S. Sen, R. R. Choudhury, and S. Nelakuditi. CSMA/CN: carrier sense multiple access with collision notification. IEEE/ACM Transactions on Networking (ToN), 20 (2):544–556, 2012.
[19] W. Zhou, K. Srinivasan, and P. Sinha. RCTC: Rapid concurrent transmission coordination in full duplex wireless networks. In Proc. of IEEE ICNP, pages 1–10, 2013.
[20] S. Liu, L. Lazos, and M. Krunz. Thwarting control-channel jamming attacks from inside jammers. IEEE Transactions on Mobile Computing, 11 (9):1545–1558, 2012.

Edited by: Vinoth Kumar
Received: Aug 3, 2022
Accepted: Nov 22, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 351–361, DOI 10.12694/scpe.v23i4.2046

DESIGN AN UNCERTAIN MODEL FOR THE STUBBED GROUND PLANE BY INCREASING THE BANDWIDTH OF THE MONOPOLE ANTENNA

D. JASMINE DAVID*, K. RAMALAKSHMI†, M. SELVARATHI‡, T.JEMIMA JEBASEELI§, P. SURESH¶, AND P.DHIVYA ‖

Abstract. Uncertainty is stated as the indication of the quality of the calibration certificate. When uncertainties occurred in a stubbed ground plane, it affects the performance of every source. Hence, a thorough analysis of uncertainties on a monopole antenna is required. The proposed research work is to focus on designing a planar monopole antenna to improve the bandwidth with minimal changes on the ground plane mainly for medical applications and to reduce the uncertainty. The narrowband antenna on the ground plane is redesigned to boost up the gain and broadens the monopole antenna's bandwidth impedance. Then the ground plane is integrated with the rectangular plate. As a result the bandwidth is 42.5 GHz ahead.

Key words: Bandwidth; monopole antenna; stubbed ground plane; uncertainty, wireless communication.

AMS subject classifications. 68M14, 78A50

1. Introduction. There is a lot of interest in developing on-body monitoring devices. Now, people are more aware of and accepting of healthcare technology. A Body Area Network (BAN) is the foundational mechanism for tracking human health concerns. [1]. BANs apply to both the commercial and healthcare sectors. Despite their different applications, the criteria of devices grouped around the BAN are demanding, including miniaturization, ultra-light weightless, discreetness, cost, self-sustainability, and safety. While retaining the intended functioning, the gadgets must interfere with human mobility as little as possible [2]. These on-body devices are intended to monitor human activity and collect crucial physiological and clinical signals to monitor the health of the aged, disabled, and any other sick patients, as well as a tracking system.

2. Literature Review. Wideband systems have non-ionizing radiation qualities; higher resolution, and great penetration capacity, wideband systems have been effectively adopted into healthcare [3]. It also appears to be tolerant of modulation devices due to its very low transmit power to high-frequency range [4]. A wideband antenna is a pass filter and sends small pulses rather than a continuous carrier wave; distortion and wave spreading are also reduced, resulting in more efficient transmission. A variety of antennas for uncertain approaches have been proposed. Some of them are simple modifications to ordinary antennas that have been used for decades [5]. The modified planar monopole antenna is one such uncertain antenna that has demonstrated potential wideband performance. Planar monopole antennas have several benefits due to their simple design and omnidirectional emission pattern [6-7]. The ground plane is embedded with microstrip lines, and readily integrated with microwave circuitry. Boosting the bandwidth, many ways to alter monopole antennas have been tried [8-12]. Some of these include dielectric resonators, annular slots, inverse T of ground planes, simple bevel design, feature smoothing, shorting pins, fractal feature creation, open slot insertion, spurious feature integration, use of substrates of varying thicknesses, and truncated ground planes [13-15].

Impedance matching is performed on a small monopole antenna by inserting a stub on the radiating section of the antenna or on its ground plane. It is less difficult to implement. Different sorts of uncertain

*This work was supported by Karunya Institute of Technology and Sciences.
Assistant Professor, ECE, Karunya Institute of Technology and Sciences, Coimbatore, India (jasmine@karunya.edu)
†Associate Professor,CSE, Alliance University, Bangalore, India (ramalakshmivenkatesan@gmail.com)
‡Assistant Professor,Mathematics, Karunya Institute of Technology and Sciences, Coimbatore, India (selvarathi@karunya.edu)
§Associate Professor, CSE, Karunya Institute of Technology and Sciences, Coimbatore, India (jemima_jeba@karunya.edu)
¶Assistant Professor, AI&DS, M. Kumarasamy College of Engineering, Karur, India (psuresh82@gmail.com)
‖Assistant Professor, CSBS, M. Kumarasamy College of Engineering, Karur, India (dhivyainfo@gmail.com)

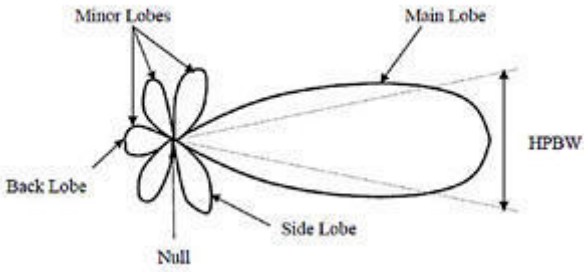


Fig. 3.1: Radiation pattern of general directional antenna.

stub structures have different outcomes [16-18]. Its primary application is to improve bandwidth and radiation characteristics.In this article, we have modified a small circular unbalanced antenna to provide an extremely broadband body health monitoring system. The radiating component has spherical and circular slots and 50 microstrip lines. The modified partial ground plane includes a smaller rectangular area that acts as an efficient impedance matching device to increase bandwidth and reduce uncertainty. It is made up of Kepton material.The stray or edge capacitance will be impacted by the microstrip thickness. Specifically, the stay capacitance of the conductor increases as conductor thickness increases.

3. Methodology. In a wireless communication system, the antenna is a vital component. A well-designed antenna can minimize system requirements and uncertainty while boosting overall performance.

3.1. Radiation Pattern. The radiation of the antenna depicts the radiation qualities in the far-field as a function of spatial coordinates defined by the elevation angle θ and the azimuth angle ϕ. This is a graph of the radiated antenna's power.

Take the anisotropic antenna as an example. An isotropic antenna radiates in the same direction in all directions. If an isotropic antenna radiates total power P, the power is spread with radius r, and density S is given in the following equation.

$$s = \frac{p}{area} = \frac{p}{4\pi r^2} \tag{3.1}$$

$$U_i = sr^2 = \frac{p}{4\pi r^2} \tag{3.2}$$

U_i denotes the isotropic antenna's radiation intensity.

The directional antenna is a more practical form of dissipating more power in one direction and less power in another. An omnidirectional antenna has a consistent radiation in E-plane but changes in H-plane. The description of the different phases of Figure 3.1 is as follows.

- PBW: The angle is determined not entirely by the half-power.
- Main Lobe: It's the largest radial direction of the lobe.
- Minor Lobe: All minor robes except major robes are classified as minor robes. The minor lobes emit radiation in an unfavorable direction.
- Back Lobe: It is the exact opposite of the minor lobe.
- Side Lobes: These are sidelobes separated by zero from the main robe. Minor robes are the most prominent of the minor robes.
- Sidelobes are not desirable in most wireless systems. Sidelobes are the largest of the sidelobes. Sidelobes are not desirable in most wireless systems. Therefore, a good antenna design should prevent small robes.

3.2. Directivity. The proportion of the force of radiation transmitted by the uncertain antenna in a ground plane to the normal power of radiation discharged by the receiving wire every which way is characterized

as the receiving wire's directivity.

$$s = \frac{U}{U_i} = \frac{4\pi U}{P} \tag{3.3}$$

where D denotes antenna's directivity and U denotes antenna's radiation intensity. The radiation intensity of an isotropic source is denoted by U_i.

The complete power transmitted is indicated by P. The bearing of the directivity isn't given all of the time. In this uncertain situation, the highest directivity is implied by the direction of maximum radiation intensity:

$$D_{max} = \frac{U_{max}}{U_i} = \frac{4\pi U_{max}}{P} \tag{3.4}$$

The greatest radiation force is meant by U_{max}. Since it is the proportion of two radiation forces, directivity is a dimensionless amount. As a result, it is often expressed in decibels (dBi). An antenna's radiation pattern may be easily utilized to calculate its directivity.

3.3. Input Impedance. The impedance presented by a receiving wire at its terminals is characterized as the voltage to flow proportion at the sets of terminals or the proportion of the pertinent parts at a given position. As a result, the impedance of the radio wire might be composed as:

$$Z_{in} = R_{in} + jX_{in}Z_s = R_s + jX_s \tag{3.5}$$

The imaginary part X is the information impedance addresses where the power is put away near the radiation field. The resistive impedance at input R is involved two segments sections: R_i is the radiation obstruction and R_L is the resistance loss. The power-related to the radiation opposition addresses the genuine power transmitted by the radio wire, while the uncertain power squandered in the misfortune obstruction is lost as hotness in the receiving wire because of dielectric or conductive misfortunes.

3.4. Voltage Standing Wave Ratio (VSWR). The impedance mismatch between the transmitter and the antenna is assessed by VSWR. Higher the VSWR, then there will be a greater disparity. VSWR compares to an ideal pair. Because of the way that most antenna design is intended for this uncertain impedance, a satisfactory receiving antenna configuration ought to have an information impedance of 50 Ω or 75Ω. A VSWR of 1:1 denotes an ideal match.

3.5. Return Loss. Return Loss concludes how much power is lost to the load. At the point when the transmitter and antenna impedance don't coordinate, hence it reflects the waves. This result leads to the arrangement of standing waves, as portrayed in the previous section.

R_L metric reflects how well the transmitter and antenna have been matched.

The expression for R_L is given in equation (3.6) to ensure that the transmitter and antenna are perfectly matched.

$$R_L = -20\log_{10}\Gamma \tag{3.6}$$

$\Gamma = 0$ and $R_L = \infty$ implying that no power is reflected.

3.6. Antenna Efficiency. The antenna efficiency is an action that thinks about how much loss is at the receiving wire's terminals as well as inside the antenna's design. These losses are caused by,

- Reflections are induced by an uncertain antenna-to-transmitter mismatch.
- I2R is the loss due to conduction.

As a consequence, the efficiency of the antenna's function is expressed as follows:

$$e_t = e_r e_c e_d \tag{3.7}$$

where e_t stands for total antenna efficiency. $e_r = (1 - \Gamma^2|)$ efficiency of reflection.

The conduction efficiency is denoted by e_c. e_c and e_d are combined to generate efficiency e_{cd}. This is expressed as follows:

$$e_{cd} = e_c e_d = \frac{R_r}{R_r + R_L} \tag{3.8}$$

where e_{cd} is the ratio of power sent to R_r to R_r and R_L is known as antenna radiation efficiency.

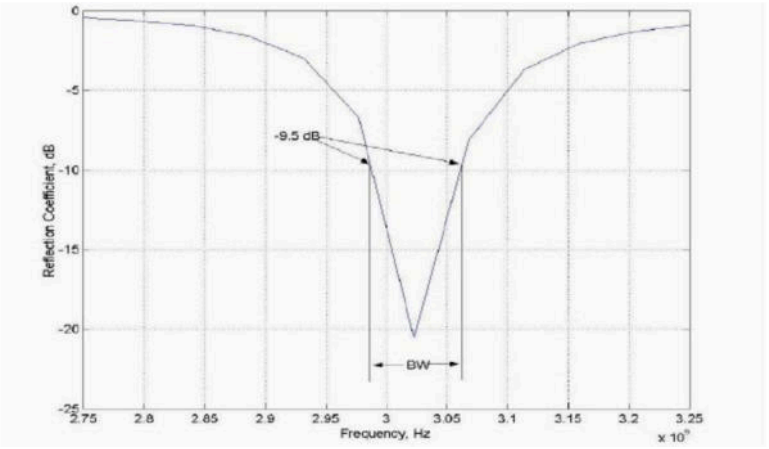


Fig. 3.2: Bandwidth measurement

3.7. Antenna Gain. Antenna gain is a quantity that strongly connected the directivity of the antenna. The amount of energy that an antenna concentrates in one direction over other emission directions is referred to as directivity. As a result, if the antenna is efficient, the directivity equals the antenna gain. The gain is generally connected to the primary flap and provided toward most elevated radiation is portrayed as follows:

$$G(\theta, \phi) = e_{cd}(\theta, \phi) \tag{3.9}$$

3.8. Bandwidth. The data transmission and the bandwidth are characterized as the usable frequencies comparable to some predetermined standard. The transfer speed of a broadband antenna is characterized as the proportion of satisfactory higher to bring down frequencies. The transmission capacity is characterized as the proportion of fluctuation across center frequency. It is defined as follows:

$$BW_broadband = \frac{f_H}{f_L} \tag{3.10}$$

$$BW_narrowband(\%) = \frac{(f_H - f_L)}{f_c} \tag{3.11}$$

where f_H denotes the maximum cut-off frequency. f_L is the lower cut-off frequency, and f_c is the center frequency. If f_c is the center frequency. If $\frac{f_H}{f_L} = 2$, an antenna is considered to be broadband.

Measuring an antenna's VSWR is one approach to determine how effectively it operates over the desired frequency range. Figure 3.2 depicts bandwidth measurement.

3.9. Beam Width. The bandwidth of the antenna is the shaft width. After determining the radiation intensity, the spots on each side of the peak intensity are identified. The half-power beam width is 3dB. Assume that the bulk of the radiated power is not divided into side lobes. Thus the directive gain is inversely proportional to the beam width. As the beam width decreases, the directive gain decrease as well.

The matching slots might still be employed on here just to alter the compatibility even if matching slots in microstrip patch are half-wave resonance. Similar to a wire mesh monopole, the monopole's bandwidth is enhanced because it is circular. The bandwidth of thick monopoles is greater than that of thin monopoles. The matching and the bandwidth are both affected by the curvature of the ground; typically, the match defines the bandwidth, thus modifying one affects another. This is a microstrip-fed monopole with an extended ground instead of a flat ground that runs perpendicular to the monopole; it is not a microstrip patches with a partial ground. The stray or edge capacitance will be impacted by the microstrip thickness. Specifically, the stay capacitance of the conductor increases as conductor thickness increases.

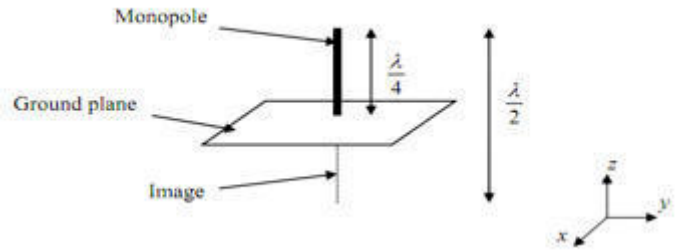


Fig. 4.1: Monopole antenna structure.

3.10. Side Lobes. There is none of the antenna emit all of their energy in a single chosen direction. Some of it will certainly be radiated in other ways. The peaks are known as side lobes, and they are generally described in decibels down from the main lobe.

3.11. Nulls. An invalid zone in a receiving antenna radiation design is viable with t minimal transmitted power. An invalid as often as possible has a smaller directivity point than the essential beam. Subsequently, the null is useful for an assortment of reasons including remembering the concealment of interfering signs for a specific direction.

3.12. Radiation Resistance. Resistance to Radiation is described as a fake or hypothetical resistance that dissipates the same amount of power as the radiated power.

$$R_r = \frac{PowerRadiated}{{I_r}^2 ms} \tag{3.12}$$

3.13. Radiation Intensity. The power emitted in a specific direction per unit solid angle is known as radiation intensity.

$$RI = \frac{r^2 E^2}{\eta_0} = r^2 P \frac{Watts}{unit} solidangle \tag{3.13}$$

Here η_0 = intrinsic impedance, r = radius of the sphere (meter) P = power radiated instantaneously, E = Electric Field Strength (V/m).

4. Monopole Antenna Design. The monopole antenna is the consequence of utilizing the structure hypothesis on a dipole [19-20]. Assuming a directing plane is set under a solitary component of length L/2 conveying a current, the component and its picture act indistinguishably from a dipole of length L, with the provision that radiation happens just over the uncertain ground plane, as portrayed by Saunders. Figure 4.1 tells the best way to construct a monopole antenna.

The top ground raises the line's capacitance, decreases its impedance, and perhaps most likely modifies the dielectric characteristic constant and the value of velocity. If the vias were too near to the edge, it could increase capacitance and interfere with other processes. They would cause a cyclic disturbance in the capacitance, perhaps lowering the maximum working frequencies. This will not occur if the vias are sufficiently distant from the edge. To add frequency sweep, the following procedures are required. First select the menu item HFSS, Analysis Setup, Add Sweep. Then, select Solution Setup: Setup1. Then, Edit Sweep Window with Sweep Type as Fast and Frequency Setup Type as Linear Count. Finally, Give Start and stop frequency as 5 GHz and 10 GHz in terms of Count as 500.

The antenna's directivity is doubled and its radiation resistance is half that of a dipole. A quarter-wave monopole can have the same properties as half-wave dipoles. Monopoles are astoundingly useful in convenient antennas, where the uncertain coordinating plane might be the body of the vehicle or the phone shell. The quarter frequency monopole regularly has an increase of 2-6dB and a data transmission of around 10%. It has

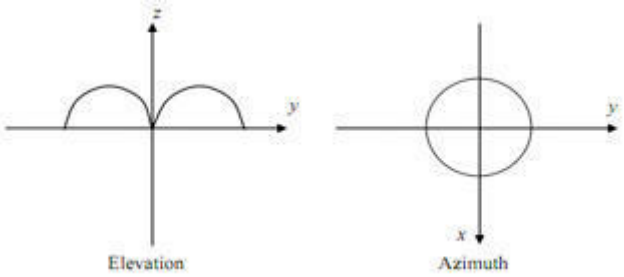


Fig. 4.2: The monopole antenna's radiation pattern.

Table 4.1: Particulars of design

PARAMETER	DIMENSIONS (in mm)	
	LENGTH	WIDTH
SUBSTRATE (FR4 epoxy)	40	60
GROUND	40	22
STRIP LINE	20	2
GAP	2	22

a 36.5 radiation resistance and directivity of 3.28. (5.16dB). Figure 4.2 portrays the radiation example of a monopole. Table 4.1 portrays the details of the design.

5. Compact Antennas Feed Analysis. The simplest extraction approach is a microstrip line feed, which has the benefit of feeding in the same plane as the radiating monopole. When integrating the big arrays of feeding network f, this way of simply attaching a 'U' patch at the edge is quite convenient. However, the patch's erroneous radiation causes issues. Prior knowledge of the feed point position is essential for impedance matching in this form of stimulation. Figure 5.1 depicts a microstrip line fed rectangular patch.

6. Experimental Results and Discussions. The High-performance Full-wave Electromagnetic System (HFES) is an inconsistent 3D volumetric model that uninvolved gadgets with the Microsoft Windows graphical interface. It incorporates representation, strong demonstration, and mechanization in a simple to learn environment to generate results for 3D-EM issues.

Ansoft spearheaded the utilization of the Finite Element Method (FEM) to electromagnetic demonstrating by including innovations like unrelated vector limited components, versatile lattice, and Adaptive Lanczos-Pade Sweep (ALPS). HFSS keeps on driving the market today with the developments like Modes-to Nodes and Full Wave Spice.

Ansoft HFSS has advanced over time because of client and industry input. It is the business standard for high-efficiency and virtual prototyping. As a result, the design is adjusted and the stub is integrated as shown in Figure 6.1 which results in the same reflection coefficient. The VSWR plot in Figure 6.1 shows the obtained UWB. The recurrence goes from 3.29 to 9.33 GHz. The recurrence band for UWB is 3.1 to 10.6 GHz.

The connectors used here are Coax, SFP/XFP, Backplane Transitions.Dielectric constant, bulk conductance, and loss tangent are employed in the battery's HFSS simulation.A reflector must be fixed quarter wavelengths distant from the origin for it to operate. As a result, waves reflected from the ground plane are in sync with and overlaying the waves being created. The reflector is quite near to the patches in the instance of a radiating patch. The waves will not be in phase if it is near to the patch.

A VSWR of 2 represents 11% of the reflected power. As demonstrated in Figure 6.2, a reflected power of 0 dB; where all of the power is reflected, but a reflected value of -10 dB shows that only 10% of the power is

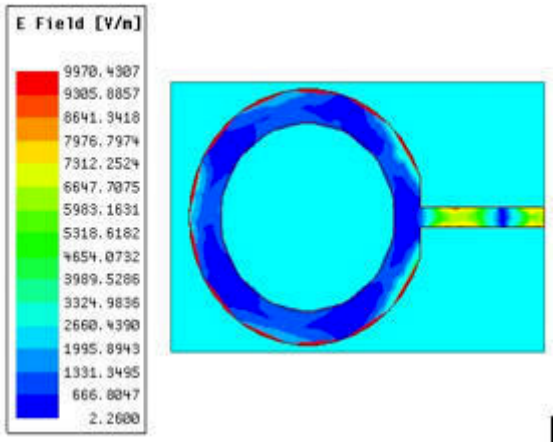


Fig. 5.1: A microstrip line feeds a rectangular patch.

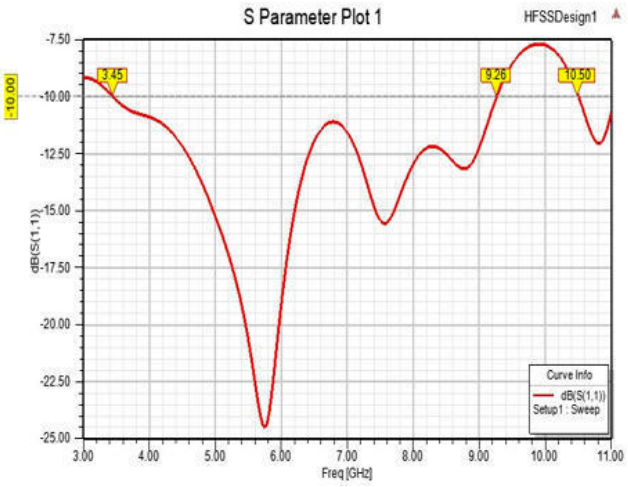


Fig. 6.1: Reflection coefficients are depicted graphically.

reflected. The VSWR would be infinite if 100% of the power was reflected.

The thinner the surface, the greater the loss factor due to increased leak of current density, and hence performance suffers significantly. Furthermore, a narrow track causes increased line impedance, which leads to loss and, as a consequence, a drop in efficiency.

Figure 6.3 depicts the current distributions at the microstrip patch. The passage of current across the ground plane is seen in Figure 6.4. Figure 6.5 shows the monopole antenna surface current: Surface current at 3.1 GHz, 5.7 GHz, 8.0 GHz, and 10.6 GHz (A-D).

It is observed that the dielectric substrate's thickness is quite thin; as a result, the height-dependent field fluctuation will be constant. When the electric field is normal to the patch's surface, the fringing effect fields at the patch's borders are similarly minimal. Five conformal antenna components are used to create the first antenna array, which yields a consistent gain of about 12 dBi with little scan loss over a wide range of scan angles. The antenna's electrical height is connected to crucial antenna properties such as gain, beam width, reflection coefficient, efficiencies, and so on.

The patch antenna has a resonant frequency. The patch and the ground create a low-impedance microstrip

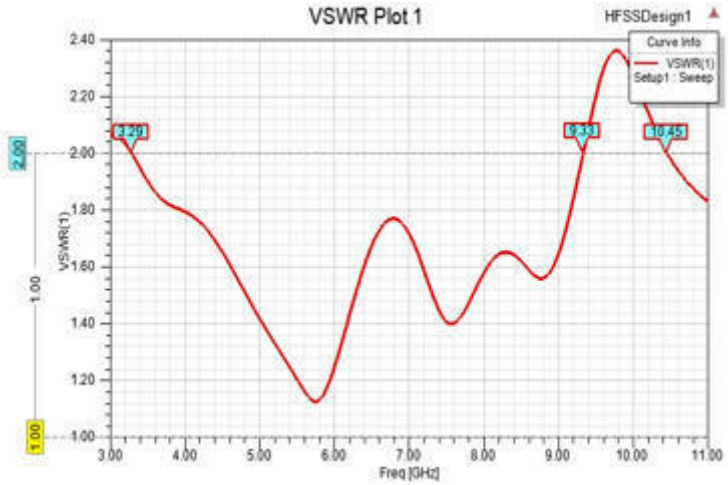


Fig. 6.2: VSWR Plot.

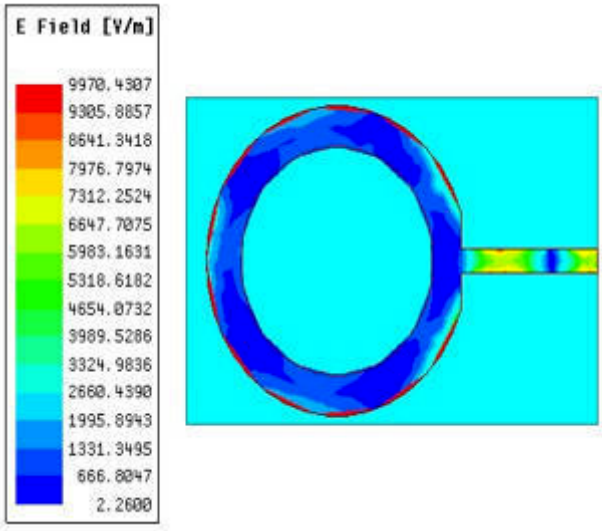


Fig. 6.3: Distributions of current at the microstrip patch.

half-wave resonator, and the fields are mostly in the dielectric underneath the patch. Because the wavelengths in the dielectric are shorter, the length of the patches must be smaller than half a wavelength in the atmosphere. The resistance of the resonator is determined by the size of the patches and the dielectric. Therefore for the same resistance, the patch must be narrower for large dielectric constants, or the dielectric must be thicker.

Table 6.1 shows the obtained values of different performance parameters of the proposed antenna model. The obtained results of Peak Gain = 2.374932, Peak Realised Gain = 2.367537, and Peak System Gain=2.367537.

7. Conclusion. A monopole broadband antenna is proposed in this plan for enlarging the data transfer capacity. A partial ground plane change builds the data transmission from 3 GHz to 37.26 GHz by 170%. Accordingly, the proposed system is accomplished the UWB at frequencies to 10.6 GHz. The antenna properties

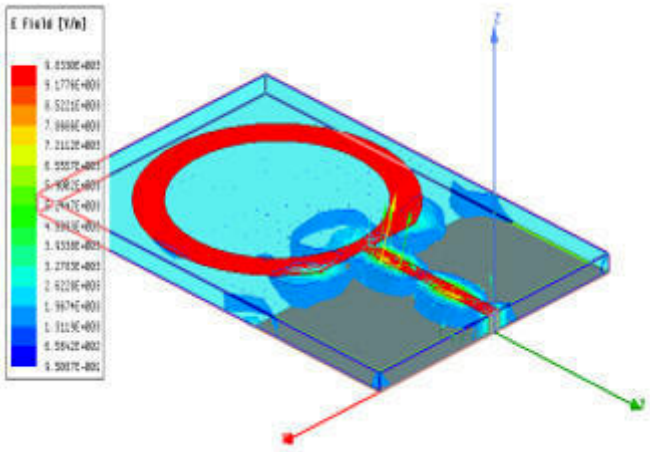


Fig. 6.4: Flow of current through the ground plane.

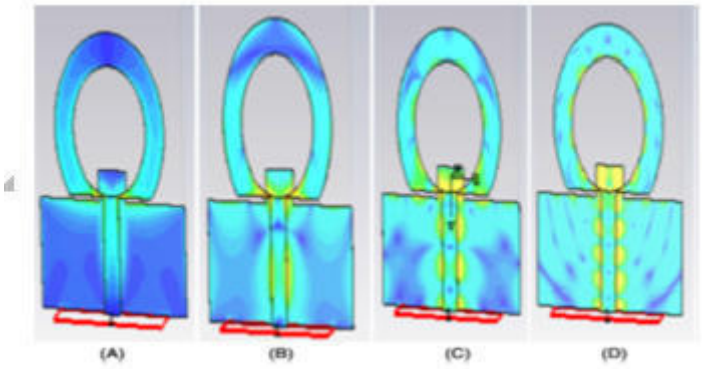


Fig. 6.5: Monopole antenna surface current at 3.1 GHz, 5.7 GHz, 8.0 GHz, and 10.6 GHz (A-D).

Table 6.1: Performance Metrics for the frequency with 5.76 GHz.

QUANTITY	VALUE
Max U	188.398178 mW/sr
Peak Directivity	2.693408
Radiated Power	879.011566 mW
Accepted Power	996.886391 mW
Incident Power	1.000000 W
Radiation Efficiency	88 %
Front to Back Ratio	9.051799
Decay Factor	0.000000
System Power	1.000000 W

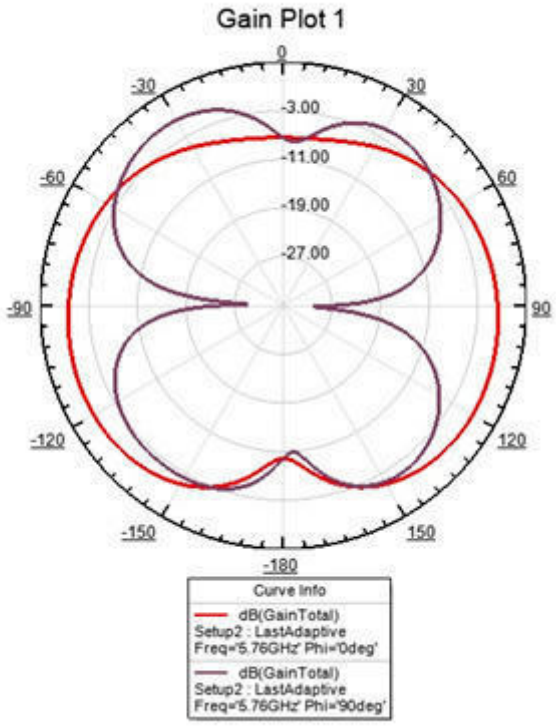


Fig. 6.6: The obtained gain plot.

are being tried tentatively, and it has been found that the antenna is a brilliant competitor for profoundly wideband applications. This plan explores the possibility of bandwidth improvement on an uncertain stubbed ground plane by the circumstance of a stub on the edge of the partial ground plane as an impedance matching instrument for the antenna. Since the antenna is being proposed for use in clinical applications, it should be analyzed to comprehend the impact of human body proximity, which will be accounted for in future correspondence.

REFERENCES

[1] IEEE 802.15 Working Group for WPAN. 2000.
[2] Soh PJ, Vandenbosch GAE, Ooi SL, Husna MRN. Wearable dualband Sierpinski fractal PIFA using conductive fabric. Electron Lett. 2011; 47(6):365-367.
[3] Soh PJ, Boyes SJ, Vandenbosch GAE, Huang Y, Ooi SL. On-body characterization of a dual-band, all-textile PIFA. Prog Electromagn Res. 2012; 129:517-539.
[4] Iwasaki H. A circularly polarized small-size microstrip antenna with a cross slot. IEEE Trans Antennas Propag. 1996;44(10): 1399-1401.
[5] Lilja J, Salonen P, Kaija T, de Maagt P. Design and manufacturing of robust textile antennas for harsh environments. IEEE Trans Antennas Propag. 2012; 60(9):4130-4140.
[6] Park S, Jayaraman S. Enhancing the quality of life through wearable technology. IEEE Eng Med Biol Mag. 2003; 22(3):41-48.
[7] Declerq F, Rogier H. Active integrated wearable textile antenna with optimized noise characteristics. IEEE Trans Antennas Propag. 2010; 50(9): 3050-3054.
[8] Lee HJ, Ford KL, Langley RJ. Switchable on/off body communication at 2.45 GHz using textile microstrip patch antenna on stripline, Electron Lett. 2012; 48(5):254-256.
[9] Axisa F, Schmitt P, Gehin C, Delhomme G, McAdams E, Dittmar A. Flexible technologies and smart clothing for citizen medicine, home healthcare, and disease prevention. IEEE Trans Inf Technol. Biomed. 2005; 9(3):325-336.

[10] Yang LQ, Giannakis GB. Ultra-wideband communications: an idea whose time has come. IEEE Signal Process Mag. 2004; 21: 26-54.
[11] Powell J, Chandrakasan A. Spiral slot patch antenna and circular disc monopole antenna for 3.1–10.6 GHz ultra-wideband communication, Proc 2004 ISAP, Sendai, Japan, 2004: 85–88.
[12] Ammann MJ, Chen ZN. Wideband monopole antennas for multiband wireless systems. IEEE Antennas Propag Mag. 2003;45(2):146-150.
[13] Liang J, Chiau CC, Chen X, Parini CG. Analysis and design of UWB disc monopole antennas. 2004 IEE Seminar on Ultra Wideband Communications Technologies and System Design, 2004: 103-106.
[14] Kalyan Mondal, Axial Ratio (AR) and Impedance Bandwidth (IBW) enhancement of Circular Polarized (CP) monopole antenna, AEU - International Journal of Electronics and Communications, Volume 134, 2021.
[15] Narinder Sharma, Sumeet Singh Bhatia, Ultra-wideband fractal antenna using rhombus shaped patch with stub loaded defected ground plane: Design and measurement, AEU - International Journal of Electronics and Communications, Volume 131, 2021.
[16] Hussein Alsariera, Zahriladha Zakaria, Azmi bin Awang Md Isa, New broadband L-shaped CPW-fed circularly polarized monopole antenna with asymmetric modified ground plane and a couple series-aligning inverted L-shaped strip, AEU - International Journal of Electronics and Communications, Volume 118, 2020.
[17] Utsab Banerjee, Anirban Karmakar, Anuradha Saha, Piyali Chakraborty, A CPW-fed compact monopole antenna with defected ground structure and modified parasitic hilbert strip having wideband circular polarization, AEU - International Journal of Electronics and Communications, Volume 110, 2019.
[18] Rohit Gurjar, Dharmendra K. Upadhyay, Binod K. Kanaujia, Amit Kumar, A compact modified sierpinski carpet fractal UWB MIMO antenna with square-shaped funnel-like ground stub, AEU - International Journal of Electronics and Communications, Volume 117, 2020.
[19] Xiaohua Tan, Weimin Wang, Yongle Wu, Yuanan Liu, Ahmed A. Kishk, Heng Wang, Enhancing isolation and bandwidth in planar monopole multiple antennas using thin inductive line resonator, AEU - International Journal of Electronics and Communications, Volume 117, 2020.
[20] Rohit Mathur, Santanu Dwari, Compact CPW-Fed ultrawideband MIMO antenna using hexagonal ring monopole antenna elements, AEU - International Journal of Electronics and Communications, Volume 93, Pages 1-6, 2018.

Edited by: Vinoth Kumar
Received: Aug 13, 2022
Accepted: Nov 18, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 363–376, DOI 10.12694/scpe.v23i4.2047

DESIGN AND DEVELOPMENT OF A LOW-COST SENSOR IOT COMPUTING DEVICE FOR GREENHOUSE GAS MOMITOR FROM SELECTED INDUSTRY LOCATIONS

IBRAHIM HAMIDU *, BENJAMIN AFOTEY †, AND ZAKARIA AYATUL-LAHI‡

Abstract. The objective of the study is to develop low-cost IoT based sensor to monitor real-time greenhouse gases (GHG) emissions data from selected industry locations (city blocks) in a top-down approach. Three (3) industry locations were selected within the Suame Industrial complex (the largest single cluster of artisanal engineering and light manufacturing in Sub Saharan Africa and even Africa) which has no reported GHG emissions data. A GHG monitor was developed using Atmega328 microcontroller and a sim800I GSM module was used to collect a 24-hour real-time minute-by-minute emissions data from the selected industry locations. A MQ-4 (methane/natural gas sensor), MQ-135 (Nitrous Oxide sensors) and DHT22 (temperature and humidity sensor) were used in the GHG monitor design. The GHG of concern were carbon dioxide, methane and nitrous oxide. A total of 3627 emissions data were collected and analyzed from the three (3) industry locations. Location 3 had the highest average carbon dioxide emissions of 508.11 ppm, followed by location 2 with 477.31 ppm with the least emissions in location 1 with 472.51 ppm which are above the global carbon dioxide average of 414.7 ppm. The average methane emission was highest in location 1 with 0.1599 ppm (1599 ppb), followed by location 3 with 0.1366 ppm (1366 ppb) with the least average methane emission of 0.1358 ppm (1358 ppb) in location 2 which are slightly below the global methane average of 1895.7 ppb. The MQ-135 nitrous oxide sensor reported zero emissions data throughout the deployment at the various industry locations which indicated the nitrous oxides emission in the selected sample site is negligible or below the detectable range of the sensor.

Key words: Low-cost-sensor IoT design, GHG monitor, Real-time emissions data, Regression Analysis, Ghana.

AMS subject classifications. 68M14, 68M18

1. Introduction.

Greenhouse gases (GHG) are considered the main contributor to global warming and climate change [1] but the inefficiencies in methods used in the GHGs estimation has created huge emissions and sink data gaps.Numerous Sub Saharan African countries such as Ghana lack the capacity to effectively conduct Air Quality monitoring and meeting the Air Quality standards [10].

An estimated 75% of people in developed countries live in cities as compared to 35 % of people in developing countries [24]. There is a projected increase in urban (city) dwellers, with an estimated 66% of the total human population living in cities by 2050 [11]. The movement of majority of the population into urban cities will lead to significant increase in energy demand [11] and a drastic increase in air population within Sub-Saharan Africa [25]. Greenhouse Gases which are air pollutants absorb and radiate heat back to the earth surface increasing the atmospheric temperature [21].

Air pollution which result from the release of harmful solid, liquid or gaseous suspended particles has been estimated to cause over 28,000 deaths in Ghana in 2018 [25]. Findings suggest that among the impact of climate change in Ghana, there is a high temperature projection and low rainfall in the years 2020, 2050, and 2080 and a projected desertification rate of 20,000 hectares per annum [4].

The main estimation approaches of GHG estimation are the bottom-up and top-down approaches.Top-down approach involves direct ambient or atmospheric emissions concentration measurement whereas bottom-up approach involves the direct emissions measurement from the individual emissions sources [2].The EPA in Ghana, responsible for the monitoring and reporting of GHG emissions uses the IPCC 2006 guidelines and software with National and international GHG inventory system [5]. The IPCC 2006 guidance is based on

*Department of Chemical Engineering, Kwame Nkrumah University of Science and Technology, Private Mail Bag, University Post Office, KNUST – Kumasi, Ghana (hamiduibrahim75@gmail.com)

†Department of Chemical Engineering, Kwame Nkrumah University of Science and Technology, Private Mail Bag, University Post Office, KNUST – Kumasi, Ghana. (ba_2021@yahoo.com).

‡Private Research, ZAKS Electronics, Post Office Box WE 586, Asawase Kumasi, Ghana (ayatullahizakaria@gmail.com)

bottom-up models (activity and impact factor based) which requires several assumptions and yields uncertainties in the results.

Low Cost Sensors (LCS) over time have experienced tremendous developments and application in greenhouse gas accounting and reporting at lower levels (city blocks) [31]. However, the use of low cost sensor in Ambient Air Quality monitoring is limited to Particulate matter only in Ghana evidenced in the Ghana Urban Air Quality Project (GHAir) [10, 25]. The use of the low cost sensors in the GHAir project has also faced limitations in reporting monitored data due to poor Wi-Fi connectivity and power supply in deployed areas [25].

The objective of this research is to determine ambient greenhouse gases within selected industry locations within the Suame Industrial Complex using a Top-Down approach. In this research study, a low-cost-sensor IoT based GHG monitor to report real-time GHG emissions data was designed and developed. The developed monitor was then deployed within selected industry locations within the Suame Industrial Complex to collect a 24-hour minute-by-minute GHG emissions data. A statistical tool (Minitab Statistical Software 20.3 (64 bit)) was used in analyzing reported/collected GHG emissions data.

2. Literature Survey. The phenomenon of air pollution falls under SDG 3 – Good Health and Wellbeing and specifically SDG 3.9. SDG 3.9 targets the substantial reduction in illnesses and death associated with the release of hazardous pollutants. Air pollution also affects SDG 11 – Sustainable Cities and Communities, and Specifically Target 11.6. Target 11.6 emphasizes the reduction of the environmental impacts of cities on people to enhance a safe and inclusive human settlement [25]. Greenhouse Gases such as Ozone, Methane, Carbon Dioxide, Nitrous Oxide, and Water Vapour absorb and emit radiant energy in the earth atmosphere within the Thermal Infrared Spectrum/region [21]. Greenhouse gases are predominantly made of Carbon Dioxide and Methane [32] and are emitted/released mostly through human-made (anthropogenic) activities mainly associated with fossil fuel use [21]. The absorption and emission of radiant energy leads to Greenhouse effect which in excess can lead to catastrophic effect due to increase in atmospheric temperature to unbearable limits [21]. Anthropogenic methane sources account for 60% whereas Natural gas sources account for 40% of the global methane emissions [18].

The uncertainty of CO2 emissions from Land Use according to the 4th Assessment Report is estimated at $\pm$ 2933 MtCO2 at the 1990 global level [33] with an average temperature increase over land by $1.320.04^oC$ [34]. The average temperature increase of the ocean is also estimated as $0.590.006^oC$ [34]. The increase in atmospheric temperature poses a threat to aquatic and human survival [32]. There has been an estimated atmospheric temperature increase of 0.7^oC from 1961 to 1990 [34].

Carbon Dioxide is the highest contributor to Global Warming and the most significant greenhouse gas [13]. The concentration of Carbon Dioxide increased from 280 ppm at a pre-industrial level value to 410 ppm in 2020 [35]. Although the methane increase within the atmosphere which constitute 16% of GHG emissions is not as high as carbon dioxide, it however contributes as much as carbon dioxide due to its Global Warming Potential [18]. Without the Greenhouse effect, the earth atmospheric temperature will be at an unbearable -18^oC [21], and in excess (due to the drastic increase in GHG concentration) will also lead to unbearable high atmospheric temperature. Although the global GHG emissions slowed after 2014, the emissions generally increased from 2010 to 2018 and peaked at 58 GtCO2eq in 2018, 11% above 2010 levels [36, 37].

In order to mitigate the effects of greenhouse gases due to sustained increase in GHG emissions, several international agreements have been formed. Among them is the Paris agreement which seek to limit the global atmospheric temperature increase to 2^oC and possibly 1.5^oC above preindustrial era levels [36]. Among the estimation techniques used in GHG accounting are the top-down and bottom-up approaches.

Top-down estimates uses meteorological data like wind speeds and directions to obtain emissions data through inverse modeling by projecting atmospheric transportation. Inverse modelling is based on source-receptor relationship atmospheric data based [38]. The Top-down approach which monitors the atmospheric emissions provides an emissions aggregate from a location but does not identify emissions sources unless tracers are identified for specific emissions [2].

Bottom-up analysis includes life-cycle analysis [39]. Bottom up estimation is from emissions inventory which are heterogeneous from one location to another and prone to uncertainties, errors and manipulations. An example of the bottom-up approach is the emissions factor approach as used in the IPCC guidelines [38].

The emissions data on the global scale reported by countries using the bottom-up approach is approximately less than half of the emissions data from the top-down approach [38]. Omissions of emissions sources account for the weakness of the bottom up approach [2].

The effectiveness of the GHG accounting and reporting methodologies rely on drivers and motives which are characterized by the level of development of a country - whether a country is develop or developing. Developed country have much obligations to gather, validate, model data and used in estimating and reporting whereas developing countries have voluntary obligation and most lack the required data. The outcome of GHG accounting and reporting in developing countries is characterized by the lack of necessary data, expertise, resources to enhance the effective evaluation of GHG. Lack of the up to date generation data also makes the results from the various models used in GHG accounting and reporting incoherent [40].

Among the four main GHG (bottom-up) accounting methodologies are the IPCC method for national accounting, corporate level accounting (municipalities), Life Cycle Analysis/Assessment and Carbon Trading methodologies. The results from these methods varies significantly when applied to the same activity and geographies based on the assumption made [40].

Model based GHG estimation (GHG accounting and Reporting) such as the IPCC method is widely accepted. The Tier-1 of the IPCC model is used for national inventory reporting [5] whereas the Tier-2 can be applied to various sub sectors such as emissions from waste sources [41]. Other models such as the LandGEM method and the Triangular method were applied by [41] in estimated landfill gas (methane).

The authors of [31] assessed in their research the adaptation of the IPCC model for small cities in carbon monitoring. In their research, the use of the IPCC model was asserted as ineffective in estimating carbon emissions on small cities level due to lack of timely real time monitoring data. Lack of intelligent energy monitoring system (due to high cost) to monitor and report carbon inventory data was identified among the limitation of adapting the IPCC model for small cities. The last identified limitation is the inability to universally adapt the IPCC model for different levels in small cities due to the differences in estimation method of monitoring data and long term data [31].

Among the development in the monitoring of GHG within the earth atmosphere in recent times involved the use of Satellites. Satellites are used to fill the spatial GHG monitoring gaps within the earth's atmosphere. The Thermal And Near infrared Sensor for Carbon Observation (TANSO) launched by Japan in 2009 which aboard Greenhouse gases Observing SATellite (GOSAT) was the first satellite instrument designed specifically to record methane and Carbon Dioxide concentrations [13]. Other satellites such as the Orbiting Carbon Observatory - 2 (OCO-2) launched by NASA in 2014 and the TanSat (also known as CarbonSat) launched by the Chinese Space Program in 2016 are dedicated to monitor GHG (carbon dioxide) in the Earth's atmosphere [35]. Other instruments such as the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) launched in 2002 and lost in 2012 by the European Space Agency (ESA) which aboard the ENVIronmental SATellite (ENVISAT) could retrieve GHG concentrations [13].

Low Cost Sensors (LCS) are characterized by their low purchase pricing and operation cost as compared to the purchase and operating cost of other sensors/analyzers [17]. Among the LCS are the electrochemical gas sensors, semiconducting gas sensors, and the optical particle monitors [17, 21].

One of the main advantages and motivations for using Low Cost Sensors is in spatial resolution increase of Ambient air quality. This enhances the detection of variability below a regional/city level, to lower levels such as city-block or below. The size and low cost of sensors provides the opportunity for deploying in small footprint locations and in mobile/semi-mobile sampling schemes. Current models techniques cannot compete at lower levels of spatial resolutions due to high dynamic variability of the pollutants, and external hindrances such as wind, humidity etc… which is overcome by the use of LCS. Mostly Low Cost Sensors marketed as a single pollutant sensor exhibit cross-sensitivity to other pollutants [8].

Among the various approach used in sensor selectivity are temperature, filters, catalyst and specific adsorption. Temperature changes affects the spontaneity/rate of a reaction (the activity of a reducing agent) thereby affecting the selectivity of the sensor. Methane burns at high temperature compared to the low temperature (room temperature) at which CO and hydrogen burns. Hence at elevated temperatures (high above room temperature) at which the gases burn rapidly, it is likely gases such as hydrogen and CO will not be detected [42].

Specific materials that allow specific gases to react can be used to enhance sensor selectivity. Hygroscopic salts are used to attract/concentrate water for humidity sensors. Another example is the use of sulfanilic acid to enhance the selectivity of NO2 [42].

Metal Oxides as used in sensor development varies the surface conductance in a gaseous atmosphere. Various metal oxides are used as catalyst or dopants to enhance to response to some specific gases. Tin oxide (which is the most studied material) modifies the response spectrum of the sensor to gases by enhancing suitable changes of the catalyst or dopant deposited. Among the dopants/catalyst are palladium, platinum and cadmium [43].

3. Materials and Methods.

3.1. Selected Sample Site. The selected sample site for this project is the Suame Industrial Complex. The Suame Industrial Complex mainly known as Suame Magazine comprises over 12,000 enterprises and over 200,000 artisans involved in operations such as equipment and machinery manufacturing, foundries, vehicle repairs, materials workings etc. The Suame Magazine spans 900,000 square meters, 1.8 km long and 0.32 km wide with an estimated cluster nucleus perimeter of 7 km. [3, 14, 20, 29].

Three (3) industry locations were selected within the selected sample site. The selected location 1(Lat: 6.7211, Long: -1.6291) is the Abu-Dia Company Limited with operations in metal foundry. Location 2 (Lat: 6.7113, Long: -1.6306) is KOSAMO Limited with operations in HVAC operations and duct fabrication. Location 3 (Lat: 6.7117, Long: -1.6422) is Presank Company Limited with operations in waste handling and recycling.

3.2. Selected pollutants of concern. This project is limited to Carbon Dioxide, Methane and Nitrous Oxide due to the quantities release and their potency in trapping heat within the earth atmosphere

Carbon dioxide can potentially be produced from methane oxidation [22] and is the most oxidized state of Carbon, which is used as a chemical reagent and a phosgene substitute [44]. The average global carbon dioxide emissions level in 2021 was 414.7 ppm, a 2.66 ppm increase from the 2020 average. A 35 billion tons of carbon dioxide were emitted into the atmosphere in 2021, and these carbon dioxide emissions have the potential of trapping heat within the atmosphere for thousands of years [23].

Methane is a stable gas which starts decomposing into elements at 785^oC [45]. A graph of the average global methane emissions per year is shown. As observed by [23], the global average methane emissions in 2021 was 1895.7 ppb, a 15% higher than the 1984 – 2006 period and a 162% higher than the pre-industrial level. The methane atmospheric residence time is about 9 years but is 25 times more potent in trapping heat in the atmosphere, and a short-term climate change influencer as compared to carbon dioxide [23].

Nitrous Oxide is a colourless, slightly pleasant sweet odour and taste and non-flammable gas which is acknowledged as an undesired by-product in exhaust gases from lean NOx trapping [15]. Agriculture dominates the global anthropogenic sources of N2O [22]. Among the sources of Nitrous Oxide are industries, fertilizer production, and wastewater treatment plant [15, 28]. Nitrous Oxide is currently the largest destroyer of ozone in the stratosphere than any other reactive chemical [28]. The Global Warming Potential of Nitrous Oxide is between 265 to 298 over a 150 years' time series [15].

3.3. Development of an IoT GHG monitor. The block diagram of the monitor design is shown in figure 3.1. The various components including the sensors used in the monitor design and development are shown in figure 3.2. The sensors used in this project are MQ-135 (nitrous oxide sensor), MQ-4 (methane sensor), SCD-30 (carbon dioxide, temperature and humidity sensor), and DHT22 (temperature and humidity sensor).

MQ-135 is a Tin Dioxide (SnO_2) semiconductor based gas sensor effective in detecting Nitrous Oxide, Ammonia, Benzene, Smoke, alcohol and Carbon Monoxide [1, 21]. MQ-4 is composed of Al_2O_3 ceramic tube, Tin Dioxide (SnO_2) sensitive layer, measuring electrode and heater that detects the concentration methane/natural gas between 0 ppm - 10000 ppm [6].

Sensiron SCD-30 operates on the non-dispersive infrared(NDIR) principle which measures CO_2 between 0 - 40000 ppm ($\pm$ 30 ppm + 3 % accuracy for range between 400 ppm – 10000 ppm), temperature between - 40 to 120 ^{o}C, and relative humidity between 0 % - 100 % [19]. NDIR sensors measure the CO_2 levels by measuring the amount of infrared light reflected by the CO_2 [9].

DHT22 contains a thermistor for dry bulb temperature measurement and a capacitive sensor for moisture/humidity measurement. The DHT22 measures temperature between -40 to +80 o C (0.5 o C accuracy)

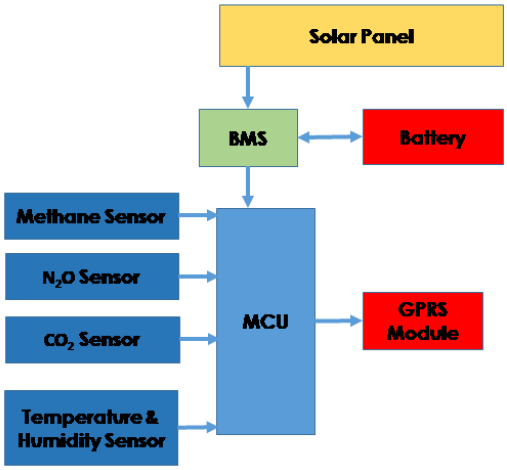


Fig. 3.1: Block diagram of the various units of the GHG monitor

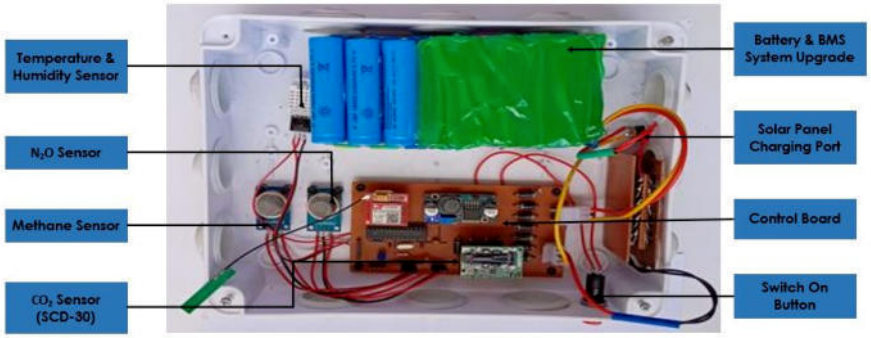


Fig. 3.2: Various components in the developed monitor

and relative humidity from 0% to 100% (2% accuracy) [46, 47].

The DHT22 sensors measure 4% more accurate temperature readings and 18% more accurate humidity readings than its DHT11 [48]. The various sensors were calibrated using libraries from the sensor manufacturers designated for the working conditions of the sensors based on the various location. The calibration of the sensor was to enhance better performance and to enhance the accuracy of the monitored emissions.

Among the other various components incudes;

1. Battery pack (19.6 Ah)
2. A battery Management System
3. A 50 W solar panel
4. An Atmega 328 microcontroller unit
5. A sim800I GSM module
6. Im2596 buck converter
7. A 28 pin IC socket
8. 1N5408
9. A 2.54 male and female pin headers
10. 16 MHz crystal
11. A 22 Pico farad, 100 Nano farad capacitors

Fig. 3.3: Various components in the developed monitor

Fig. 3.4: Images of the deployment at a. location 1 (lat: 6.7211, long: -1.6291); b. Location 2 Images of the deployment at a. location 1 (lat: 6.7211, long: -1.6291); b. Location 2

12. 10 k ohms, 20 k ohms, and a 1 k ohm resistors
13. A Vh3.96 connector

The sim800I GSM module was integrated into the monitor to enhance the transmission of the monitored emissions data onto a Thing Speak cloud platform. The Thing Speak (by Mathworks) IoT analytic platform was selected for this project due to the easiness in data download from the platform, easy monitoring and visualization of live reported data on the platform. A private access was created. Screenshots of the mobile view interface of the ThingSpeak cloud platform is shown in figure 3.3.

3.4. GHG monitor deployment. The developed monitor was deployed in location 1(Lat: 6.7113, Long: -1.6306) from 2022-07-19 (18:00:10 GMT) to 2022-07-20 (17:59:42 GMT) and a total of 1286 data collected within the 24 hours' deployment. The monitor was then deployed in location 2 (Lat: 6.7211, Long: -1.6291) with 1246 total data collected within 24 hours deployed period. The developed was lastly deployed at location 3 (Lat: 6.7117, Long: -1.6422) with 1095 data collected in the 24 hours deployed period. Imaged of the monitor deployed in location 1,2 and 3 is shown in figure 3.4.

3.5. Statistical analysis of collected GHG emissions data. Among the statistical analysis performed on the collected data are descriptive statistics analysis, Time series plots, Surface plots, Correlation and Regression analysis of the data.

A correlation coefficient (r) as used in correlation analysis is the measure of the linear relationship between two (2) variables. The Pearson product-moment correlation coefficient (r) is a parametric measure of the linear association of normal distribution between two (2) variables. Conversely to the Pearson coefficient. [26].

Monitored environmental factors (temperature and humidity) are correlated with the monitored greenhouse

Table 4.1: Summary of descriptive statistics of reported pollutant emissions data by the various selected sensors in various locations.

Location (Lat, Long)	Descriptive Statistics	Pollutant/Analyte (ppm)		
		CO2 (SCD-30)	CH4 (MQ-4)	N20 (MQ-135)
Location 1 (Lat, Long)	Mean	472.51	0.1599	0
	Min	421.00	0.0800	0
	Max	513.00	0.5100	0
	Standard Deviation	11.59	0.08542	0
Location 2 (Lat, Long)	Mean	477.31	0.1358	0
	Min	435.00	0.0800	0
	Max	515.00	0.3700	0
	Standard Deviation	12.44	0.05857	0
Location 3 (Lat, Long)	Mean	508.11	0.1366	0
	Min	439.00	0.0800	0
	Max	585.00	04800	0
	Standard Deviation	11.43	0.05602	0

gases in this research study. Temperature and relative humidity were selected in the correlation investigation with the various emissions data due their influence in sensor performance and greenhouse gases concentration and pollutant chemistry as outlined in literature [8, 12, 16, 27, 30].

Regression Analysis is modeling of a set of numerical data using a set of methods and relates a dependent (response) variable to one or more independent (explanatory) variables [26].

Minitab Statistical Software 20.3 (64 bit) was the only statistical software used for all the data analysis. The listed analysis was done on the 3627 data collected in the final deployment.

The time-stamped minute-by-minute reported emissions data made it necessary to conduct time series plots of the reported emissions data from the various locations. The time series plot was also used to investigate performance of DHT22 and SCD-30 sensor.

4. Findings and Results.

4.1. Descriptive statistics of reported data. The descriptive statistics (mean, minimum value (min), maximum value (max) and the standard deviation) of the reported carbon dioxide, methane and nitrous oxide emissions data are shown in table 4.1. The descriptive statistics of the temperature and humidity reported values by the DHT22 and SCD-30 sensors are shown in table 4.2.

4.2. Correlation analysis. The effects of temperature and humidity was investigated on the monitoring of the greenhouse gases. As indicated in by [8] that these factors (temperature and humidity) affect sensor performance. The examination of temperature and humidity effects was also investigated on the dispersion or production/emission of carbon dioxide and methane as investigated in correlation analysis by [27] and [12]. The correlation analysis of carbon dioxide, and methane with temperature and humidity is shown in table 4.3.

4.3. Regression Analysis. Based on the correlation analysis as shown in table 4.2, regression analysis of carbon dioxide and methane (as response variable) against temperature and humidity (as independent) variable was investigated, with the corresponding coefficient of determination (R^2) as shown in table 4.4 and 4.5. The DHT22 sensors temperature and humidity data was used in all the regression analysis.

The regression analysis of temperature against humidity was investigated as shown in table 4.6.

4.4. Time Series Plots. Time series plot was used to visualize the reported emissions data. The time series plot of carbon dioxide and methane emissions data, and environmental factors (temperature and humidity) data for the various deployed locations are shown in figure 4.1 to 4.6 below.

Time series plot of reported data in location 1 (lat: 6.7211,-1.6291)

Time series plot of reported data in location 2 (lat: 6.7113, -1.6306)

Table 4.2: Summary of descriptive statistics of reported environmental factors data by the various selected sensors in various locations

Location (Lat, Long)	Descriptive Statistics	Environmental Factors			
		Temperature (o C)		Humidity (%)	
		(SCD-30)	(DHT22)	(SCD-30)	(DHT22)
Location 1 (Lat, Long)	Mean	28.20	26.47	62.91	70.57
	Min	24.40	22.90	37.79	36.50
	Max	36.48	33.90	75.97	87.50
	Standard Deviation	3.14	3.11	10.97	14.94
Location 2 (Lat, Long)	Mean	27.95	26.77	65.00	70.24
	Min	23.04	23.40	48.20	51.00
	Max	34.62	32.30	80.56	84.50
	Standard Deviation	3.34	2.56	10.07	9.99
Location 3 (Lat, Long)	Mean	30.69	28.20	62.51	72.62
	Min	27.65	25.30	39.58	46.00
	Max	40.01	35.40	71.77	85.80
	Standard Deviation	2.43	2.47	7.51	11.03

Table 4.3: Summary of correlation details from the various locations

95 % Confidence Interval Pearson Correlation			
Location Details		Temperature	Humidity
Location 1 (lat: 6.7211, long: -1.6291) 1286 Data	CO2	-0.258	0.252
	CH4	0.146	-0.144
	Temperature		-0.996
Location 2 (lat: 6.7113, long: -1.6306) 1246 Data	CO2	-0.704	0.696
	CH4	-0.097	0.113
	Temperature		-0.991
Location 3 (lat: 6.7117, long: -1.6422) 1095 Data	CO2	-0.234	0.258
	CH4	0.096	-0.074
	Temperature		-0.993

Table 4.4: Summary of regression analysis of carbon dioxide against temperature and humidity from the various deployed locations

Location details	Regression equation	R2 (%)
Location 1 (lat: 6.7211, long: -1.6291) 1286 Data	CO2 = 607.0 - 3.61×Temperature - 0.552× Humidity	7.02
Location 2 (lat: 6.7113, long: -1.6306) 1246 Data	CO2 = 586.1 - 3.802×Temperature - 0.100× Humidity	49.55
Location 3 (lat: 6.7117, long: -1.6422) 1095 Data	CO2 = 177.2 + 7.03×Temperature + 1.828× Humidity	10.00

Time series plot of reported data in location 3 (lat: 6.7117, -1.6422)

5. Discussion. Carbon dioxide emissions were averagely the highest in location 3 with 508.11 ppm concentration followed by location 2 with 477.31 ppm concentration and lastly location 1 with 472.51 ppm. Location 3

Table 4.5: Summary of regression analysis of methane against temperature and humidity from the various deployed locations

Location details	Regression equation	R2 (%)
Location 1 (lat: 6.7211, long: -1.6291) 1286 Data	CH4 = -0.124 + 0.00833×Temperature + 0.00090× Humidity	2.15
Location 2 (lat: 6.7113, long: -1.6306) 1246 Data	CH4 = -0.769 + 0.01922×Temperature -+0.00555× Humidity	2.57
Location 3 (lat: 6.7117, long: -1.6422) 1095 Data	CH4 = -1.414 + 0.03560×Temperature -+0.00753× Humidity	4.14

Table 4.6: Summary of regression analysis of temperature against humidity from the various deployed locations

Location details	Regression equation	R2 (%)
Location 1 (lat: 6.7211, long: -1.6291) 1286 Data	Temperature = 41.0898 - 0.207153× Humidity	99.26
Location 2 (lat: 6.7113, long: -1.6306) 1246 Data	Temperature = 44.6179 - 0.254177× Humidity	98.19
Location 3 (lat: 6.7117, long: -1.6422) 1095 Data	Temperature = 44.3276 - 0.222073× Humidity	98.54

also recorded the highest minimum carbon dioxide concentration (439 ppm), followed by location 2 (435.00 ppm) and the least in location 1 (421 ppm) which are all above the 2021 global carbon dioxide concentrationaverage in the atmosphere of 414.7 ppm as observed by [23].

The average methane concentration was highest in location 1 (0.1599 ppm equivalent to 1599 ppb) followed by location 3 (0.1366 ppm equivalent to 1366 ppb) and lastly location 2 (0.1358 ppm equivalent to 1358 ppb) as shown in table 5.1. An average methane concentration recorded in all locations are slightly below the global atmospheric methane average. However, the maximum recorded methane emissions of 0.5100 ppm (5100 ppb), 0.3700 ppm (3700 ppb) and 0.4800 ppm (4800 ppb) in location 1, 2 and 3 respectively are above the 2021 global atmospheric methane average (1895.7 ppb) as observed by NOAA, (2022). A 0.08 ppm minimum methane concentration in all deployed industry locations indicates a methane background concentration of 0.08 ppm (80 ppb) within all the selected industry location and the sample site as a whole.

The reported emissions values of the MQ-135 nitrous oxide sensor were zero (0) throughout the entire deployments as shown in all the reported data as summarized also in table 4.1. This indicates that the nitrous oxide emissions are not at detectable levels to be detected by the MQ-135 nitrous oxide sensor. It is also evident according to [15]; [22]; [28] that nitrous oxide sources are mainly agriculture activity related, and was logical that the nitrous oxide emissions within all deployed industry locations within the sample site Suame Industrial Complex are zero (0).

The difference in the average temperature data reported by the DHT22 sensor and SCD-30 sensor are less than 2 ^{o} C for location 1 and 2, and less than 2.5 ^{o} C for location 3 as shown in table 4.2. The SCD-30 sensor reported the higher temperature values in all locations. This shows the accuracy of reported temperature data by the sensors. However, the difference in reported humidity data by DHT22 and SCD-30 sensors was less than 8 % for location 1 and 2, and approximately 10 % in location 3. The DHT22 sensor reported the much higher values of humidity in all locations. This shows the precision in the reported data by the SCD-30 and DHT22 sensors.

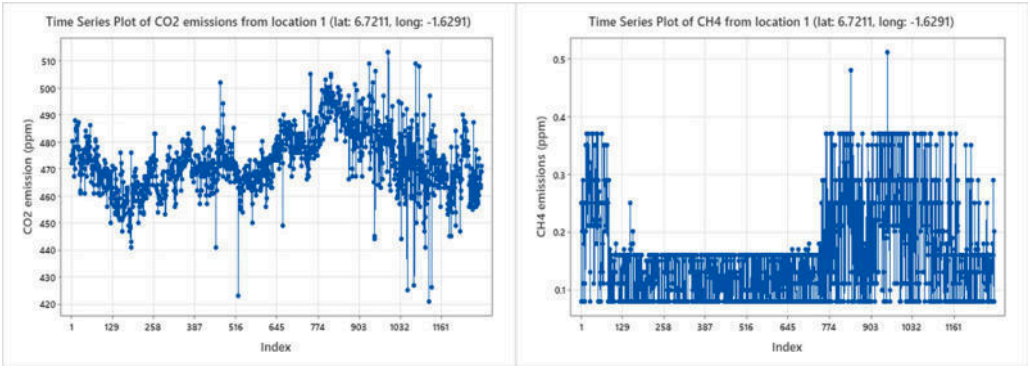


Fig. 4.1: Time series plot of reported carbon dioxide and methane emissions data from location 1 (lat: 6.7211, long: -1.6291)

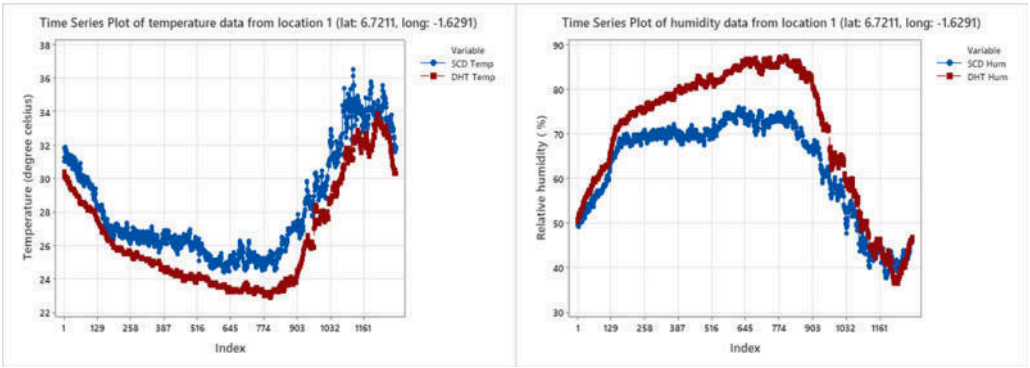


Fig. 4.2: Comparison of temperature and humidity data from location 1 (lat: 6.7211, long: -1.6291) by SCD-30 sensor and DHT22 sensor using time series plot.

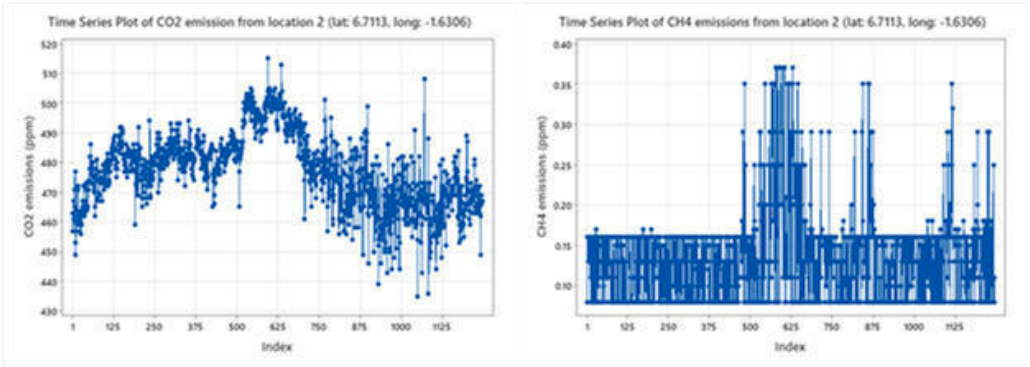


Fig. 4.3: Time series plot of reported carbon dioxide and methane emissions data from location 2 (lat: 6.7113, long: -1.6306)

As shown in table 4.3, carbon dioxide has a weaker negative correlation and weaker positive correlation with temperature and humidity respectively in location 1 and 3. However, the correlation of carbon dioxide with temperature and humidity in location 2 was moderately negative and positive of -0.704 and 0.696 respectively.

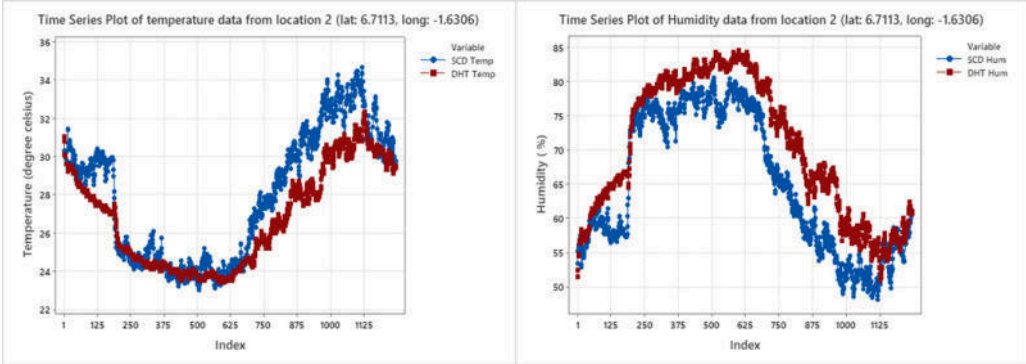


Fig. 4.4: Comparison of temperature and humidity data from location 2 (lat: 6.7113, long: -1.6306) by SCD-30 sensor and DHT22 sensor using time series plot.

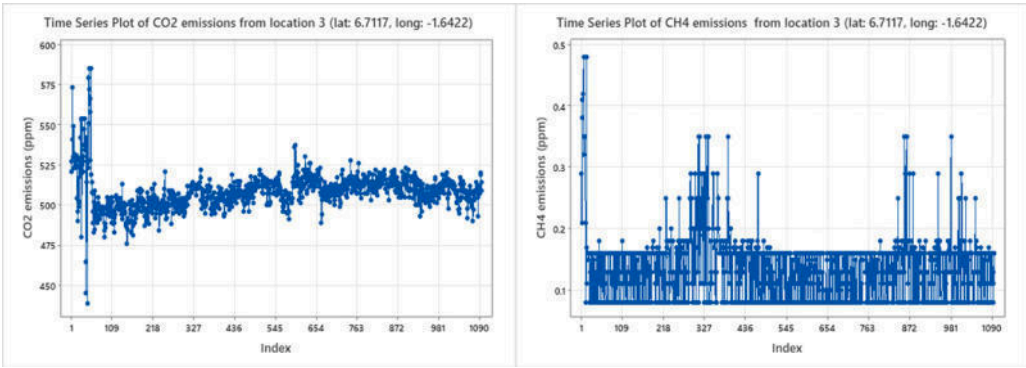


Fig. 4.5: Time series plot of reported carbon dioxide and methane emissions data from location 3 (lat: 6.7117, long: -1.6422)

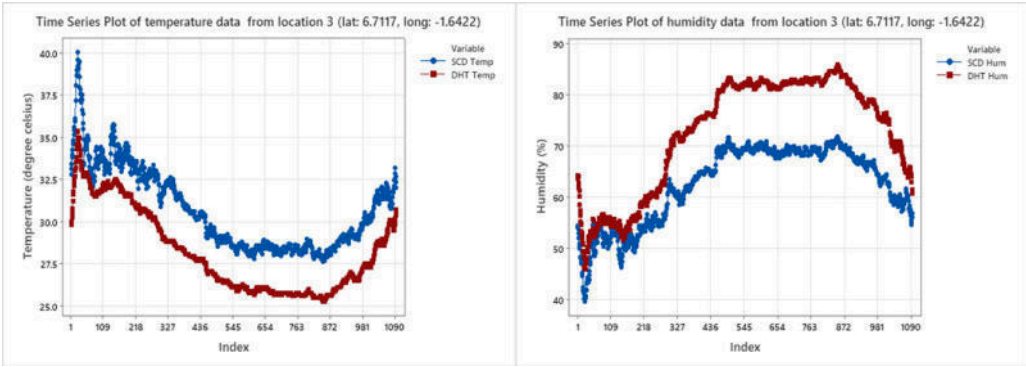


Fig. 4.6: Comparison of temperature and humidity data from location 3 (lat: 6.7117, long: -1.6422) by SCD-30 sensor and DHT22 sensor using time series plot.

This shows that either the dispersion/emissions of carbon dioxide decreases with rise in temperature and increases with increase in relative humidity. This deviate from the positive correlation of carbon dioxide and humidity as observed by [30] in their carbon dioxide emissions correlation with temperature and humidity

investigation. This also shows that the dispersion/emission of carbon dioxide is affected differently in various localities and hence much studies has to be done in various environments.

The correlation of methane with temperature and humidity was weaker (almost negligible) in all three (3) industry locations. Methane showed a negative correlation with temperature and a positive correlation with humidity in location 1 and 3. The correlation of methane in location 2 was positive with humidity and negative with temperature as also observed by [27] in their research. The variation of the correlation also suggests that, the correlation is dependent on the particular environment and much studies has to be done is GHG pollutant emissions/dispersion.

The correlation between temperature and humidity was very high in all three (3) locations with over 99 % of all the data completely correlating.

Based on the correlation analysis as shown in table 4.3 above, regression analysis was investigated with the corresponding coefficient of determination (R^2). As shown in table 4.5 above, the regression equation between methane emissions as the response variable (dependent variable) and temperature and humidity as the independent variables had the least coefficient of determination (R^2) values less than 5 % in all location. The almost negligible coefficient of determent (R^2) can be tallied with the almost negligible Pearson correlation coefficient less than 0.2 in all locations.

The coefficient of determination (R^2) of the regression equations with carbon dioxide emissions as the response variable (dependent variable), and temperature and humidity as the independent variable was less than or equal to 10 % for location 1 and 2 and approximately moderate (49.55 %) for location 2 as shown in table 4.4 above. This shows that temperature and humidity have an effect on carbon dioxide dispersion as observed in correlation research by [30, 12] in an indoor environment.

Much higher coefficient of determination (R^2) above 98 % was observed for the regression equation between temperature and humidity for all three (3) location as shown in table 4.6 as also observed by [30]. The temperature-humidity regression equation can be used in heat transfer analysis (temperature-humidity equations) as a calibration equation for temperature and humidity measurements/sensors.

The Sensiron SCD-30 carbon dioxide sensor, the DHT22 (AM2302) temperature and humidity sensor, and the MQ-4 methane sensor had a stable performance throughout its deployment.The MQ-135 nitrous oxide sensor reported zero emissions data which shows the nitrous oxide emissions are not in the detectable range of the sensor.

6. Conclusion. All the carbon dioxide emissions data collected from the various locations are above the global carbon dioxide emissions average of 414.7 ppm, hence the need for policy consideration to significantly reduce carbon dioxide emissions within the selected industrial hubs. The average methane emissions were 0.1599 ppm (1599 ppb), 0.1358 ppm (1358 ppb) and 0.1366 ppm (1366 ppb) from location 1, 2 and 3 respectively slightly below the global average of 1895.7 ppb. The maximum recorded methane emissions was 0.5100 ppm (5100 ppb), 0.3700 ppm (3700 ppb) and 0.4800 ppm (4800 ppb) which are above the global methane emissions average of 1895.7 ppb.

The use of sim800I GSM module instead of Wi-Fi provided a reliable data transmission during the deployment and solves the data transmission problem faced by the Ghana Ambient Air Quality Project (GHAir). The use of a 50 W solar panel and 19.6 Ah battery pack provided a reliable electrical power for the monitor without power blackout/outages.The selected sensors Sensiron SCD-30, MQ-4, DHT22 were able to withstand the ambient condition within the selected industry locations within Suame Industrial Complex and monitored effectively the carbon dioxide, methane, temperature and humidity.

7. Acknowledgement. The authors are thankful to KNUST Engineering Education Project (KEEPS) for providing funds for carrying out the research work.

REFERENCES

[1] Abbas, F. N., *Capable of Gas Sensor MQ-135 to Monitor the Air Quality with Arduino uno*, International Journal of Engineering Research and Technology, 13(10), 2020, pp.2955–2959.

[2] Allen, D. T., *ScienceDirect Review Methane emissions from natural gas production and use : reconciling bottom-up and top-down measurements*, Current Opinion in chemical Engineering, 5, 2014, pp. 78-83.

[3] Amrago, D., Dachniewski, M., & Danquah, B., *APPRENTICESHIP: TRENDS EXPERIENCES AND EXPECTATIONS OF KUMASI SUAME MAGAZINE INFORMAL INDUSTRIAL ENCLAVE IN GHANA*, History of Suame Magazine in Kumasi. Journal of Management Challenges, 1(1), 2020.

[4] Asante, F. A., & Amuakwa-mensah, F. , *Climate Change and Variability in Ghana: Stocktaking*, Climate, 3, 2015, pp.78–99.

[5] Benefoh, D. T., & Amoah, A. , *Status of Monitoring Reporting Verification Scheme in Ghana (Issue November)*, 2017.

[6] Bhattacharjee, D. , *Design and Development of Wireless Sensor Node*, International Journal on Computer Science and Engineering, 02(07), 2010, pp.2431–2438.

[7] Bouwman, A. F., Hoek, K. W. Van Der, & Olivier, J. G. J. , *Uncertainties in the global source distribution of nitrous oxide inventories of N20*, Journal of Geophysical Research, 100(D2), 1995, pp.2785–2800.

[8] Clements, A. L., Griswold, W. G., Rs, A., Johnston, J. E., Herting, M. M., Thorson, J., Collier-oxandale, A., & Hannigan, M. , *Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary)*, Sensors, 17, 2017, pp.1–20.

[9] Dey, T., Ferrein, A., Frauenrath, T., Reke, M., & Schiffer, S. , *CO2 Meter: A do-it-yourself carbon dioxide measuring device for the classroom* The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA), 2021, pp.1–8.

[10] Gameli Hodoli, C., Coulon, F., & Mead, M. I. , *Applicability of factory calibrated optical particle counters for high-density air quality monitoring networks in Ghana*, Heliyon, 6(6), 2020, e04206.

[11] Geng, Y., Chen, W., Liu, Z., Chiu, A. S. F., Han, W., Liu, Z., Zhong, S., Qian, Y., You, W., & Cui, X. , *A bibliometric review: Energy consumption and greenhouse gas emissions in the residential sector*, Journal of Cleaner Production, 159(800), 2017, pp.301–316.

[12] Gładyszewska-fiedoruk, K. , *Correlations of air humidity and carbon dioxide concentration in the kindergarten*, Energy & Buildings, 62, 2013, pp.45–50.

[13] Guo, M., Wang, X., Li, J., Yi, K., Zhong, G., & Tani, H. , *Assessment of Global Carbon Dioxide Concentration Using MODIS and GOSAT Data*, Sensors, 12(16368 - 16389), 2012, pp.16368–16389.

[14] Jaarsma, T., Maat, H., Richards, P., & Wals, A. , *The role of materiality in apprenticeships: The case of the Suame Magazine, Kumasi, Ghana*, Journal of Vocational Education and Training, 63(3), 2011, pp.439–449.

[15] Jablonska, M., & Palkovits, R. , *Catalysis Science & Technology*, Catalysis Science and Technology, 2016, pp.1–17.

[16] Jagatha, J. V., Klausnitzer, A., Chac, M., & Laquai, B. , *Calibration Method for Particulate Matter Low-Cost Sensors Used in Ambient Air Quality Monitoring and Research*, Sensors, 21(3960), 2021, pp.1–27.

[17] Karagulian, F., Barbiere, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., Redon, N., Crunaire, S., & Borowiak, A. , *Review of the Performance of Low-Cost Sensors for Air Quality Monitoring*, Atmosphere, 10(506), 2019, pp.1–41.

[18] Karakurt, I., Aydin, G., & Aydiner, K. , *Sources and mitigation of methane emissions by sectors: A critical review*, Renewable Energy, 39(1), 2012, pp.40–48.

[19] Kemp, A. H., & Kemp, S. A. H. , *Real Time Wireless Sensor Network based Indoor Air Quality Monitoring System*, IFAC PapersOnLine, 52(24), 2019, pp.324–327.

[20] Kodom, K., Preko, K., & Boamah, D. , *X-ray fluorescence (XRF) analysis of soil heavy metal pollution from an industrial area in Kumasi, Ghana.*, Soil and Sediment Contamination: An International Journal, 0383. 2012,

[21] Mane, S. A., Nadargi, D. Y., Nadargi, J. D., Aldossary, O. M., Tamboli, M. S., & Dhulap, V. P. , *Design , Development and Validation of a Portable Gas Sensor Module: A facile Approach for Monitoring Greenhouse Gases*, Coatings, 10(1148), 2020, pp.1–11.

[22] Mosier, A. R., Halvorson, A. D., & Reule, C. A. , *Net Global Warming Potential and Greenhouse Gas Intensity in Irrigated Cropping Systems in Northeastern Colorado*, Journal Environmental Qualoity, 35, 2006, pp.1584–1598.

[23] NOAA. , *Increase in atmospheric methane set another record during 2021*, National Oceanic and Atmospheric Administration; US Department of Commerce. 2022.

[24] Özden, Ö., Dö, T., & Kara, S. , *Assessment of ambient air quality in Eski ş ehir , Turkey*, Environment International, 34, 2008, pp.678–687.

[25] Sewor, C., A. Obeng, A., & Amegah, K. , *Commentary: The Ghana Urban Air Quality Project (GHAir): Bridging air pollution data gaps in Ghana*, Clean Air Journal, 31(1), 2021, pp.1–2.

[26] Shi, R., & Conrad, S. A. , *Correlation and regression analysis*, Annals of Allergy, Asthma and Immunology, 103(4), 2009, pp.S35–S41.

[27] Srivastava, R. K., Sarkar, S., & Beig, G. , *Correlation of Various Gaseous Pollutants with Meteorological Parameter (Temperature, Relative Humidity and Rainfall)*, GLOBAL JOURNAL OF SCIENCE FRONTIER RESEARCH: H ENVIRONMENT & EARTH SCIENCE, 14(6), 2014, pp.57–65.

[28] Wuebbles, D. J. , *Nitrous Oxide: No Laughing Matter*, Science, 56(2009), 2014, pp.56–57. https://doi.org/10.1126/science.1179571

[29] Yeebo, Y. , *Suame Magazine Photo Essay: Suame Magazine*, Humanity: An International Journal of Human Rights, Humanitarianism, and Development, 7(1), 2016, pp.111–121.

[30] YuLiang, L., Xiang, N., YuanXi, W., & Wei, Z. , *Study on effect of temperature and humidity on the CO 2 concentration measurement Study on effect of temperature and humidity on the CO 2 concentration measurement*, IOP Conference Series: Earth and Environmental Science PAPER, 81, 2017, pp.1–6.

[31] Zhang, H., Zhang, J., Wang, R., Huang, Y., Zhang, M., Shang, X., & Gao, C. , *Smart carbon monitoring platform under IoT-Cloud architecture for small cities in B5G*, Wireless Networks, 2, 2021, pp.1–17.

[32] Jeffry, L., Yin, M., Nomanbhay, S., Mofijur, M., Mubashir, M., & Loke, P., *Greenhouse gases utilization : A review*,

Fuel, 301, 2021.
[33] Herzog, T., *World Greenhouse Gas Emissions*, 2009.
[34] Kumar, A., Singh, P., Raizada, P., & Mustansar, C., *Science of the Total Environment Impact of COVID-19 on greenhouse gases emissions : A critical review*, Science of the Total Environment, 806, 2022.
[35] Boesch, H., Liu, Y., Tamminen, J., Yang, D., Palmer, P. I., Lindqvist, H., Cai, Z., Che, K., Noia, A. Di, Feng, L., Hakkarainen, J., Ialongo, I., Kalaitzi, N., Karppinen, T., Kivi, R., Kivimäki, E., Parker, R. J., Preval, S., Wang, J., … Chen, H., *Monitoring Greenhouse Gases from Space*, Sensors, 13, 2021.
[36] Lamb, W. F., Wiedmann, T., Tubiello, F. N., Salvatore, M., Lamb, W. F., Wiedmann, T., Pongratz, J., Andrew, R., Crippa, M., Olivier, J. G. J., Wiedenhofer, D., Mattioli, G., Khourdajie, A. Al, & House, J., *A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018*, Environmental Research Letters, 16, 2018.
[37] Ahmad, M., Beddu, S., binti Itam, Z., & Alanimi, F. B. I. , *State of the art compendium of macro and micro energies*, Advances in Science and Technology Research Journal, 13, 2019.
[38] Flerlage, H., Velders, G. J. M., & Boer, J. De., *Chemosphere A review of bottom-up and top-down emission estimates of hydrofluorocarbons (HFCs) in different parts of the world*, Chemosphere, 283, 2021.
[39] Creutzig, F., Popp, A., Plevin, R., Luderer, G., Minx, J., & Edenhofer, O., *Reconciling top-down and bottom-up modelling on future bioenergy deployment*, Nature Climate Change, 2(5), 2021, pp. 320-327.
[40] Friedrich, E., & Trois, C., *Quantification of greenhouse gas emissions from waste management processes for municipalities – A comparative review focusing on Africa*, Waste Management, 31(7), 2011, pp. 1585–1596.
[41] Yodi, I. W. K. S., & Afifah, A. S., *Estimation of Green House Gas (GHG) emission at Telaga Punggur landfill using triangular, LandGEM, and IPCC methods Estimation of Green House Gas (GHG) emission at Telaga Punggur landfill using triangular, LandGEM, and IPCC methods*, Journal of Physics, 1452, 2020, pp. 1-8.
[42] Morrison, S. R. , *Selectivity in semiconductor gas sensors*, Sensors and Actuators, 12, 1987, pp. 425-440.
[43] Sberveglieri, G. , *Recent developments in semiconducting thin-film gas sensors*, Sensors and Actuators, 23, 1995, pp. 103-109.
[44] Sakakura, T., Choi, J., & Yasuda, H., *Transformation of Carbon Dioxide*, Chemical Reviews, 107(6), 2007, pp. 2365–2387.
[45] Crabtree, R. H., *Aspects of Methane Chemistry*, Chemical Reviews, 1994, pp. 987-1007.
[46] Koestoer, R. A., Pancasaputra, N., & Roihan, I., *A sirelativemple calibration methods of humidity sensor DHT22 for tropical climates based on Arduino data acquisition system*, AIP Conference Proeedings, 2019, pp. 1-8.
[47] Mihai, B. , *How to use the DHT22 sensor for measuring temperature and humidity with the Arduino Board*, DE GRUYTER OPEN, 2016, pp. 22-25.
[48] Adhiwibowo, W., Daru, A. F., & Hirzan, A. M., *Temperature and Humidity Monitoring Using DHT22 Sensor and Cayenne API*, TRANSFORMTIKA, 17(2), 2020, pp. 209-214.

Edited by: Vinoth Kumar
Received: Aug 16, 2022
Accepted: Nov 29, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 377–388, DOI 10.12694/scpe.v23i4.2048

ACUTE MYELOID LEUKEMIA MULTI-CLASSIFICATION USING ENHANCED FEW-SHOT LEARNING TECHNIQUE

KANNAN VENKATESH *, S. PASUPATHY †, AND S.P. RAJA ‡

Abstract. Acute Myeloid Leukemia (AML) is a form of the condition that is fatal and has a high mortality rate. It is characterised by abnormal cells growing rapidly inside the human body. The conventional method for detecting AML seems to be examining the blood sample manually under a microscope, which is a manual and cumbersome task that also requires well-trained medical expertise for efficient identification. On the other hand, considering medical diagnosis, the capacity to classify medical images faster and accurate is essential. The classification of medical images my currently be accomplished using a range of methodologies including Machine Learning (ML), Deep Learning (DL) and Transfer Learning (RL). While these approaches are effective for large datasets, they can take a while and not ideal for small datasets. In recent years, advances in Deep Convolutional Neural Networks (DCNN) have made it possible and produce more accurate and promising outcome while processing a medical image. However, the paradigm that DCNN use for training includes a large number of annotations in order to prevent overfitting and produce promising results. Obtaining large-scale semantic annotations in clinical operations might be problematic in some cases, particularly biological expertise knowledge is needed. It is also regular occurrence in scenarios where only a small number of annotated classes are accessible in some circumstances. At this context, in order overcome the drawback of traditional approach a framework has been developed which comprises of Enhanced Few-Shot Learning Technique integrated Base Classifier (Feature Encoder)-EFLTBC. The proposed model has built using base classifier and meta-learning block, and it optimized the better results. To diagnose AML, the doctor must count the number of white blood cells and red blood cells and see if there are any abnormal health conditions in that using a microscope. However, obtaining an accurate result takes time and effort. To address these issues, the proposed Novel AML detection model employing is used in this study. Base classifier utilizing ResNet-18 pretrained model and meta learning block has computed using the average feature of every samples. Also, the dataset that we used consisting of three classes includes Normal monocytes, Abnormal monocytes, Lymphocyte and Experimental results outperform various existing deep learning technique with the accuracy of 97%, recall of 96.55% F1-Score of 96.65% and precision of 96.60.

Key words: Transfer Learning, Deep Convolutional Neural Networks, Enhanced Few-Shot Learning Technique integrated Base Classifier, Meta learning, base classifier.

AMS subject classifications. 68T05

1. Introduction.

One of the hallmarks of acute myeloid leukemia (AML) is the proliferation of immature progenitors inside the bone marrow, which ultimately leads to a lack of hematopoietic function [1]. While invasion of other organs including the brain and the lungs is rare and observed only in instances with elevated blast numbers in the blood, peripheral blood involvement is perhaps the most frequent tissue affected. At minimum 20% of the myeloblasts in the bone marrow (or plasma) must have myeloid ancestry to qualify as AML, as per the World Health Organization (WHO). As an alternative to greater than 20% threshold, AML can be diagnosed without reference to blast percent in occurrences with Nucleophosmin 1-mutated AML and Acute Promyelocytic Leukemia (APL). Acute undifferentiated Leukemia (AUL) is a kind of Leukemia which has more than 20% blasts absent markers and is often handled as AML [2].

For forecasting survival rates (SR) as well as event-free survival, research have shown that both hereditary and clinical features which are significant. There are three genetic mutations that have been linked to poor outlook and reduced survival rates: RUNX1 (ASXL1), TP53 (ASXL1), and ASXL1. The combination of a TP53 mutant and a complicated karyotype leads in the worst prognosis [3,4]. Typically, it has already been hypothesized that 75 percent of variances are due to genomic instability, and the remaining 25% are due to

*Department of Computer Science and Engineering, Annamalai University, Chidambaram,608002, India (venkiur@gmail.com).
†Department of Computer Science and Engineering, Annamalai University, Chidambaram,608002, India (pathyannamalai@gmail.com)
‡School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, 632014, India (avemariaraja@gmail.com)

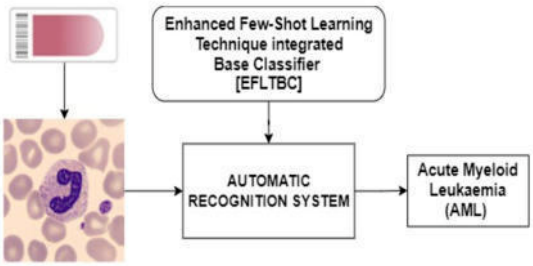


Fig. 1.1: General workflow of Research

clinical, therapeutic, and behavioral characteristics. In the 37 models that have been tested thus far, remission or life span could only be predicted in 76% to 81% of instances [5]. In light of this, it is imperative that we find more accurate prognostic factors.

Forecasting of clinical outcome using machine learning (ML) approaches has grown increasingly common. This method can be used to forecast the prognosis of cancer, for instance [6]. An artificial intelligence technique known as "machine learning" fo- cuses on developing computer algorithms that can be updated as new data comes to light. Because of this, models built from earlier data can be utilized for diagnosis and detection. In this context, Deep learning models also a powerful and robust algorithm that are effectively applied in important research areas like medical im- age analysis, data processing, financial modelling, and spam detection [7, 8]. Deep learning techniques that specialize in visual detection and segmentation include con- volutional neural networks (CNNs) [9]. There are seven different layers in LeNet-5, comprising maximum pooling, convolutional, and fully connected (FC) [10], which is based on CNN. A CNN-based system known as AlexNet employs eight layers, combin- ing maximum pooling and convolution neural networks [11,12]. In the Massive Visual Recognition Problem, AlexNet was stated the best artificial neural model for image classification. Many studies have suggested CNN-based algorithms for the diagnosis of Leukemiain this context, [13-16] focused on the identification of Acute Lymphatic Leukemia (ALL). There were two studies that focused on the detection of Acute Myeloid Leukemia (AML): P. Tiwari [17] and S. Rodrigues et al [18]. According to this study, there has been very little research done on the automatic identification of Acute Myeloid Leukemia (AML). The well deep learning method, AlexNet, can be applied in a variety of fields, including medical image analysis. Medical images including such CT scans, X-rays, MRI, PET, ultrasonic, and hematologic images can be analyzed and detected Alex Net.

2. Background work. Cells that protect the body from illness, infectious diseases, and protozoans, such as bacteria, viruses, and parasites are known as white blood cells (WBCs). In the bone marrow, hematopoietic stem cells produce WBCs. In general, there are five major types of WBC's exists includes neutrophils, monocytes, eosinophils, basophils and lymphocytes. It various one another depends on their physical and functional features. All these has been explained simply as given as follow:

Lymphocytes: It helps to develop antibodies in order to prevent from numerous microorganisms includes bacteria, viruses and different harmful invaders. Also, B cells, T cells and natural killer cells which are its further subcategory.

Monocytes: Bacteria can be broken down more easily because of its longer lifespan. The core purpose of the monocyte cell is to provide a piece of microbe to the T cell so that it can recognise it again in potential attacks.

Eosinophils: Immune responses such as hives and hay fever as well as diseases of the nervous and collagenous systems, the spleen, parasitic inflammation, and other types of allergies are all boosted by it. Nevertheless, parasites like tapeworms and hookworms are their fundamental prey.

Basophils: Antigen and allergic reaction are largely mediated by basophils, which release histamine (a sign of allergy) to dilate blood vessels and aid the immune system.

Neutrophils: Whenever there is a threat of infection, this is the first line of defence. Fungus and bacteria are

the primary targets of their attack. This cell type is commonly found in the early levels of inflammatory response, where its death creates pus in the body.

Types of Leukemia: Leukemia is a haematological illness and form of blood cancer that impairs the immune system of human by producing an excess of malignant WBCs (White Blood Cells) [19]. Acute Myeloblastic (Myeloid) Leukemia (AML), Chronic Myeloblastic (Myeloid) Leukemia (CML), Acute Lymphoblastic (Lymphocytic) Leukemia (ALL), and Chronic Lymphoblastic (Lymphocytic) Leukemia are the four major subtypes of Leukemia(CLL). The following is a brief overview of the four major types of Leukemia[20]: **AML**: This is the most prevalent type of leukaemia, affecting both kids and adults. It has an effect on the human body's WBCs, causing them to develop abnormally. Acute cancers are caused by an excess of unusual WBCs present over the human body.

ALL: Although it is the another very common form of cancer in children, it can affect adults as well. ALL is caused by a rise in lymphocytes (a type of white blood cell) in the human.

CML: It is a type of white blood cell cancer wherein immature white blood cells or blast cell lines form and multiply uncontrolled way in the human body. Adults are primarily affected by CML.

CLL: It is a specific type of cancer that attacks B lymphocytes or B cells. B cells assist the human body in fighting infection; however, cancerous B cells are incapable of fighting infection. CLL is typically found in older adults.

3. Related work and Literature Survey. E. Tuba et al. [20] presented a com- prehensive survey of WBC categorization methods for analyzing fatal diseases includ- ing blood cancer, AIDS, and Leukemia by examining microscopic images of haema- tology reports. Additionally, they highlighted several critical challenges associated with the WBC classification stage, including WBC structure, quality of image, cell differentiation, and time consumption of classification process [21]. J. Su, et al. classified WBCs utilizing neural network models with forward and backward propa- gation. After analyzing two complicated blood cells obtained via extendable images, the neural network was fed the 16 another very crucial components of that cell as input. After segmentation, training has been given with neural network consisting of 50% data and used the remaining data for testing. They classified WBCs with a 96% accuracy. The papers [22-25] noted that identifying WBCs is challenging caused by uneven coloration and anomalous illumination of clinical specimens. They presented a system for WBC classification that incorporates histogram dispersion, K-means clustering, a watershed method, and CNN. They noted a 95.81% accuracy rate for their suggested hybrid approach [27]. Umamaheswari and Geetha provided an overview of leukemia, its sub - types, and the various automated segmentation techniques used during machine learning to detect Leukemiain microscopic examination. They came to two significant conclusions at the conclusion of their reviews, the first among one is Hybrid techniques combining machine learning and image processing may improve Leukemia detection. And the second one is, A benchmark dataset is required to iden- tify periodic advancements to the proposed schemes [28]. Thanh et al. proposed a comprehensive model based on CNNs for the early screening of acute Leukemia disease. The experimental findings for the proposed approach include the first classification process, which represents an impressive result in terms of separating normal and ab- normal cells. Their suggested technique classified Leukemia cells with an accuracy of 96.6% [14]. The methodology based on deep segmentation of CNN for identifying Leukemia blood cells in microscopic examination. Their proposed models were capable of classifying WBCs in blood smear images with an accuracy of 93.94%.

It's suggested a blood cell microscopical classification algorithm based on a combination of CNN and RNN to address the heavily reliant link problem among significant features of an images. They concluded that, the combination of CNN and RNN model is more effective and reliable than other Convolutional neural models. The methodology for detecting blood cells relying on Double Convolutional Layer Deep Neural network and performance in comparison to that of SVM and Naive bayes Classifier models. They demonstrated that their envisaged DCLNN model outperforms SVM and Naive Bayes technique. The envisioned a Decision - Making framework (DMF) based on K-means clustering and panel preference to classify Leukemia cells in microscopic examination. They carried out experiments using a variety of publicly available and benchmark data sources and then validated their findings with expert physicians. They revealed that their proposed model got an accuracy of 99.517 percent [29]. AlexNet is a capable and quite well deep learning technique that is effective in a variety of research areas, most notably medical image processing. AlexNet employs eight distinct layers,

among which are optimum pooling, convolutional, and fully connected layers. AlexNet is capable of analysing and detecting significant features in a variety of medical im- ages, including CT scans, X-rays, MRIs, PET scans, ultrasounds, and hematological images.

This study offers a brand-new approach to categorizing acute myeloid leukemia. As a result of the fact that medical images are by their very nature noisier and more diverse than the original convolutional frame, a convolutional block extracts basic feature knowledge and concentrates on the entire image. [30] With the use of a 1x1 convolutional block, we were able to extract spatial and channel properties from AML blood smears. These changes allow us to give more weight to data that contributes to classification. There are two parts to our strategy. The first is a pre-trained ResNet-18 network enhanced with an additional layer and a meta-learning component that quantifies the averaged feature of every sample before categorizing query set features based on the closest centroid and co- sine metric distance between both. We hope to better grasp the difference between whole-classification and meta-learning by isolating the inconsistencies. After training a classifier just on basis classes, we replace the final FC layer with a 1x1 convolutional layer that is reliant on the base classes. [31] When a new class is being tested, the centroids are obtained by determining the mean immersion of the support samples, and then query samples are assigned to the centroid with nearest cosine distance.

4. Related work and Literature Survey. This research presents a novel method to the classification of Acute Myeloid Leukemia. By nature, medical images are noisier and much more diversified than the original convolutional frame, therefore a convolutional block extracts basic feature knowledge and focuses on the complete image. With the use of a 1x1 convolutional block, we were able to extract spatial and channel properties from AML blood smears. These changes allow us to give more weight to data that contributes to classification. There are two parts to our strategy. The first is a pre-trained ResNet-18 network augmented with an extra layer and a meta-learning component that quantifies the averaged feature of every sample, then classifies query set features are based on the closest centroid and cosine metric distance between both. We hope to better grasp the difference between whole-classification and meta-learning by isolating the inconsistencies. After training a classifier just on basis classes, we replace the final FC layer with a 1x1 convolutional layer that is reliant on the base classes. When a new class is being tested, the centroids are obtained by determining the mean immersion of the support samples, and then query samples are assigned to the centroid with nearest cosine distance. The few-shot learning paradigm we've presented is shown in Figure 1.1.

A. Base Classifier

Here pre-trained ResNet-18 architecture is used as a base classifier. A cosine nearest centroid method is used in few-shot learning technique which helps to training the classifier along with losses has been associated with respect to base classes. In order to obtaining an encoder 'g', fully connected layer is removed from the pretrained model also 1×1 convolutional layer is integrated on top of the pre-trained model. The responsibility of encoder 'g' is to map every input with respect to its embedding which helps to provide training the classifier corresponding to all the base classes along with traditional cross-entropy loss. Let consider the supported set for the few-shot task 'S'. For every class present over the dataset is represented as 't' and the corresponding few-shot samples over the class is taken as 'S_t'. Base classifier determines the average embedding 'X_t' which is known as class centroid and the equation is given as follow:

$$W_t = \frac{1}{|S_t|} \sum_{b \in S_t} g(b) \tag{4.1}$$

In order a provide a sample of query 'b' is used in a few-shot task. It helps to determine the probability of samples present over the class 't' with respect to the cosine similarity along with centroid.

$$p\left(B = t \mid b\right) = \frac{\exp(\langle g\left(b\right), W_t \rangle)}{\sum_{t'} \exp(\langle g\left(b\right), W_{t'} \rangle)} \tag{4.2}$$

The cosine similarity of two vector values in the blood smear image is represented by the value $\langle g\left(b\right), W_t \rangle$. Equation (3) states that cosine similarity is obtained as

$$\langle g\left(b\right), W_t \rangle = \frac{g\left(b\right).W_t}{\|g(b)\| * \|W_t\|} \tag{4.3}$$

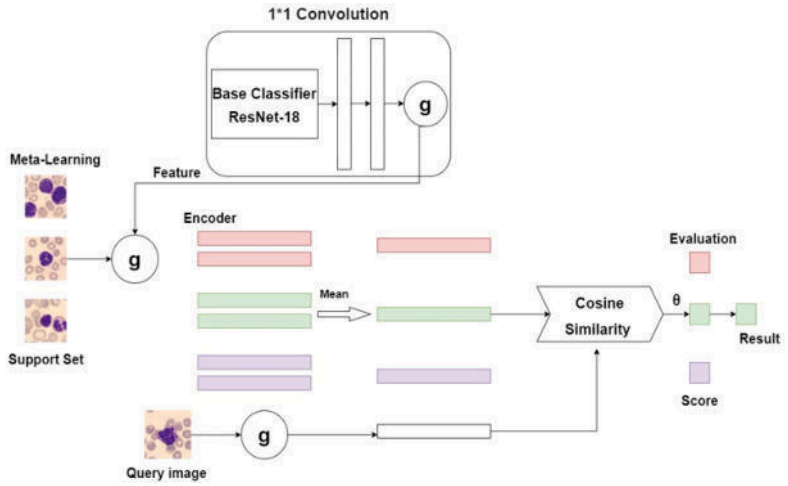


Fig. 4.1: Proposed Workflow

B.Meta-Learning

The Meta-Learning is represented in Figure 4.1. To begin, the classification training stage requires training a base classifier, which entails training a classifier on all possible base classes by removing its last FC layer and replacing it with a 1x1 convolutional layer to achieve the final classifier g [32-33]. The meta-learning stage is the next step in optimizing the model for the pre-trained base classifier evaluation measure. Classification-trained feature encoder f is used to sample from training samples in base classes in order to carry out N-way K-shot tasks (with NxQ query samples). The projected probability distribution for each sample in the query set given in Equation (2) is then used to determine the loss for each task in order to compute the loss for each of the N classes established in Equation(1) using the centroids. When p is used in conjunction with the labels assigned to the samples in the query set, this loss is calculated as a cross-entropy loss. During training, each training batch can contain a number of tasks, with the average loss being calculated for each task.

The confidence interval for cosine similarity is [1, 1], hence when determining logits, it may be useful to normalize the value beforehand employing the SoftMax function throughout training. The cosine similarity and scalar are multiplied by the learnable scalar 'θ' in training to get the prediction of probability.

$$p\left(B = t \mid b\right) = \frac{\exp(\theta . \langle g\left(b\right), W_t \rangle)}{\sum_{t'} \exp(\theta . \langle g\left(b\right), W_{t'} \rangle)} \tag{4.4}$$

To find out if the meta-learning aim is still advantageous when implemented to a comprehensive classification model of classification in Acute Myeloid Leukemia classification, this study sets out to investigate. Using the support set 'S' as input, a few-shot task evaluates the average features for samples of every class and then uses cosine similarity as a proximity metric to classify a specimen in the query set. An additional learnable scalar is provided to adjust the cosine similarity of a basic classifier using Meta-Learning, further optimizing the convergence pre-trained base classifier. To better classify AML, we now have the ability to take into account their spatial and communication channels.

5. Experimentation and Result Discussion.

5.1. Dataset. This study analyzed two distinct datasets, all of which were derived from publicly accessible databases. The first dataset included in this research comprised of 4,000 blood smear acute myeloid test results obtained from two distinct sources which is available in repository [34][1]. Among the dataset, 1000 of which

[1]https://data.mendeley.com/v1/datasets/snkd93bnjr/1

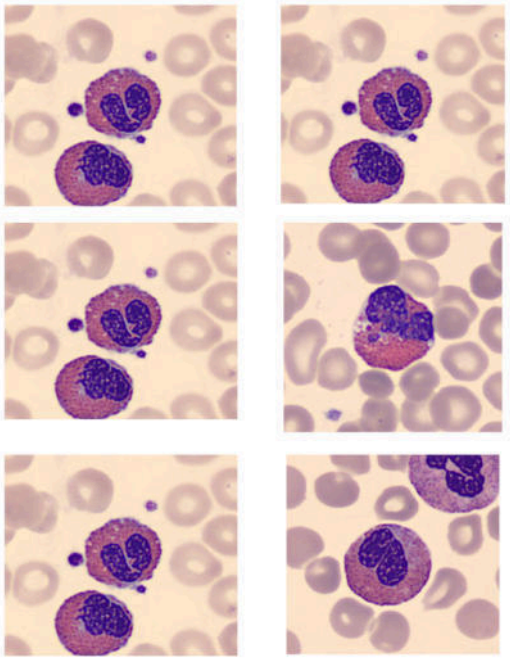


Fig. 5.1: Example of single-cell images taken from peripheral blood smears (includes all the classes).

Table 5.1: Properties of Dataset-1

Properties of dataset	**Respective values**
Image size	960 ×1080 pixel
Nature of image	RGB
Entire classes	3
Lymphocyte	1000
Normal monocytes	1600
Abnormal monocytes	1400

contains lymphocytes, 1600 of which contains normal monocytes and 1400 contains abnormal monocytes. The data has been pre-classified by subject-matter experts. And the second dataset that we have used in our research work which is accesses from publicly available repository. The total dataset consisting of 18,384 images. And the entire images holds both AML infected patients and healthy individuals. Totally Eight classes which are existing in this dataset which includes neutrophil, promyelocyte, metamyelocyte, eosinophil, basophil, erythroblast, smudge cell, lymphocytes. Each class consisting of different number of images. For instance, class neutrophil comprises maximum of 8,583 images and class lymphocytes comprises of only twenty images. The dataset which already reclassified by subject-matter expertise which is publicly available in the repository [35][2].

The datastore is depicted in Fig. 5.1. The dataset's properties are listed in Table 5.1 and Table 5.2 accordingly.

5.2. Implementation. We use ResNet-18 as the grounds classifier, which become trained on the ImageNet database. For training phase of classification, we use the SGD optimizer with such a learning rate of 0.1, a

[2]https://wiki.cancerimagingarchive.net/plugins/servlet/mobile?contentId=61080958#content/view/61080958

Table 5.2: Properties of Dataset-2

Properties of dataset	**Respective values**
Image size	960 ×1080 pixel
Nature of image	RGB
Entire classes	8
neutrophil	8,583
promyelocyte	1630
metamyelocyte	1443
eosinophil	2305
basophil	1440
erythroblast	2326
smudge cell	657
lymphocytes	20

momentum of 0.9, and a decay factor of 0.1. ResNet-18 has a weight decay of 0.0001. For the meta-learning step, we use the SGD optimizer with a momentum of 0.9. We fixed the rate of learning throughout the classification process is 0.001. Also, every training batch includes a small quantity of few shot tasks for the purpose of calculating the average loss. The cosine scaling parameter is initially set to a value of 10. It is crucial to evaluate performance of classification in image classification interrogations in order acquire scientific evidence for the study's findings. Otherwise, the classification study might be abandoned in the classroom, leaving it inadequate.

It is common practice in image classification research to use a variety of performance evaluation criteria that are widely used in other similar studies. There are a number of them, including accuracy, precision, specificity, and sensitivity. Typical performance criteria in image classification investigation are employed in this study to examine the classification reliability and accuracy. It is also used to evaluate models' performance by looking at the AUC of the receiver operation characteristic curve (AUC of ROC). The formulas and related numbers for each of these metrics are provided below.

$$Accuracy = \frac{True\ positive + True\ Negative}{True\ positive + True\ Negative + False\ Positive + False\ Negative}$$

$$Specifity = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

It is typical practice in few-shot learning to employ the episode training strategy. We use a two-way 25-shot technique for classification-1, using 25 query images for every class in training. The training set is divided into three categories, from which we draw a random sample of twenty images from each. A five-way, 25-shot approach with 25 training images for each class is used to complete the classification-2 task. The training set is divided into nine categories, from which we draw a random sample of twenty images for each.

Table 5.3 depicts the performance of our proposed work. The performance evaluation measures that we used includes accuracy, specificity, precision and sensitivity. Table 5.3 exactly point out that our suggested few shots techniques employ good performance for all the two datasets. In classification task-1, Normal monocytes achieved a 100% accuracy.

Table 5.3: Performance of our Proposed technique

Dataset	**Classes**	**Accuracy (%)**	**Specificity**	**Sensitivity**	**Precision**
Classification task-1	Lymphocytes	99.07632	0.982111	0.979462	0.974312
	Normal monocytes	100	1	1	1
	Abnormal monocytes	97.99993	0.984378	0.963458	0.962361
Classification task-1	Neutrophil	98.0578	0.98635	0.972674	0.973124
	Promyelocyte	97.7366	0.98238	0.931233	0.9165
	Metamyelocyte	97.12548	0.991453	0.923453	0.962375
	Eosinophil	96.82245	0.963446	0.932221	0.887671
	Basophil	97.12492	0.983418	0.911818	0.93116
	Erythroblast	96.22348	0.981353	0.923453	0.962375
	Smudge cell	94.83245	0.973446	0.95341	0.897671
	Lymphocytes	97.14452	0.964418	0.904818	0.94316

Table 5.4: Result Comparison of our proposed model with various existing models

Models	**Classification Task-1**		**Classification Task-2**	
	Accuracy (%)	**AUC**	**Accuracy (%)**	**AUC**
AlexNet	90.13	0.8776	85.13	0.8331
Dual Path Network	88.91	0.8745	84.54	0.8469
ResNet-50	93.97	0.9348	77.23	0.8212
VGG-16	87.29	0.9356	89.67	0.8908
DenseNet169	87.23	0.8613	90.99	0.9122
Constellation Net	84.38	0.8575	92.33	0.9332
Proposed Model	**97.87**	**0.9696**	**94.76**	**0.9793**

We make a comparison between our approach and that of transfer learning. Models from AlexNet [33], VGG16, ResNet-50 and DenseNET169 [33] are modified with the final classification layer and then used. Furthermore, our proposed methodology outperformed the constellation Net [32], an approach based on a few shots. A summary of results from two datasets is shown in Table 5.4.

It is shown in Figures 5.2 and 5.3 that classification task-1 has an accuracy and loss value. The testing set's loss is diminishing over time, as can be shown. Though there is a difference in training and test accuracy between epochs 20 and 40, it improves after 50. Results for classification task 2 are shown in Figures 5.4 and 5.5.

To train and evaluate our classifiers on the AML dataset, we start with the pretrained ResNet-18 as a foundation classifier. Examining Table 5.4, it is obvious that our proposed model is more accurate than other models. Comparing our approach to pre-trained models, the findings show that it has enormous benefits. In the case of pre-trained models, the ImageNet dataset is frequently used as the training dataset. However, the natural images on ImageNet differ from the medical ones, resulting in disappointing results on a regular basis. It was built on top of an already-trained model and retrained via meta-learning. Relatively poor outcomes are obtained when using the Dual Path Network, which combines ResNeXt and DenseNet. This may be due to the medical image's features and the embedding module used. The result of this is that our model performs better ever more densely connected networks.

Graphical representations of the accuracy comparison between two datasets are depicted in Figures 5.6 and

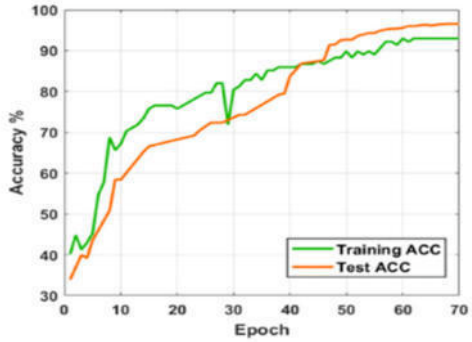


Fig. 5.2: Classification task-1 Accuracy

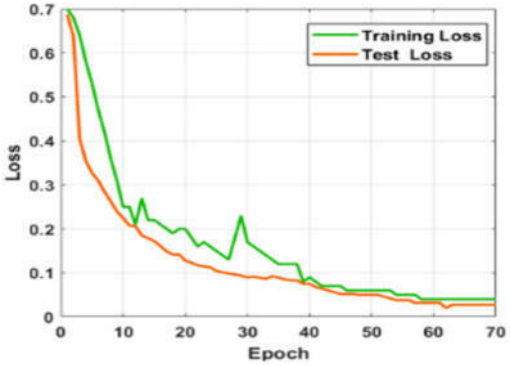


Fig. 5.3: Classification task-1 Loss

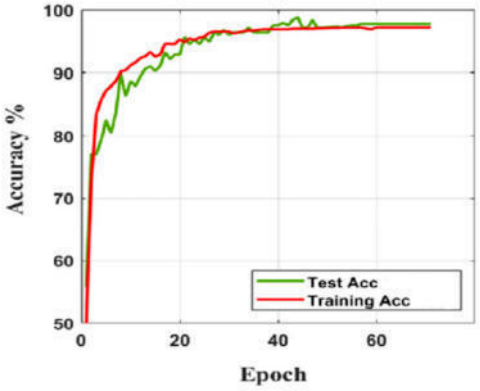


Fig. 5.4: Classification task-2 Accuracy

5.7. We can clearly see that our proposed model outperforms another model in this trial. In order to increase nonlinear features without sacrificing resolution, we use 1x1 convolution. It could also be used to alter the size

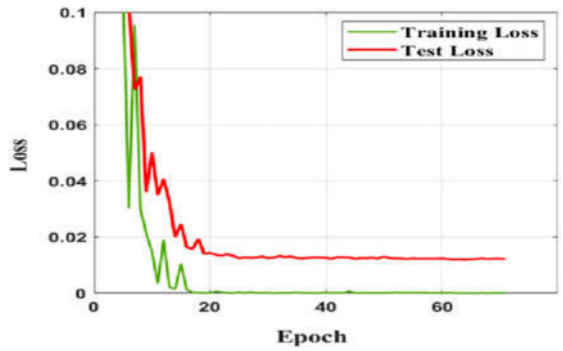


Fig. 5.5: Classification task-2 Loss

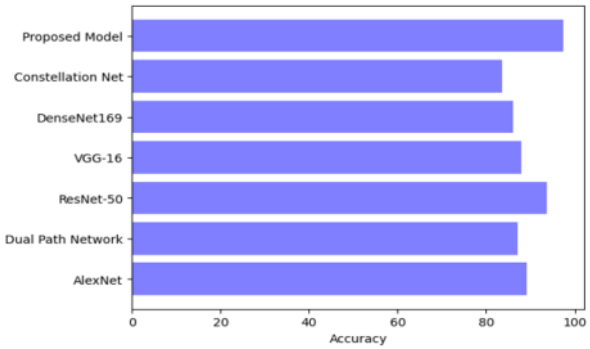


Fig. 5.6: Comparison results for Classification task-1

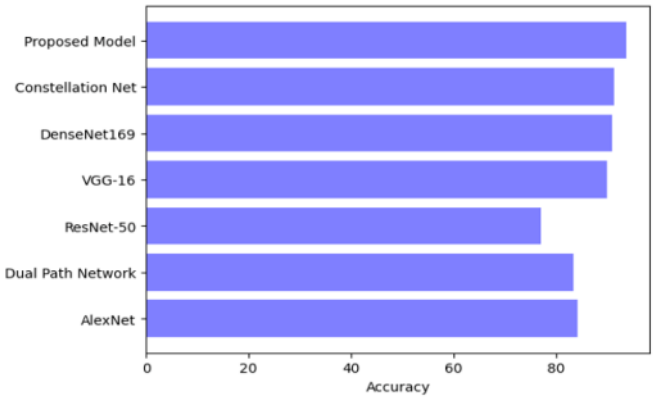


Fig. 5.7: Comparison results for Classification task-2

of an object, making it smaller or larger. During the feature extraction process, it makes it easier to take into account spatial and channel information.

6. Conclusion. When the white blood cells (WBCs) in the blood become can- cerous, it is known as acute myeloid leukemia (AML). We use the few-shot learning approach to solve the problem of classifying AML-impacted blood sample images. We proposed a framework based on a combination of pre-trained base classifiers and meta-learning. It is significantly easier and more convenient than other approaches. Incorporating a basic classification top on the pretrained ResNet-18 and 1x1 convolution kernel to the base module significantly improves the model's presen- tation potential. Our investigations are based on the AML dataset, which we use to compare two datasets. From the experimentation outcome, our proposed model achieves the classification accuracy of 97.87% and 94.76% in both classification tasks respectively. When building our model, we use the cosine similarity distance to esti- mate how far off the trained meta-learning model and the query image are from each other. This distance keeps spatial and channel information in blood smear samples intact. We will experiment with various distance metrics in order to conduct future research.

REFERENCES

[1] Pollyea D.A., Kohrt H.E., and Medeiros B.C., Acute myeloid Leukemiain the elderly: a review. British journal of haematology, 2011, p. 524–542.

[2] Estey E.H., Acute myeloid leukemia: 2019 update on risk-stratification and management. American Journal of Hematology, 2018, p. 1267–1291.

[3] Arber D.A., et al., The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood, 2016, p. 2391–2405.

[4] Do ner H., et al., Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood, 2017, p. 424–447.

[5] Kourou K., et al., Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 2015. 13: p. 8–17.

[6] Omondiagbe, D.A., S. Veeramani, and A.S. Sidhu. Machine Learning Classification Techniques for Breast Cancer Diagnosis. in IOP Conference Series: Materials Science and Engineering. 2019. IOP Publishing.

[7] E. Nazari, A. H. Farzin, M. Aghemiri, A. Avan, M. Tara, and H. Tabesh, "Deep learning for acute myeloid leukemia diagnosis," Journal of Medicine and Life, vol. 13, pp. 382–387, 2020.

[8] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov, "Applications of deep learning in biomedicine," Molecular Pharmaceutics, vol. 13, no. 5, pp. 1445–1454, 2016.

[9] P. Bharati and A. Pramanik, "Deep learning techniques-RCNN to mask R-CNN: a survey," in Proceedings of the Computational Intelligence in Pattern Recognition, pp. 657– 668, Springer, Xiamen, China, May 2020.

[10] C.-W. Zhang, M.-Y. Yang, H.-J. Zeng, and J.-P. Wen, "Pedestrian detection based on improved LeNet-5 convolutional neural network," Journal of Algorithms & Computational Technology, vol. 13, 2019.

[11] M. Z. Alom, T. M. Taha, C. Yakopcic et al., "The history began from alexnet: a comprehensive survey on deep learning approaches," 2018.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[13] N. Kumar, N. Kaur, and D. Gupta, "Major convolutional neural networks in image classification: a survey," in Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), pp. 243–258, Chandigarh, India, June 2020.

[14] T. T. P. Thanh, C. Vununu, C. Vununu, S. Atoev, and S.-H. Lee, "Blood cell image classification using convolutional neural network," International Journal of Computer eory and Engineering, vol. 10, no. 2, pp. 54–58, 2018.

[15] S. L. a. S. Daware, "Analysis of white blood cell segmentation techniques and classification using deep convolutional neural network for leukemia detection," HELIX, vol. 8, pp. 4519– 4524, 2018.

[16] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining convolutional neural network with recursive neural network for blood cell image classification," IEEE Access, vol. 6, pp. 36188–36197, 2018.

[17] P. Tiwari, J. Qian, Q. Li et al., "Detection of subtype blood cells using deep learning," Cognitive Systems Research, vol. 52, pp. 1036–1044, 2018.

[18] S. Rodrigues, B. Majhi, and P. K. Sa, "Texture feature-based classification on microscopic blood smear for acute lymphoblastic leukemia detection," Biomedical Signal Processing and Control, vol. 47, pp. 303–311, 2019.

[19] S. Mishra, B. Majhi, P. K. Sa, and L. Sharma, "Gray level cooccurrence matrix and random forest based acute lymphoblastic leukemia detection," Biomedical Signal Processing and Control, vol. 33, pp. 272–280, 2017.

[20] E. Tuba, I. Strumberger, N. Bacanin, D. Zivkovic, and M. Tuba, "Acute lymphoblastic leukemia cell detection in microscopic digital images based on shape and texture features," in Proceedings of the International Conference on Swarm Intelligence, pp. 142–151, 2019.

[21] J. Su, S. Liu, and J. Song, "A segmentation method based on HMRF for the aided diagnosis of acute myeloid leukemia," Computer Methods and Programs in Biomedicine, vol. 152, pp. 115–123, 2017.

[22] S. Agaian, M. Madhukar, and A. T. Chronopoulos, "Automated screening system for acute myelogenous leukemia detection in blood microscopic images," IEEE Systems Journal, vol. 8, no. 3, pp. 995–1004, 2014.

[23] H. Kutlu, E. Avci, and F. Ozyurt, "White blood cells detection and classification based on regional convolutional neural networks," Medical Hypotheses, vol. 135, Article ID 109472, 2020.

[24] V. Lights, Leukemia, 2019, https://www.healthline.com/ health/leukemia.
[25] K. AL-Dulaimi, J. Banks, V. Chandran, I. Tomeo-Reyes, and K. Nguyen Thanh, "Classification of white blood cell types from microscope images: techniques and challenges," Microscopy Science: Last Approaches on Educational Programs and Applied Research, vol. 8, pp. 1–10, 2018.
[26] M. Z. Othman, T. S. Mohammed, and A. B. Ali, "Neural network classification of white blood cell using microscopic images," International Journal of Advanced Computer Science and Applications, vol. 8, pp. 99–104, 2017.
[27] L. Lin, W. Wang, and B. Chen, "Leukocyte recognition with convolutional neural network," Journal of Algorithms & Computational Technology, vol. 13, 2018.
[28] D. Umamaheswari and S. Geetha, "Review on image segmentation techniques incorporated with machine learning in the scrutinization of leukemic microscopic stained blood smear images," in Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering, pp. 1773–1791, 2018.
[29] A. S. Negm, O. A. Hassan, and A. H. Kandil, "A decision support system for acute Leukemiaclassification based on digital microscopic images," Alexandria Engineering Journal, vol. 57, no. 4, pp. 2319–2332, 2018.
[30] S. Q. Salih, H. K. Abdulla, Z. S. Ahmed, N. M. S. Surameery, and R. D. Rashid, "Modified alexnet convolution neural network for covid-19 detection using chest x-ray images," Kurdistan Journal of Applied Research, vol. 10, pp. 119–130, 2020.
[31] S. Lu, S.-H. Wang, and Y.-D. Zhang, "Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm," Neural Computing and Applications, vol. 99, pp. 1–13, 2020.
[32] Weijian Xu, yifan xu, Huaijin Wang, and Zhuowen Tu. Constellation nets for few-shot learning. In International Conference on Learning Representations, 2021.
[33] Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., Feng, J.: Dual path networks. In: Guyon, I., et al. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017, pp. 4467–4475, 2017.

Edited by: Vinoth Kumar
Received: Aug 17, 2022
Accepted: Nov 30, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 389–402, DOI 10.12694/scpe.v23i4.2049

A MULTI OBJECTIVE HYBRID COLLISION-FREE OPTIMAL PATH FINDER FOR AUTONOMOUS ROBOTS IN KNOWN STATIC ENVIRONMENTS

KADARI NEERAJA* AND GUGULOTHU NARSIMHA†

Abstract. The most important field of robotics study is path planning. Path planning problem in general is an NP-complete problem. Though several attempts have been made using A*, PRM, RRT, and RRT* these algorithms explore too many nodes in the state space, not completely captured kinematic constraints, and are not optimal in real-time. In this paper, a Multi-Objective Hybrid Collision- free Optimal Path Finder (MOHC-OPF) is proposed which is an attempt to obtain a near-optimal solution by exploring fewer nodes compare to the above existing methods while considering kinematic constraints aiming to generate optimal drivable paths. The empirical study revealed that the proposed algorithm is capable of detecting static obstacles and finding a collision-free nearest-optimal, smooth and safe path to the destination in a static known environment. Multiple criteria, including path length, collision-free, execution time, and smooth path, are used to determine an optimal path.. The proposed algorithm shows efficiency in finding the shortest path length and execution time decreased in 90% of the experiments.

Key words: Path Planning, A^* algorithm, Optimal Path, Collision-Free, Kinematic Constraints.

AMS subject classifications. 68U35, 65Y04, 05C38

1. Introduction.

Path planning is the ascertainment of a collision-free path in any given environment, which may often be bestrewed in the real world [1]. Sometimes it is termed as motion planning because it helps to decide on the motion of any object within the working world. An object could be a robot which is autonomous because it utilizes the path-finding algorithm to determine its traversing states in space. A robot is referred to as a mobile robot [2]. The planned path determines if a mobile robot is capable of achieving reliable as well as efficient autonomous navigation. On that account, path planning plays a very crucial role in mobile robot navigation. With the extensive application of mobile robots is leading to research on path planning more popular. Path planning could also be explained as the procedure of dividing the desired path movement into numerous iterative steps to make discrete motions to optimize some metrics.

Environment plays an important role in path planning problems. On the basis of the nature of the environment, path planning could be classified into offline and online path planning. In the former, the data concerning the environment accommodating stationary obstacles is known in advance, termed as static known environment, which is given as input, and thereafter the path is found utilizing the algorithm [3]. While in later, an environment that has static as well as dynamic obstacles is called a dynamic environment, and the robot makes use of sensors or real-time data acquisition equipment to figure out the location of the unknown obstacle that continually moves throughout the environment [4].

In a static and known environment, the robot knows the entire information of the environment before starting its traveling. Hence, the optimal path could be computed offline preceding the movement of the robot.

A path planning algorithm is necessary to be capable of finding a path that is collision-free, smooth, and drivable in real-time. The broadly used path planning algorithms are Sampling-based algorithms and Graph-based algorithms. The Probabilistic Road Map (PRM) [2] and the Rapidly Exploring Random Trees (RRTs) [2, 5] and its variants RRT^* [6] are the archetypal sample-based algorithms. Sampling-based algorithms often provide jerky and dynamically–impossible paths. The typical Graph-based path planning algorithms are A^* [7] and its variants [3, 8, 9, 10, 11, 12, 13, 14, 15], however they are likely to yield pathways that are not smooth and do not adhere to the non-holonomic constraints of the robot or the vehicle. In the case of right-angle

*Research Scholar and Associate professor, JNTUH College of Engineering, Kukatpally, Hyderabad-500090, India (kadari.neeraja @gmail.com).

†Professor, Department of CSE, JNTUH College of Engineering, Kukatpally, Hyderabad-500090, India (narsimha06@gmail.com)

turning angle, the turning motion of the mobile robot is broken down into three steps: slowdown, turning on the spot, and movement. All of this significantly slows down the mobile robot's pace.

In this paper a novel path planning algorithm called A Multi-Objective Hybrid Collision-free Optimal Path Finder (MOHC-OPF) is proposed based on A* to determine an optimal path in a hybrid environment . It is faster, and attains efficiency near to A^*. When the turning angle is right, the mobile robot's turning motion is split into three parts: slowing down, turning on the spot, and moving over fewer nodes. Compared to existing methods, this method takes kinematic constraints into account while making optimal driveable paths.

Our contributions in this paper are as follows:

1. An algorithm named Multi-Objective Hybrid Collision-free Optimal Path Finder (MOHC-OPF) is proposed to find collision-free near-optimal drivable paths in a hybrid i.e. continuous and discrete environment, and obtain efficiency near to A* algorithm.
2. It incorporates kinematic constraints such as orientation and steering angle of the robot to generate shortest, less execution time, and smooth paths apt for real-time environments.
3. It is applied to Simple and Complex static environments. The success rate is 100%.
4. The proposed technique shows efficient results in known static environments.

The rest of the paper is organised as follows. Section 2 examines exisiting path-planning methods for autonomous robots. Section 3 puts forward the proposed methodology and underlying algorithm for path planning in the static known environment. Section 4 includes experimental findings and evaluates the performance of the proposed method. The paper is concluded in Section 5, which also offers guidelines for future work.

2. Related Works. A^* [7] and its variants [3, 8, 9, 10, 11, 12, 13, 14, 15] are a large part of earlier path planning work which produces quick solutions for discrete-state spaces. The non-holonomic limitations of autonomous robots or vehicles make it impossible to use these strategies, which lead to paths that are not smooth. An alternative approach, that can guarantee kinematic feasibility is Sampling based approach in continuous coordinates, like Rapidly Exploring Random Trees (RRTs) [2, 5] and its variants such as RRT* [1, 16].

The configuration space is divided into cells by grid-based planners to map it into the grid form. For path planning, these planners use discrete approaches. As a collection of contiguous cells, they produce a path. However, due to their ineffective search efficiency, Dijkstra [17, 18, 19] and Extended Dijkstra [20] were not suitable for real-time path-planning applications. A* is a well-known Grid-based planner that was inspired by Dijkstra variations [7]. However, A* generates unrealistic paths. These restrictions on the development of grid-based planners have been addressed by showcasing the various A* variants [3, 8, 9, 10, 11, 12, 13, 14, 15]. Some of them are Time efficient A* [12] has shortened the time by reducing the exploration of state space. Iterative-deepening A* (IDA*) [3, 15] addressed large memory requirements issue. Nevertheless, it commonly leads to revisiting the same nodes, lengthening the exploration period and ensuring the optimality of grid-based algorithms like A* up to the grid resolution. However, they generate sharp turn paths which are not suitable for non-holonomic robots or vehicles.

Sampling-based algorithms like the Rapidly-exploring Random Tree (RRT)[2] has been proposed as an effective solution to computationally challenge motion planning problems. Although the RRT method is proven to be able to locate a workable solution rapidly, there is no assurance that it will be the optimal one. The authors developed a novel algorithm called RRT* [1] to get around this restriction. It maintains the same properties of the traditional RRT algorithm in terms of computation of feasible solutions and computational complexity while guaranteeing asymptotic optimality, i.e., virtually certain convergence of the solution returned by the algorithm to an optimal solution.

Larsen et al. describe a system that can plan routes for industrial robots. We compare sampling-based methods like PRM or RRT with methods like genetic algorithms that are based on Computational Intelligence (CI). Both sampling-based planners and genetic algorithms could be shown to be useful for planning the paths of 6-DOF industrial robots. On the one hand, RRT and TRRT are much faster than the genetic planner, but on the other hand, the paths found by the genetic planner are shorter, even though RRT* took 1200s to calculate. In real production situations, it seems that probabilistic planners are better to use because the paths are much smoother [31].

In many research problems, there are often too many things to take into account when making the final

choice. These things are basically features, which are variables. The more features there are, the harder it is to see what the training set looks like and then work on it. Sometimes, most of these things are related to each other, which makes them redundant. This is where algorithms that reduce the number of dimensions come in. Dimensionality reduction is the process of getting a set of principal variables from a group of random variables to cut down on the number of variables that need to be considered. It can be split into two parts: selecting features and getting information about them. Gunupudi et.al [32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54] discusses various dimensionality reductions and data transformation techniques used for various learning methods in their proposed research work.

Rybus et al. in thier proposed method creates a path for the arm that avoids collisions and leads to the gripper being in the right place and facing the right way and the spacecraft being in the right attitude. The main benefit of the proposed plan is that it might be possible to reach all of these goals at the same time. Experiments done on the planar air-bearing microgravity simulator (for a flat case) and numerical simulations done with the Monte Carlo method show that the method presented can be used in the real world (for a spatial case). In the planar case, the best path is compared to the path made by the Rapidly-exploring Random Trees (RRT) algorithm, which works in both directions. The proposed method makes a path that is shorter and easier to follow than the RRT method [55].

Santos et al. suggests a way to do multiobjective optimization that takes into account more than one part of motion. The effect of the mass of the manipulator is looked at. Then, the effect of multiple goals, such as the satellite's position, the arm's ability to be moved, and the maximum torque, on the best path is figured out. Also, the position of the end-effector, the avoidance of a collision between the arm and the spacecraft, and the reduction of torque requirements are thought to be goals to be minimised, depending on how uncertain the berthing box is. The numerical procedure uses a machine learning strategy that can learn from both training data and mission tasks. During an inverse kinematics analysis, when the Cartesian position is the input parameter and the estimated joint angle is the output, it is used. This information makes it more likely that the optimization process will end up with the right value for the joint angle. When there are five or more samples, the learning strategy works well to estimate the answer, and the result gets better as more data is added to the analysis. The fact that the numerical experiments looked at many different scenarios, metrics, and parameters proves that the proposed method works and is strong [56].

Tan et al., in their contribution, show a strong model using predictive control-based strategy for following the path of 4WS4WD vehicles when there are disturbances from the outside. Model predictive control (MPC) and control allocation are both parts of the strategy, which has an upper-lower structure. The main goal of this project is to make path tracking more stable and reliable by putting an MPC algorithm in the upper layer. In the offset model, the design of the controller takes into account both general disturbances caused by allocation errors and sudden disturbances caused by an outside force. Using the offset model, you can turn the robustness constraints into a linear matrix inequality to get a robust MPC control law. The Lyapunov stability theorem shows that the control law is stable in situations with more than one disturbance. By comparing the proposed robust algorithm to a similar path-tracking control algorithm and testing it on different types of uneven ground, it was found that the proposed algorithm was able to handle disturbances in the system well.[57]

Tzafestas et al. show a System on Chip (SoC) for autonomous, non-holonomic mobile robots to follow a path. The SoC is made up of a Xilinx Microblaze soft processor core, a parameterized Digital Fuzzy Logic Controller (DFLC) core, and a flow control algorithm. The authors made a fuzzy path tracking algorithm that works with the fuzzy controller. The FPGA board with the SoC was connected to a real differential-drive Pioneer 3-DX8 robot, which was used in field tests to see how well the tracking scheme worked overall. Quantization problems and the limits set by the way the system is set up are also talked about [58].

3. Proposed Path Planning System. The Path planning problem could be stated as "Given a map, find the cost-optimal path from start-state S to a goal-state G subject to the kinematic-constraints that are given by the non-holonomic robot.

The block diagram of path planing system is given in Fig. 3.1. The path finding criteria receives the data such as Path length, duration of Time taken to find a path and path is being smoothened from Optimization Criteria. The four main key elements of path planning system are:

1. *Environment*: A static environment is known in advance and represented as occupancy grid map.

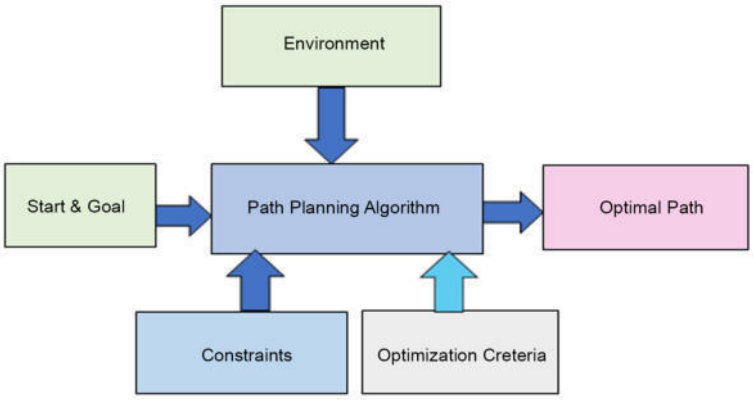


Fig. 3.1: Block Diagram for the proposed path planning system

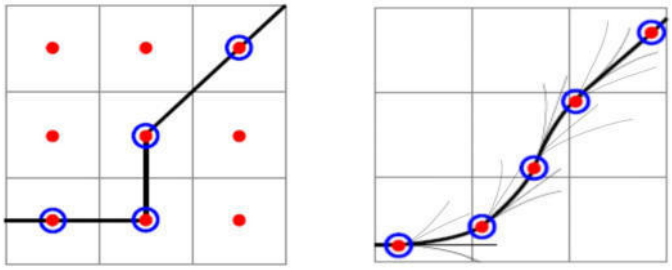


Fig. 3.2: a) Using traditional A* b) Using Proposed Method (MOHC-OPF)

2. *Constraints*: The constraints are kinematic constraints of non-holonomic car-like robot or vehicle.
3. *Optimization criterias*: Path length, Duration of time taken to find a path and Path smoothening.
4. *Path planning algorithm*: It is the main key element of the whole path planning system which finds a solution for a given path planning problem.

In this article an attempt is made to propose a new path planning algorithm named as "A Multi Objective Hybrid Collision-free Optimal Path Finder (MOHC-OPF)". It is a variation of traditional A*.

3.1. The working Principle of MOHC-OPF. The application of the traditional A^* algorithm [3] is restricted to discrete state spaces given in Fig.3.2(a). In case of a simple four or eight connected grid, it would need the robot or vehicle to turn on the spot, but in the presence of nonholonomic constraints it is not possible. The proposed algorithm, MOHC-OPF, is a variation of the traditional A* algorithm to overcome its limitations. It is done by incorporating the kinematics of the car-like robot or vehicle to predict the motion of the robot that is depending on the steering-angle in a continuous search-space. In the proposed algorithm, each continuous state is given by (x, y, θ), where (x, y) is the location and θ is the orientation of the robot or vehicle. It helps the path-finder to opt the right successor state which a nonholonomic robot or vehicle is able to follow. The states are expanded by one of the five steering-actions: maximum-left, left, maximum-right, right and no- steering actions, illustrated in Fig.3.2(b), which results in an arc of a circle with a minimum turning-radius on the basis of the simple car-like robot's or vehicle's constraints. Fig. 3.1.b. illustrates how the MOHC-OPF algorithm opts next states on the basis of these actions.

3.2. The proposed Algorithm MOHC-OPF. The algorithm, MOHC-OPF, uses 2 lists: Open-list and Closed-list. During the searching process they keep track of the states just like in the traditional A*. The open-list (OL) contains the neighbours of states that are already expanded during the search. The closed-list (CL) contains all states that are conclusively processed. The step-by-step description of the algorithm MOHC-OPF

is given below:

Output: the near optimum path for an autonomous car-like mobile robot from the given starting position to the desired target position.

Step 1: Initialize Open list (OL) and Close list (CL)

Step 2: Set NODE to be the start state

Step 3: Find the 5 neighbors of NODE for all 5 steering angles. The kinematic model of the robot with a global location of (X, Y, θ) was calculated by using a simple car like model as follows (LaValle, S.M. Planning Algorithms, 2006 [6]).

$$\begin{aligned} x &= u\ cos\ \theta \\ y &= u\ sin\ \theta \\ \psi &= \frac{U_s}{L\ tan\ \theta} \\ \rho &= \frac{L}{tan\ \psi_\theta} \end{aligned} \tag{3.1}$$

where u - denotes action havnig {0,1}, ψ is the steering angle and ρ_{min} is the minimum turning radius and L is distance between the front and rear axles of a simple car.

Step-4: If any of the neighbors is a goal state, then quit.

Step 5: For each neighbor is not lead to collision then Estimate cost of each neighbor using

$$Cost\_function\ f\ =\ g\ +\ h\ +\ additional\_cost \tag{3.2}$$

where g is distance from start node to current node and h is the value the predicted distance to the goal additional cost required for switching heading.

Step 6: Set NODE to the neighbor with the minimum cost_function f value and take the corresponding action on the map. Store the old NODE in a open list along with the f values.

Step 7: Repeat from step 3 until Open list is empty.

The pseudo code of MOHC-OPF Algorithm overview is given in Algorithm 1.

3.3. The Function. The minimum cost estimated value from any-node to the goal-node in the map is the function. It helps reduce the no.of nodes navigated.

Hence, the efficiency of the algorithm is directly impacted by the choice of function. It uses the Euclidean distance.

$$h = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \tag{3.3}$$

3.4. Multi Objective functions for Optimization:. In this work multiple objectives are taken into consideration to optimize the found path.

3.4.1. Cost Function. It computes the expense of driving from the current-position to an adjacent-position. This cost f is the sum of the cost from start node to current node(g) , the estimated cost (h) to the goal from current node and the additional cost for changing the angle.

Cost function $f = g + h+$ additional cost for switching orientation.

At each state in the algorithm min-cost next-state is added to the path and it guarantees that the final path is minimum cost.

3.4.2. Path Length. The final path is a collection of path segments $P = P_1, P_2, ..., P_n$.

So, the final path length is the sum of all path segments which connect Start and Goal states through intermediatory states.

$$Path\ Length\ =\ \sum P_i\ where\ i\ =\ 1\ to\ n \tag{3.4}$$

At each state minimum cost next-state is selected and minimum turning radius curves guarantee near shortest length path.

Algorithm 1 Pseudo code of Algorithm MOHC-OPF

Require: Start location x_{start}, *//initial point for robot path*
Goal Location x_{goal}, *//final point for robot path*
Binary Occupancy Grid Map *//representation in the form of binary matrix*
Ensure: $OL = \Theta$ *//Initialization of Open List: OL*
$CL = \Theta$ *//Initialization of Close List: OL*
Pred(x_{strt}) $\longleftarrow$ null *//initialization of prediction to* x_{start}
OL.push(x_{strt}) *//Adding current prediction to Open List: OL*
while $OL \neq 0$ **do** *//Continue as long as Open List (OL) is not empty*
x $\longleftarrow$ OL.popMin() *//Calculating next feasiable new x*
CL.push(x) *//Adding x to Close List*
if $round(x) = round(x_{goal})$ **then**
continue; *//checking if x is destination*
else
for $u \ \epsilon \ U(x)$ **do** *//if destination not reached, continue to predict new x to reach destination*
$x_{next} \longleftarrow f(x, u)$ *//calculating next x point as* x_{next}
if $collision_{check}(x) \ = \ False$) & x_{next} *not in* CL **then** *//checking if new x do not have an hurdle*
$g \longleftarrow g(x) + l(x, u)$ *//Initialization of Open List: OL*
if $(x_{next} notin OL) or (g < g(x_{next}))$ **then** *//calculating and validating next hop*
$Pred(x_{next}) \longleftarrow x$ *//moving robot to next hop*
$g(x_{next}) \longleftarrow g$ *//marking next hop*
$h(x_{next}) \longleftarrow (x_{next}, x_{goal})$ *//adding to current x to path travelled*
$cost(x_{next}) \longleftarrow g(x_{next}) + h(x_{next}) +$ *//calculating cost of travel*
additional cost for switching orientaion
$OL.push(x_{next}, cost(x_{next}))$ *//updating final cost*
else*//in the case validation fails for new x*
$OL.decreaseKey(x_{next})$ *//rolling back to previous x point*
end if
end if
end for*//predict new x till* $x_{goal} is reached$
end if
end while*//repeat this process till* $x_{goal} is reached$

3.4.3. Execution Time. Time taken to execute the proposed method for finding a path. At each node continuous state is rounded off to discrete state to avoid increase in the size of the search graph. We have used tic and toc in MatLab, to measure the performance of our execution.

3.4.4. Smooth Path. In this algorithm at each step next-node is selected based On the kinematic constraints which results in minimum turning radius curves. So, the generated final path is smooth.

4. Experimental results. In MATLAB 2021a on Windows10 64-bit with Intel core i5 NIVIDIA G5 processor extensive experiments were carried out to examine the performance of the proposed method the new Multi Objective Hybrid Collision free Optimal Path Finder (MOHC-OPF) and it was observed that the performance was improved significantly. All simulations were done on various static known maps of sizes 25 X 25, 50 X 50 and 100 X 100 for various start and goal positions.

The performance measures that are considered for performance evaluation of the proposed method are Path-Length-Mean and Average-Execution-Time, Standard-Deviation and Direct-Euclidian Distance between given start and goal locations on the map. Every time it generates same path, a deterministic path planning approach will find the same path when the scenario and parameters are the same, like A*. So, Standard deviation is zero. Optimized path length and execution time are recorded, since the results will be compared with the performance of existing methods for discussion.

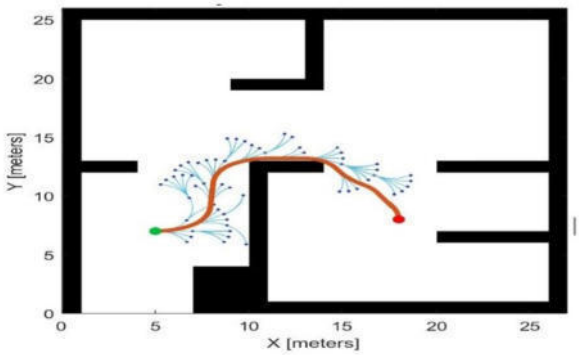


Fig. 4.1: Generated Near optimal path

Table 4.1: Results for Simple map

	Proposed method MOHC-OPF
Path Length Mean	18.7918
Path Length Std	0
Execution Time	0.29964
Direct Path Length	12.1440

Table 4.2: Results for Complex Map

	Proposed method
Path Length Mean	64.5589
Path Length Std	0
Avg Execution Time	0.457936
Direct Path Length	51.856

For the experimentation we have considered three sample environments one is simple and the other two are complex known environments. The criteria for distinguishing simple and complex is taken based on the matrix size and the number of decisions to make. In our case we have taken a matrix size of 25X25 and number of hurdles in the static known environment to 3 as simple map. In the other case, either higher matrix dimension or higher number of hurdles is categorized as complex known static environment map.

4.1. Case study one – Simple map. The proposed method MOHC-OPF was applied on simple static known environment of size 25 X 25 that contains static known obstacles. Fig 4.1. shows generated near optimal path for this environment.

The Table 4.1 shows the Path_Length_Mean and Average-Execution-Time results for Simple map.

4.2. Case Study Two: Complex Map. In case two, a more complex case is presented; the environment contains static known obstacles and it is a complex maze. Fig 4.2. shows a near optimal path generated by MOHC-OPF for this complex environment.

The Table 4.2. shows the Path_Length_Mean and Average-Execution-Time results for Complex map.

4.3. Case Study Three : Package pickup in Warehouse scenario. In case three, a Package pickup in warehouse scenario is presented and contains static known obstacles. Fig 4.3 shows MOHC-OPF generated near optimal path to go to package pickup location based on known static obstacles.

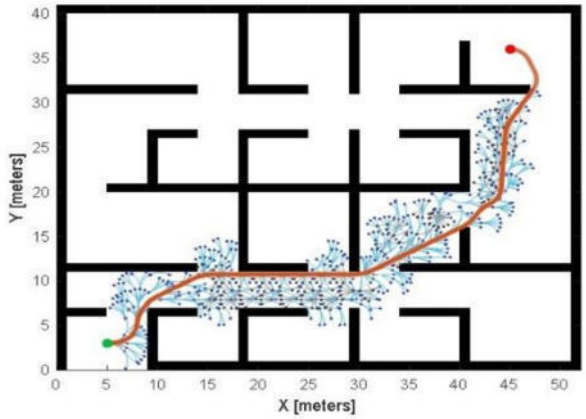


Fig. 4.2: Generated Near optimal path

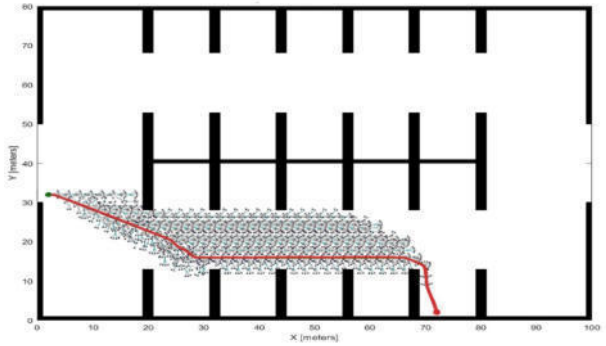


Fig. 4.3: MOHC-OPF- Near optimal path

Table 4.3: Results for Package pickup in warehouse scenario

	Proposed method MOHC-OPF
Path Length Mean	85.09725
Path Length Std	-
Avg Execution Time	0.927327
Direct Path Length	76.174

The Table 4.3. shows the Path_length mean and execution time results for this map.

4.4. Performance Evaluation. In this section, we present the performance efficiency of proposed algorithm MOHC-OPF and compare it with existing methods PRM, RRT and RRT*on the above 3 case studies.

Rigorous experimentation has been done on proposed method and existing methods PRM, RRT and RRT*. The performance comparison has shown in the following tables and graphs. The Proposed Method MOHC-OPF has given optimal path length nearer to A* when compared to PRM, RRT and RRT*. Execution time of MOHC_OPF is comparatively less.

4.4.1. Case Study 1: Simple Static Known Map. The table 4.4 shows the results of 500 iterations of proposed method as well as existing methods PRM, RRT and RRT*. These approaches proposed in the

Table 4.4: Performance efficiency Comparison in Simple Map

Planner/Performance Metric	**Path length Mean**	**Path Length Standard Deviation**	**Average Execution Time**
PRM	37.31334	8.25209	0.312554
RRT	25.77003	9.2276	1.312895
RRT Star	43.45421	3.315177	1.005047
Proposed Method for MOHC_SPF	18.79182	-	0.29964

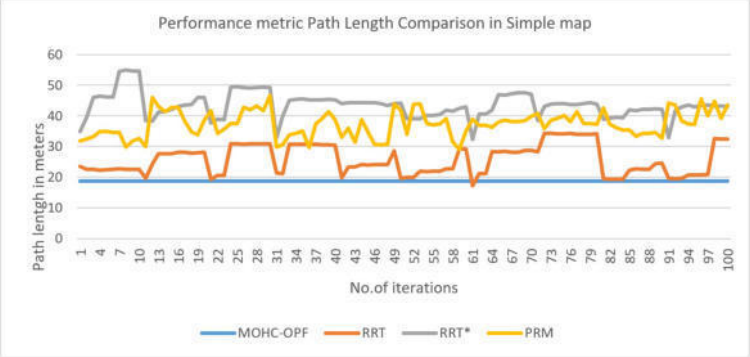


Fig. 4.4: Performance metric Path Length Comparison in Simple map

literature survey are tested with our sample space and the results obtained are compared. It shows the comparison of performance efficiency of the proposed method MOHC-OPF with the existing systems PRM, RRT and RRT* in terms of the performance metrics Path length and Execution Time. These methods are applied on Simple static known map. The values clearly shows that the proposed method has given shortest path length and less execution time taking paths.

Fig. 4.4 displays the performance graph of the proposed method and existing methods PRM, RRT and RRT* across number of iterations and path length. The blue line indicates path lengths generated by the proposed method MOHC-OPF. It generated shortest length paths in almost all iterations out of 100 iterations compared to other existing methods in Simple static known map.

Fig. 4.5 shows the performance graph of the proposed method and existing methods PRM, RRT and RRT* in terms of execution time for 100 iterations. The blue line indicates execution time taken by the proposed method MOHC-OPF. It took very less execution time in almost all iterations of 100 iterations compared to existing methods in Simple static known map.

Case Study 2: Complex Static Known Map Table 4.5 displays the comparison performance of proposed method and also existing methods PRM, RRT and RRT*. The Proposed method and these exiting methods applied 500 times on Complex Static Known Map. The comparison of performance efficiency is shown in the performance metrics Path length and Execution Time. The values clearly exhibits that the proposed method is performing efficiently when the complexity of map is increased.

Fig. 4.6 shows the efficient performance of the proposed method compared to existing methods PRM, RRT and RRT* in 100 iterations. The proposed method performance is indicated by blue line and it generated shortest length paths in all iterations compared to other existing methods in Complex static known map.

Fig. 4.7 displays the performance efficiency in terms of execution time. The proposed method MOHC-OPF is represented by blue line generated less time taking paths in comparison to existing methods PRM, RRT and RRT in almost all 100 iterations in Complex static known map.

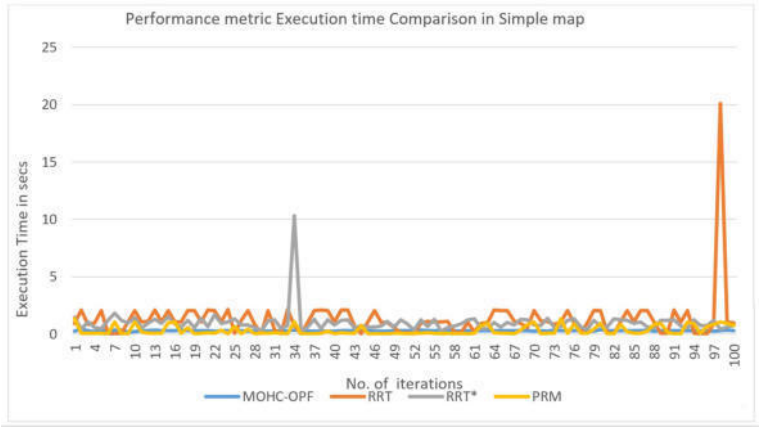


Fig. 4.5: Performance metric Execution time Comparison in Simple map

Table 4.5: Performance efficiency comparison in complex map

Planner/Performance Metric	**Path length Mean**	**Path Length Standard Deviation**	**Average Execution Time**
PRM	71.87536	7.380644	0.815483
RRT	107.5507	9.2276	2.435469
RRT Star	95.49642	24.78597	2.420371
Proposed Method for MOHC_SPF	64.5589	-	0.45973

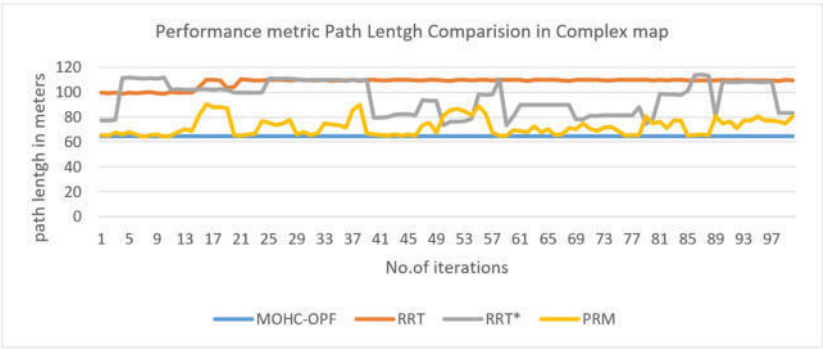


Fig. 4.6: Performance metric Path length Comparison in Complex map

Case Study Warehouse 3: Static Known Map for Package Pickup The Proposed method and the exiting methods PRM, RRT and RRT* applied 500 times on Warehouse scenario. Table 4.6 shows the comparison of performance metrics Path length and Execution Time of the proposed method as well as the existing methods. The proposed method generated shortest length and shortest time taking paths. The proposed method has exhibited efficient performance for this difficult warehouse map.

Fig 4.8 shows the efficient performance of the proposed method compared to existing methods PRM, RRT and RRT* in 100 iterations. The proposed method performance is indicated by blue line and it generated

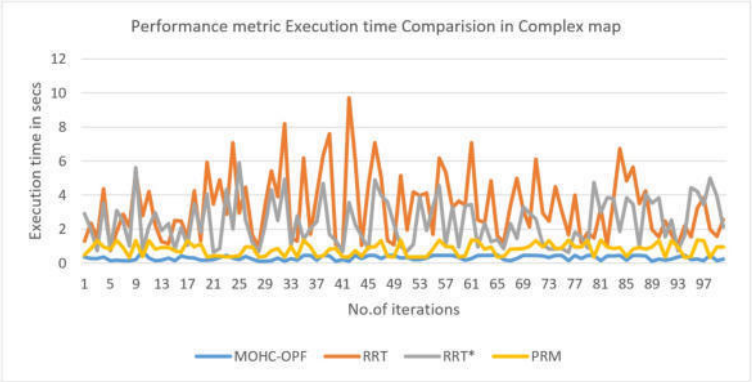


Fig. 4.7: Performance metric Execution time Comparison in Complex map

Table 4.6: Performance efficiency comparison in complex map

Planner/Performance Metric	**Path length Mean**	**Path Length Standard Deviation**	**Average Execution Time**
PRM	114.6131	16.5637	1.283141
RRT	186.1902	33.929	3.722676
RRT Star	95.49642	21.03764	3.489829
Proposed Method for MOHC_SPF	85.09725	-	0.927327

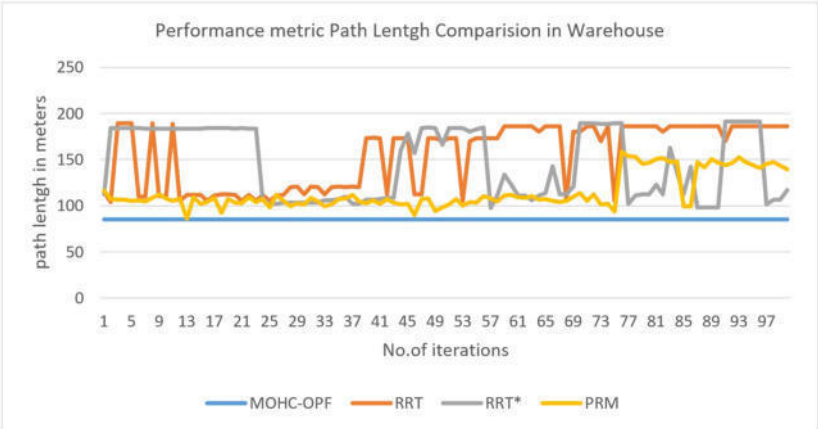


Fig. 4.8: Performance metric Path length Comparison in Warehouse map

shortest length paths in all iterations compared to other existing methods in Complex static known map.

The proposed method MOHC-OPF performance compared to existing methods PRM, RRT and RRT* in 100 iterations as shown in Fig 4.9. The proposed method has given the shortest time taking paths in almost all iterations compared to other existing methods in Warehouse static known map.

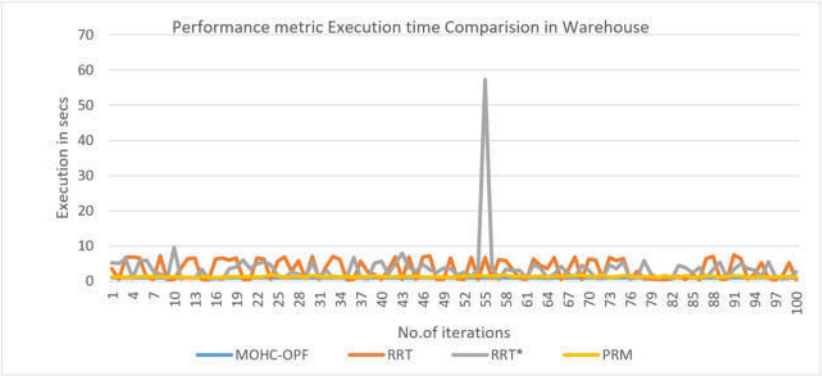


Fig. 4.9: Performance metric Execution time Comparison in Warehouse map

5. Conclusion and future work. In this paper, a new Multi Objective Hybrid Collision-free Optimal Path Finder (MOHC-OPF) is proposed to deal with the optimal path planning for autonomous mobile robots in static known environments. This is fundamentally works like A* but respects kinematic constraints of the robot. Hence, the drivable smooth paths are guaranteed to be generated which are needed for real time scenarios. The success rate for finding a smoother collision free optimized path is 100%. A Multiple Objective Optimization such as Path length, Execution Time, Cost function and Path smoothening are achieved by MOHC-OPF. In the experiments carried out during this work, it is clearly hlobserved that by using the proposed method recorded overall reduction of 15% in path length and 20% in execution-time in our experiments compared to existing methods PRM, RRT and RRT*. The proposed method has shown higher performance efficiency in complex environments compared to these existing methods

The proposed method MOHC-OPF can be adopted to environments which includes dynamic obstacles in future. It can also be applied to higher dimension environments.

REFERENCES

[1] S. Karaman, E. Frazzoli , *Incremental sampling-based algorithms for optimal motion planning*, in Proc. Robotics: Science and Systems (RSS), 2010

[2] 5. S. Karaman, E. Frazzoli, LU-*Optimal kinodynamic motion planning using incremental sampling-based methods*, in Proc. IEEE Conf. on Decision and Control (CDC), Dec. 2010.

[3] S. M. LaValle, *SRapidly-Exploring Random Trees: A New Tool for Path Planning*, 1998. Available online: http://lavalle.pl/papers/Lav98c.pdf.

[4] I. Noreen, A. Khan, Z. Habib, *A Optimal Path Planning using Memory Efficient A**, In Proceedings of the IEEE International Conference on Frontiers of Information Technology, Islamabad, Pakistan, 19–21 December 2016; pp. 142–146.

[5] P. E. Hart, N. J. Nilsson and B. Raphael, *A Formal Basis for the Determination of Minimum Cost Paths*, in IEEE Trans. Syst. Sci. Cybern. 1968, 4, 100–107.

[6] S.M. LaValle, *Planning Algorithms*, Cambridge University Press: Cambridge, UK, 2006. (1983), pp. 275–283.

[7] C.W. Warren , *Fast path planning using modified A* method*, In Proceedings of the IEEE International Conference on Robotics and Automation, Atlanta, GA, USA, 2–6 May 1993..

[8] L. Cheng, C. Liu, B. Yan, *Improved hierarchical A-star algorithm for optimal parking path planning of the large parking lot*, In Proceedings of the 2014 IEEE International Conference on Information and Automation (ICIA), Hailar, Hulun Buir, China, 28–30 July 2014; pp. 695–698.

[9] R. Korf, *Depth First Iterative Deepening: An Optimal Admissible Tree Search*, AI Journal 27 (1): 97-109, 1985.

[10] František. Ducho, Andrej. Babinec, Martin. Kajan, Peter. Beno, Martin. Florek, Tomáš. Fico, Ladislav. Jurišica, *Path planning with modified A star algorithm for a mobile robot* , In Proceedings of the Elseveir, International Conference on the Modelling of Mechanical and Mechatronic Systems MMaMS 2014.

[11] Y. Zhang, G. Tang, L. Chen, *Improved A* Algorithm For Time-dependent Vehicle Routing Problem*, In Proceedings of the 2012 International Conference on Computer Application and System Modeling, ICCASM 2012, Taiyuan, China, 27–29 July 2012.

[12] C. Liu, Q. Mao, X. Chu, *An improved A-Star algorithm considering water current, traffic separation and berthing for*

vessel path planning, Applied Sciences, vol. 9, no. 6, pp. 1–17, 2019.

[13] Guruji. Akshay Kumar, Agarwal. Himansh, D.K. Parsediya, *A Time-efficient A* Algorithm for Robot Path Planning*, Procedia Technology, 23, p144–149.

[14] Ismail, AL-Taharwa, Alaa Sheta, Mohammed Al-Weshah, *A mobile robot path planning using genetic algorithm in static environment*, Journal of Computer Science 4.4 (2008): 341-344.

[15] R. Korf, *Optimal Path Finding Algorithms*, Search in AI. Edited by Kanal, L., and Kumar, V., Springer Verlag, Symbolic Computation: 200-222, 1988.

[16] B. Hu, Z. Cao, M. Zhou , *An Efficient rrt-based framework for planning short and smooth wheeled robot motion under kinodynamic constraints*, IEEE Transactions on Industrial Electronics, vol. 68, no. 4, pp. 3292–3302, 2021.

[17] M. Barbehenn , *A Note on the Complexity of Dijkstra's Algorithm for Graphs with Weighted Vertices* , IEEE Transactions on Computers, Vol. 47, No.2, 1998.

[18] T. Dudi, R. Singhal,R. Kumar , *Shortest Path Evaluation with Enhanced Linear Graph and Dijkstra Algorithm*, In Proceedings of the 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE); 2020; pp. 451–456.

[19] A. Ammar, H. Bennaceur,I. Chaari, A. Koubaa, M. Alajlan, *Relaxed Dijkstra and A* with linear complexity for robot path planning problems in large-scale grid environments*, Soft Comput. 2015, 20, 1–23.

[20] O.A. Gbadamosi, D.R. Aremu , *Design of a Modified Dijkstra's Algorithm for finding alternate routes for shortest-path problems with huge costs*, In Proceedings of the 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS), Lagos, Nigeria, 18–21 March 2020; pp. 1–6.

[21] G. Qing, Z. Zheng, X. Yue , *Path-planning of automated guided vehicle based on improved Dijkstra algorithm*, In Proceedings of the 2017 29th Chinese control and decision conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 7138–7143.

[22] M. Noto, K. Univ, H. Sato , *A Method for the Shortest Path Search by Extended Dijkstra Algorithm*, In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Nashville, TN, USA, 8–11 October 2000; Volume 3, pp. 2316 – 2320.

[23] X. Cui, H. Shi , *A* - based Pathfinding in Modern Computer Games*, In Proceedings of the IEEE International Conference on Robotics and Automation, Atlanta, GA, USA, 2–6 May 1993.

[24] L. KavrakiP. M. Gibson, *Probabilistic roadmaps ˇ for path planning in high-dimensional configuration spaces*, IEEE Trans. on Robotics and Automation.

[25] O. Souissi, R. Benatitallah, D. Duvivier, A. Artiba, N. Belanger, P. Feyzeau, *Path planning: A 2013 survey*, In Proceedings of the 2013 International Conference on Industrial Engineering and Systems Management (IESM), Rabat, Morocco, 28–30 October 2013; pp. 1–8.

[26] T. Mac, C. Copot, D. Tran, R. Keyser, *approaches in robot path planning: A survey*, Robot. Auton. Syst. 2016, 86.

[27] Liu, Shuang, Sun. Dong, *Minimizing Energy Consumption of Wheeled Mobile Robots via Optimal Motion Planning*. IEEE/ASME Transactions on Mechatronics-2014, 19(2), p401–411.

[28] J. Latombe, *Motion planning: A journey of robots, molecules, digital actors, and other artifacts*, Int'l J. of Robotics Research, vol. 18, no. 11, pp. 1119–1128, 1999..

[29] H.M. Choset, *Principles of Robot Motion: Theory, Algorithms, and Implementation*, The MIT Press: Cambridge, MA, USA, 2005.

[30] B. Chen, G. Quan, *NP-Hard Problems of Learning from Examples*, In Proceedings of the 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Jinan, China, 8–20 October 2008; Volume 2, pp. 182–186.

[31] L. Larsen, J. Kim, M. Kupke, A. Schuster, *Automatic Path Planning of Industrial Robots Comparing Sampling-based and Computational Intelligence Methods*, Procedia Manufacturing, Volume 11, 2017, Pages 241-248, ISSN 2351-9789.

[32] G. R. Kumar. M. Nimmala, G. Narsimha, *An Approach for Intrusion Detection Using Novel Gaussian Based Kernel Function*, Journal of Universal Computer Science, Volume 22, Issue 4, 2016, pp 589-604 ISSN: 0948-6968

[33] G. R. Kumar. M. Nimmala, G. Narsimha, *Intrusion Detection – A Text Mining Based Approach", International Journal of Computer Science and Information Security (IJCSIS)*, Special issue on "Computing Application", Volume 14, 2016, pp 76-88

[34] G. R. Kumar. M. Nimmala, G. Narsimha, *A Novel Similarity Measure for Intrusion Detection using Gaussian Function*, Technical Journal of the Faculty of Engineering, TJFE, Volume 39, Issue 2, 2016, pp 173-183

[35] G. R. Kumar. M. Nimmala, G. Narsimha, *CLAPP: A self constructing feature clustering approach for anomaly detection*, Future Generation Computer Systems, Volume 74, 2017, Pages 417-429

[36] G. R. Kumar. M. Nimmala, G. Narsimha, *A Feature Clustering Based Dimensionality Reduction For Intrusion Detection (FCBDR)*, IADIS International Journal on Computer Science & Information Systems . 2017, Vol. 12 Issue 1, p26-44.

[37] G. Narsimha, G. R. Kumar. M. Nimmala, *Optimising business intelligence results through strategic application of software process model*, Int. J. of Intelligent Enterprise, 2017 Vol.4, No.1/2, pp.128 – 142

[38] K. R. Gunupudi. M. Nimmala, G. Narsimha, *An Evolutionary Feature Clustering Approach for Anomaly Detection Using Improved Fuzzy Membership Function: Feature Clustering Approach for Anomaly Detection*. International Journal of Information Technology and Web Engineering (IJITWE), 14(4), 19-49. doi:10.4018/IJITWE.2019100102

[39] G. Narsimha, K. R. Gunupudi. M. Nimmala, *UTTAMA: An Intrusion Detection System Based on Feature Clustering and Feature Transformation*, Foundations of Science, 2019, https://doi.org/10.1007/s10699-019-09589-5

[40] K. R. Gunupudi. M. Nimmala, G. Narsimha, *An Evolutionary Feature Clustering Approach for Anomaly Detection Using Improved Fuzzy Membership Function: Feature Clustering Approach for Anomaly Detection*. International Journal of Information Technology and Web Engineering, vol. 14, no. 4, Oct. 2019, pp. 19–49. DOI.org (Crossref), https://doi.org/10.4018/IJITWE.2019100102.

[41] G. Narsimha, G. R. Kumar. M. Nimmala, *Optimising business intelligence results through strategic application of software process model*, International Journal of Intelligent Enterprise, Jan 2017, Vol. 4, Issue 1-2, pp. 128-142
[42] G. R. Kumar. M. Nimmala, G. Narsimha, *An approach for Intrusion Detection using Text Mining Techniques*. In Proceedings of the The International Conference on Engineering & MIS 2015 (ICEMIS '15). ACM, New York, NY, USA, , Article 63 , 6 pages. DOI: http://dx.doi.org/10.1145/2832987.2833076
[43] G. Narsimha, G. R. Kumar. M. Nimmala, *Strategic Application of Software Process Model to Optimize Business Intelligence Results*. In Proceedings of the The International Conference on Engineering & MIS 2015 (ICEMIS '15). ACM, New York, NY, USA, , Article 44 , 6 pages. DOI: http://dx.doi.org/10.1145/2832987.2833053
[44] G. Narsimha, G. R. Kumar. M. Nimmala, *Intrusion Detection Using Text Processing Techniques: A Recent Survey*. In Proceedings of the The International Conference on Engineering & MIS 2015 (ICEMIS '15). ACM, New York, NY, USA, , Article 55 , 6 pages. DOI: http://dx.doi.org/10.1145/2832987.2833067
[45] G. Narsimha, G. R. Kumar. M. Nimmala, *An improved k-Means Clustering algorithm for Intrusion Detection using Gaussian function*. In Proceedings of the The International Conference on Engineering & MIS 2015 (ICEMIS '15). ACM, New York, NY, USA, , Article 69 , 7 pages. DOI: http://dx.doi.org/10.1145/2832987.2833082
[46] G. Narsimha, G. R. Kumar. M. Nimmala, *An approach for intrusion detection using fuzzy feature clustering*, 2016 International Conference on Engineering & MIS (ICEMIS), Agadir, 2016, pp. 1-8. doi: 10.1109/ICEMIS.2016.7745345
[47] G. Narsimha, G. R. Kumar. M. Nimmala, *Text mining based approach for intrusion detection*, 2016 International Conference on Engineering & MIS (ICEMIS), Agadir, 2016, pp. 1-5. doi: 10.1109/ICEMIS.2016.7745351
[48] G. Narsimha, G. R. Kumar. M. Nimmala, *Design of novel fuzzy distribution function for dimensionality reduction and intrusion detection*, 2016 International Conference on Engineering & MIS (ICEMIS), Agadir, 2016, pp. 1-6.
[49] M. Nimmala, G. R. Kumar. G. Narsimha, *Optimization of Access Points in Wireless Sensor Network: An Approach towards Security*, Intelligent Systems in Cybernetics and Automation Theory, 2015, Volume 348, pp 299-306
[50] G. R. Kumar. M. Nimmala, G. Narsimha, *Evolutionary approach for intrusion detection*, 2017 International Conference on Engineering & MIS (ICEMIS), Monastir, 2017, pp. 1-6. doi: 10.1109/ICEMIS.2017.8273116
[51] M. Nimmala, G. R. Kumar. G. Narsimha, *A fuzzy measure for intrusion and anomaly detection*, 2017 International Conference on Engineering & MIS (ICEMIS), Monastir, 2017, pp. 1-6. doi: 10.1109/ICEMIS.2017.8273113
[52] A. Cheruvu, G. R. Kumar. M. Nimmala, G. Narsimha, *Feature Clustering for Anomaly Detection Using Improved Fuzzy Membership Function*. In Proceedings of the Fourth International Conference on Engineering & MIS 2018 (ICEMIS '18). ACM, New York, NY, USA, Article 35, 9 pages. DOI: https://doi.org/10.1145/3234698.3234733
[53] A. Shadi, M. Nimmala, G. R. Kumar. G. Narsimha, *A recent survey on challenges in security and privacy in internet of things*. In Proceedings of the 5th International Conference on Engineering and MIS (ICEMIS '19). ACM, New York, NY, USA, Article 25, 9 pages. DOI: https://doi.org/10.1145/3330431.3330457
[54] K. R. Gunupudi, M. Nimmala, G. Narsimha, *Similarity function for intrusion detection*. In Proceedings of the 5th International Conference on Engineering and MIS (ICEMIS '19). ACM, New York, NY, USA, Article 28, 4 pages. DOI: https://doi.org/10.1145/3330431.3330460
[55] T. Rybus, M. Wojtunik, F. L. Basmadji, *Optimal collision-free path planning of a free-floating space robot using spline-based trajectories*. Acta Astronautica, Volume 190, 2022, Pages 395-408, ISSN 0094-5765, https://doi.org/10.1016/j.actaastro.2021.10.012.
[56] R. R. Santos, D. A. Rade, I. M. da Fonseca, *A machine learning strategy for optimal path pl*. Acta Astronautica, Volume 191, 2022, Pages 41-54, ISSN 0094-5765, https://doi.org/10.1016/j.actaastro.2021.10.031.
[57] Q. Tan, C. Qiu, J. Huang, Y. Yin, X. Zhang, H. Liu, *Path tracking control strategy for off-road 4WS4WD vehicle based on robust model predictive control*. Robotics and Autonomous Systems, Volume 158, 2022, 104267, ISSN 0921-8890 https://doi.org/10.1016/j.robot.2022.104267.
[58] S.G. Tzafestas, K.M. Deliparaschos, G.P. Moustris, *Fuzzy logic path tracking control for autonomous non-holonomic mobile robots: Design of System on a Chip*. Robotics and Autonomous Systems, Volume 58, Issue 8, 2010, Pages 1017-1027, ISSN 0921-8890, https://doi.org/10.1016/j.robot.2010.03.014.

Edited by: Vinoth Kumar
Received: Aug 19, 2022
Accepted: Nov 8, 2022


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 403–420, DOI 10.12694/scpe.v23i4.2051

SEGMENTATION AND PRE-PROCESSING OF INTERSTITIAL LUNG DISEASE USING DEEP LEARNING MODEL

PRIYANKA YADLAPALLI *AND D. BHAVANA †

Abstract. Medical image processing involves using and examining 3D human body images, which are most frequently acquired through a computed tomography scanner, to diagnose disorders. Medical image process- ing helps radiologists, engineers, and clinicians better comprehend the anatomy of specific patients or groups of patients. Due to recent advancements in deep learn ing techniques, the study of medical image analysis is now a quickly expanding area of research. Interstitial Lung Disease is a chronic lung disease that worsens with time. This condition cannot be completely treated when the lungs have been damaged. Early detection, on the other hand, aids in the control of the disease. It causes lung scarring as a result. The first methodology characterizes lung tissue utilizing first order statistics, grey live occurrence, run length matrices, and fractal analysis. It was suggested by Uppaluri et al in one instance. In the pre-processing step, patients' CT scans are presented using various color map models for better understanding of data-set. and also for determining the patients final Force Vital Capacity and Confidence values using a Pytorch model with leaky relu activation function. These variables can be used to determine whether a person has a disease. Segmentation is a crucial stage in employing a computer assisted diagnosis system to estimate interstitial lung disease. Accurate segmentation of aberrant lung is essential for a trustworthy computer-aided illness diagnosis. Using separate training, validation, and test sets, we proposed an efficient deep learning model using Unet architecture and Densenet121 to segment lungs with Interstitial Lung Disease. The proposed segmentation model distinguishes the exact lung region from the ct slice background. To train and evaluate the algo rithm, 176 sparsely annotated Computed Tomography scans were utilized. The training was completed in a supervised and end to end manner. Contrary to current approaches, the suggested method yields accurate segmentation results without the requirement for re-initialization. We were able to achieve an accuracy of 92.59 percent after training the proposed model with Nvidia's CUDA GPU.

Key words: Medical Image Processing, Interstitial Lung Disease, Force Vital Capacity, Pytorch, Densenet121

AMS subject classifications. 68T05

1. Introduction.

Deep learning is a component of machine learning, an artificial intelligence subset. Broadly speaking, "artificial intelligence" refers to techniques that enable computers to mimic human behaviour. All of this is made possible through machine learning, which consists of a collection of algorithms that are taught using data. A subset of machine learning called "Deep Learning" is inspired by the structure of the human brain. Deep Learning systems continuously analyse data using a predetermined logical framework in an effort to make decisions that are comparable to those made by humans. To accomplish this, deep learning uses neural networks, a multi-layered architecture of algorithms. Advances in deep learning for medical image analysis have created new opportunities for morphological and textural pattern discovery in images using only data [2].

Deep learning models can fix issues that machine learning models cannot is shown in Fig.1.1. and deep learning has been given credit for all recent developments in artificial intelligence. There would be no self-driving cars or personal assistants like Alexa and Siri without deep learning. The Google Trans- late app wouldn't have changed from 10 years ago, and neither Netflix nor Youtube would know what movies or TV shows we enjoy or dislike. These developments are all driven by neural networks. Processing power is required for deep learning in a substantial way. High performance graphics processing units (GPUs) are appropriate since they have a lot of memory and can do a lot of tasks over multiple cores. On-premises management of several GPUs, however, can put a heavy burden on internal resources and be prohibitively expensive to scale

*Asst. Professor, Department of ECE, GRIET Hyderabad,Telangana, India (bhavanaece@kluniversity.in)
†Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur District, AP 522502, India (bhavanaece@kluniversity.in)

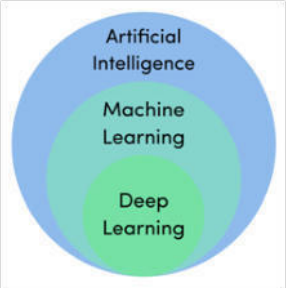


Fig. 1.1: Deep Learning

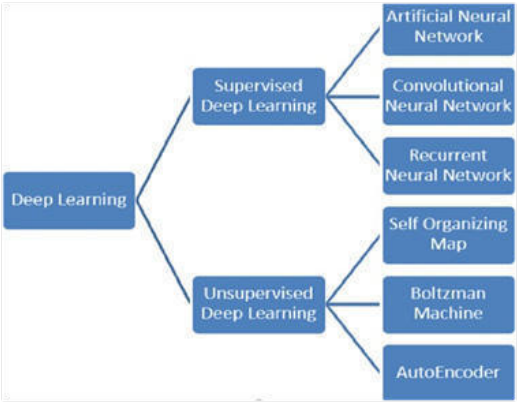


Fig. 1.2: Types of Deep Learning Algorithms

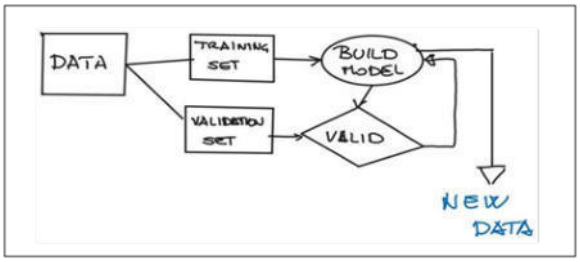


Fig. 1.3: Supervised Learning Algorithm

1.1. Deep Learning Algorithms. Deep learning algorithms are dynamically constructed to run over numerous layers of neural networks, which are simply a collection of pre-trained decision-making networks. Each of them is then processed through a series of simple layered representations before being moved to the next layer.

The most popular type of Deep Learning is Supervised Learning. During the training phase of supervised learning, a set of examples, known as training set, is submitted as input to the system. Each input is tagged with a desired output value, so the system understands what the output should be when the input is given. The brief classification of deep learning algorithms is shown in Fig.1.2.

Supervised learning shown in Fig.1.3 is the most common type of Deep Learning.

1. A set of examples known as the training set is submitted to the system as input during the training

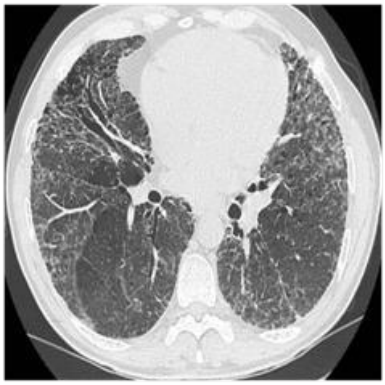


Fig. 2.1: Lung effected with ILD

phase of supervised learning
2. Each input is labelled with a desired output value so that the system knows what to expect when the input is given

2. Interstitial Lung Disease (ILD). Several disorders that result in scarring in the lungs are referred to together as "interstitial lung disease" (ILD). It is also known by the name diffuse parenchymal lung disorders. Scarring causes stiffness in the lungs, making it difficult to breathe and supply oxygen to the circulation. ILDs result in permanent lung damage that develops over time. Medication can help to delay the progression of interstitial lung disease, but many people never fully restore their lung functions [3].

The different types of Interstitial Lung Disease includes Mycoplasma pneumonia, Asbestosis, Acute interstitial pneumonitis, Hypersensitivity pneumonitis. Inhaling asbestos, a fibre which is present in building materials, can result in asbestosis, a kind of interstitial lung disease shown in Fig.2.1. Acute interstitial pneumonitis is caused due to rapidly progressive dyspnoea which can be developed from days to weeks . Hypersensitivity pneumonitis is the medical term for the condition that occurs while breathing in dust, mould, or other irritants over an extended period of time.The symptoms are: shortness of breath at rest or when exercised, cough that isn't watery, fatigue, unintentional weightloss, clubbing of fingernails and toes, pain in joints and muscles. After an injury, an abnormal healing reaction in the lungs appears to be the origin of interstitial lung disease.Our body typically produces just enough new tissues to correct the existing damage. When the repair system malfunctions in interstitial lung disease, the tissue that surrounds the air sacs gets scars and thickens. This complicates the flow of oxygen into your bloodstream. Long-term exposure to chemicals and pollution can harm your lungs. Examples include silica dust, asbestos fibres, grain dust, coal dust, and radiation treatments (lung or breast cancer). Interpretation of ILD and other lung illnesses is made simpler by the accurate view of lungs blood vessels and tissues [4]. Hence Such anomalies are frequently found using computer aided diagnosis methods [5].

Segmentation techniques shown in Fig.2.2 that were manual or semiautomatic in early CAD systems are now replaced by completely automatic segmentation algorithms. A local texture classification algorithm that goes across images and generates a map of diseases has typically been used to diagnose ILDs. Deep artificial neural networks and deep convolution neural networks have recently received a lot of interest following their outstanding performances in the Image-net Large Scale Visual Recognition Competition in 2012 [6]. These networks have been around for a while, but only recently have they been able to perform well, in large part because of the abundance of annotated data accessible, the extreme parallelism power of GPU, and a few clever design choices [7]. Recent research has also looked into the potential advantages of deep learning algorithms for medical picture processing, and the preliminary findings are encouraging [8].

2.1. Segmentation. Segmenting the lungs is a particularly tough task due to in homogeneities in the lung region and pulmonary structures with similar densities, such as arteries and veins. Several segmentation approaches have been put out in the past. These methods can be broadly categorised as rule based, knowledge

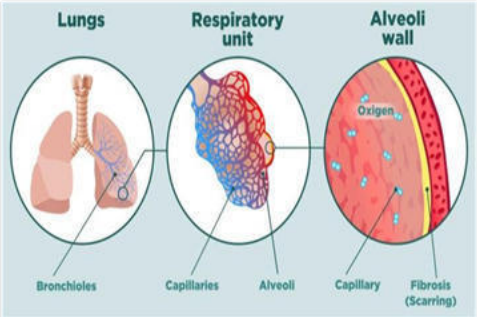


Fig. 2.2: Dissection of ILD Lung consisting of Scarring

based, pixel based, texture analysis based, and model based. Shanhui Sun and Geoffrey McLennan's [9] rule-based technique of lung segmentation identifies potential airways before applying rules to determine the lungs.

Benefits of segmentation:

1. Image segmentation is the most important medical imaging process. It uses an automatic process to extract the ROI (region of interest). Segmentation divides an image into areas of interest, such as body organs or tissue.
2. Application of classification neural network algorithms to segmented radiological images can significantly improve segmentation accuracy.
3. While manual segmentation increases computational costs, but it significantly reduces overall disease diagnosis costs.

3. Related Work. Computer aided diagnostic analysis is necessary for treatment planning, tracking the course of illness of a diseases, identifying how well a patient responds to a prescribed medication, and other purposes. Medical image modalities create images using a range of techniques for a variety of objectives. Hamzah and Mohd Faizal Mohd present a novel method for diagnosing interstitial lung disease (ILD) using high resolution computed tomography images and correlation and regression algorithms on texture measure [10].This study predicts each patient's final FVC measurements and provides a Confidence value so that different Machine Learning models can be compared and contrasted [11].

A Convolutional Neural Network (CNN) was used to classify 2D lung images as part of a novel approach for diagnosing interstitial lung disease. The suggested method obtained 94 percent accuracy and high sensitivity on a dataset of 30 CT scan images. The same dataset was classified with an accuracy of 86% using the SVM classifier [12].

In order to identify ground glass opacities (GGO) on HRCT, K.R. Heitmann and associates used neural networks and expert rules. To identify GGO in this work, a hybrid model made up of three single nets and an expert rule was constructed [13]. In order to create a consensus categorization for the entire group of IIPs (Idiopathic Interstitial Pneumonias), the American Thoracic Society (ATS) and the European Respiratory Society (ERS) worked together (ERS)[14]. It comprises classifications and definitions for clinical, radiological, and pathological conditions.

A computer-aided detection method for ILD patterns was developed by Y. Xu et al. using volumetric data from MDCT images. The most popular classifiers were Bayesian and SVM [15]. Using a multi-detector computed tomography (MDCT) dataset, Panayiotis D. Korfiatis looked into an automated technique for volumetric estimate of interstitial pneumonia patterns. The vascular tree volume and lung categorization were produced using 3D automated grey level thresholding with edge highlighting wavelet preprocessing. Then, for classification, the 3-class pattern categorization of the LP (lung parenchyma) is used [16]. For the categorization of emphysema in CT images of the lungs, Mardad.J.Ganes proposed a Textron-based classification strategy based on raw pixel representation with radial basis function kernel [17].

Rui Xu developed a method for recognising lung textural patterns using various expressions. The CT and eigenvalues of the Hessain matrices are used in this procedure [18]. For the classification of lung tissue

pattern, Adrien Depeursinge proposed near-affine-invariant texture descriptors derived from isotropic wavelet frames [19].

Jesus Loarce-Martos, Francisco Xavier Leon-Roman, and Sandra Garrote-Corral talked on improvements in home spirometry and quantitative computer tomography for the diagnosis and treatment of interstitial lung illness linked to connective tissue diseases [20]. Aasia Rehman, Dr. Majid Zaman, and Dr. Muheet Ahmed Butt reported a range of medical imaging methods in their survey[21] Panfang Hua's [22] region growth approach is unreliable in high attenuation patterns like ILD, but it performs well in the presence of noise and can correctly distinguish areas with the same properties. Azar Tolouee devised a threshold method that chooses the best threshold to distinguish the lung region from the backdrop [23]. It was discovered to be quite complex for ILD and missing several typical lung features.The fuzzy method is used for segmentation by M.Gomathi and P.Thangaraj [24], which takes less time between iterations to obtain the overall optimal solution but is unable to separate images corrupted by noise, outliers, and other artefacts. The diseases can be easily segmented with improved precision using the registration method for the lungs proposed by I. Sluimer [25]; however, the highest border placement error may occur and is only applicable in clinical practice

3.1. Dataset. Kaggle is a subsidiary of Google Limited Liability Company, which describes itself as "the home of data science." Experts in machine learning and deep learning can get together on the site. In 2010, Anthony Goldbloom and Ben Hamner launched this public data platform. San Francisco, California is the headquarters of the corporation. In a web- based data science environment, data scientists can use the kaggle platform to not only explore but also build models. We used a CT value thresholding technique to first estimate the size of the lungs and to identify normal and mild ILD lung parenchyma, which we then used to segment the lungs. Afterward, we further identified aberrant lung areas with severe ILD using texture-feature images produced from the co-occurrence matrix. Last but not least, we generated the final lung segmentation result by combining the initial lungs with the aberrant lung regions that were discovered.Following preprocessing, the datasets were divided at random into training and test sets, with 80% of the training data used to provide the segmentation models with knowledge about lung segments. Using the remaining 20% of the data, the segmentation models were evaluated. The model is trained using the train folder in order to achieve great efficiency and accuracy. After training the model, the test folder is used to observe the output in predicting the disease of the patient. In order to predict ILD in a patient, we employed the term FVC. The ratio of the lung's forced vital capacity to its forced expiratory volume in the first second is known as the FEV1/FVC. FEV1 is calculated as A x FVC + C, where A is 0.84 and C is 0.23 FVC is calculated by using spirometry, which refers to how much air can be forcibly evacuated from your lungs after taking the deepest inhale possible. A patient's FVC is evaluated at each visit, and over the course of 1-2 years, numerous follow-up visits and a photo at Week 0 are all required. The train dataset contained a complete history of measurements as well as an anonymized computed tomography scan. The only tests included in the bundle are a baseline CT scan and a Forced Vital Capacity assessment. The following variables are included in the dataset: FVC is the measured lung capacity in millilitres; weeks denotes the number of weeks preceding/following the baseline CT; patient is a unique identifier for each patient; and percent is a calculated field that displays the patient's FVC as a percentage of the typical FVC for patients with comparable features.

4. Materials and Methods. For this study, we used a Chest CT-Scan image collection from the kaggle platform. Since CNN models may be trained on X-ray images in DICOM and JPEG formats[26-27], the X-ray images from two different kaggle datasets were first converted to PNG format. Given that the two datasets had different sizes, scaling the images was the next step towards standardisation. U-Net and DenseNet 121 in this study were built with 512 x 512 pixels because the input size for various convolutional architectures vary. Deep learning algorithms have often prioritized one of the two, rather than both usability and performance. A machine learning library called Pytorch shows how these two objectives may coexist by allowing code to be used as a model, streamlining debugging, and being compatible with many well-known computing libraries. It also manages to be effective and support hardware accelerators like GPUs. Performing pulmonary function testing on individuals with ILD is difficult during the present SARS-CoV-2 outbreak[28]. Specialists have advocated spirometry and video consultations at home. In recent years, home spirometry is often used to monitor the severity of ILD. Previous research has shown that daily home spirometry can be an effective method for tracking the progression of ILD. Furthermore, it provides essential clinical data and is clear enough for the majority of

patients to make the diagnosis.

4.1. Exploratory Data Analysis (EDA). An approach to data analysis called exploratory data analysis which uses data visualisation to examine and summarise data sets. It helps us figure out how to effectively manipulate data sources to get the results we're after, making it straightforward to identify trends, assess hypotheses, and confirm assumptions. EDA is frequently used in addition to formal modelling[29] to explore what information data might reveal and to better understand the features of data sets and how they relate to one another. During the data discovery stage, EDA approaches, created in the 1970s by American mathematician John Tukey, are still often utilised. In our experiment, we used the Laplace distribution to analyse the correlation between various variables in the supplied dataset and produced graphs for the same. The analytical functions and approaches that EDA tools could conduct are as follows:

1. Dimensional reduction and grouping techniques make it easier to create graphical representations of multivariate, and high-dimensional data
2. Summary statistics and a univariate visualization[30-31] are provided for each field in the raw dataset,We can assess the relationship between the variables in the dataset and the variable of interest using bivariate visualisations and summary statistics.Multivariate visualisations are used to depict and comprehend the relationships between several data categories.
3. K-means data points are divided into K groups accord ing to how far away from the centroid of each group they are, using the unsupervised machine learning clustering process known as clustering. The data points that are grouped together are those that are closest to a given cen troid. Clustering is used in K-means market segmentation, pattern detection, and image compression.
4. Predictive models, such as linear regression, create predictions about the future using statistics and data.

4.2. Augmentations. A computer vision technique called augmentations enhances the performance of deep convolutional neural net- works. The library is useful for business, deep learning research, and machine learning competitions. Augmentations is a Python package that allows you to make quick and versatile image enhancements. Augmentations provides a simple, yet effective photo augmentation interface for a variety of computer vision applications such as object classification, segmentation, and detection. Augmentations works with a variety of data types, including photographs, medical imaging, satellite imagery, manufacturing and industrial applications, and generative adversarial networks. It operates quickly. Pytorch and Tensor Flow,[33] two deep learning frameworks, are supported.

4.3. Segmentation. Segmentation aids in determining where objects of various kinds appear in an image. Unet is just a convolutional neural network structure which grew from the CNN design[34] with little alterations. It was created to deal with medical photographs with the goal of segmenting or classifying the affected body part.

4.3.1. Unet Architecture. For all the biomedical image segmentation and other image transformation applications, the U-Net is a CNN architecture which is used . Due to its top class performance with less input, U-Net beats other convolutional models in pixel-based image segmentation. Fig.4.1 explains about Unet architecture This unique model was constructed by Olal & Ronneberger et al. Model architecture for U-Net: There are two approaches which are present in U-Net architecture for segmenting biomedical pictures. The initial approach which is also known as encoder shortens.Context is captured by the encoder using a small feature map [35-37]. Max-pooling and convolution layers like Vgg-16 make up the encoder. The other half of the structure is made up of a uniform expanding path, also referred to as a decoder. The second path uses transposed convolution to perform the precise localization. The encoder section is comprised of numerous contraction blocks. The encoder adheres to ConvNet's traditional architecture. For contraction, the network employs a repeated implementation of two 3x 3 convolutions (ReLU) and a 2 x2 max-pooling operation with stride 2. The number of feature channels doubles with each downsampling step. The extensive path consists of upsampling a feature map, a 2 x 2 convolution ("up-convolution") that reduces the number of feature channels in half, concatenation with the equivalent feature map from the skip connection, two 3 x 3 convolutions, and two ReLUs. The component feature vector is mapped using a 1x 1 convolution at the top layer. There are a total of 23 convolutional layers in the network.

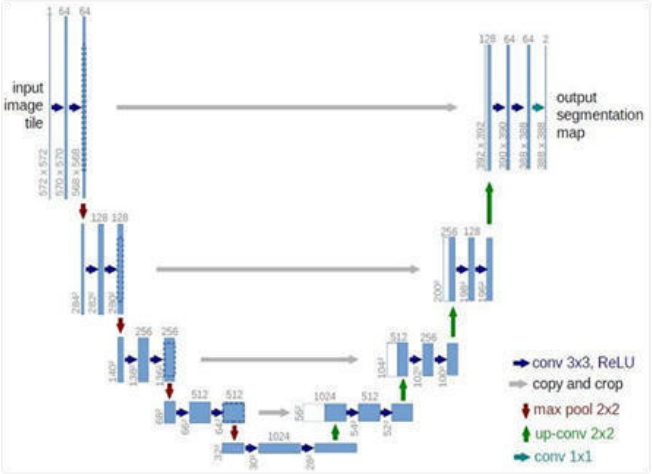


Fig. 4.1: Unet Architecture

4.3.2. Densenet121. As an encoder in our U-Net network, we used relatively simple CNN of DenseNet family.It is made up of four dense blocks with a transition layer between each pair in Fig.4.2. Each dense block has multiple convolution layers, and each transition layer has a batch normalisation, convolution, and average pooling layer. To boost nonlinearity DenseNet employs the activation function of ReLU.

Each layer of the convolutional neural network known as DenseNet shown in Fig. 4.2 is connected to every layer below it. When the composite function operation is applied, the output of the first layer becomes an input for second layer. The convolution layer, pooling layer, batch normalisation layer, and non-linear activation layer make up this composite process [38]. Densenet resolves the vanishing gradient problem, permitting the model to be trained with fewer features [39-40].

The densenet121 is: 5+ (6+12+24+16)*2 = 121
5 – Convolution and pooling layers
3 – Transition layers (6,12,24)
1 – Classification layer (16)
2 – Dense block (1*1 and 3*3 Conv)

4.3.3. Imagenet. The ImageNet project is a sizable visual database that was developed to assist in the creation of software for visual object recognition. ImageNet is a massive image database with about 14 million images in it. The management category organises and labels photos. Machine Learning and Neural Learning Networks use a big database of images to teach machines. To learn useful aspects from these training photos, we need special equipment. As a result, ImageNet features a well-organized management board that makes it suitable for supervised machine learning. ImageNet is used to import the standard weights between the inputs, hidden, and output layers in a neural network. ImageNet is a massive image database with about 14 million images in it. ImageNet is used to import the standard weights between the inputs, hidden, and output layers in a neural network

4.3.4. Learning Rate. The learning rate is the time it takes for the upgrading steps to proceed across the gradient. The Learning rate parameter aids in the updating of the gradient, with a maximum value of 1 but typically a relatively low value. Consider the analogy of a rolling ball to better comprehend the learning rate. The learning rate can be calculated using this analogy by computing the discrete steps that are not continuous.

When it comes to training a network, the Learning rate parameter Fig. 4.3 is quite important. Even when the ball is in the line of sight, there will be little development if a slow learning rate is taken into account. However, if the ball is stuck in the local minima, it has a good chance of not reaching the global minima. The

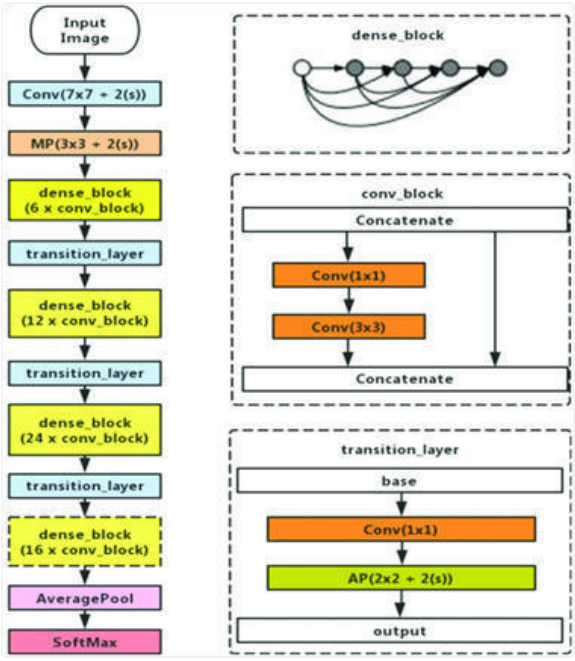


Fig. 4.2: Densenet121 Architecture

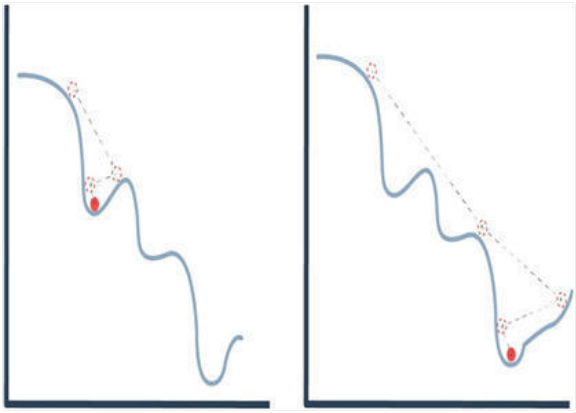


Fig. 4.3: Learning Rate (X-axis: Weight, Y-axis: Loss)

exact optima of the loss function can be attained by adjusting the precise learning rate. There are two ways to accomplish this. There are both greater and smaller steps. For every iteration, when there is a larger difference in weights it is considered as larger steps. In this method, the optimal result can be achieved in very little time, but the exact optimal value cannot be achieved. For every iteration, when there is a smaller difference in weights it is considered as smaller steps. In this method, there is a need for more epochs to get the optimal result, but they are very few chances to miss the exact optimal value

4.3.5. Early Stopping. Keras provides a function EarlyStopping(). It helps us to monitor the epochs and once set off it breaks the training. This callback is works when arguments are intialised. The monitor helps us to state the functioning and monitor and to stop the training. Calculation measurements on the test data

has prefix as $val_{eg} : val_{loss}$ for the loss on the test data. "mode" can be adjusted to either based on the use of metric which is selected if we want to increment then use maximize or max, if we want to decrement then use mini mum or min.

Usually we use minimum validation loss and maximum val_{acc}. The default mode is "auto" which means to decrement the loss or increment the accuracy. Training will get to halt when the selected metric fails to perform better. For knowing which epoch the training has halted, "verbose" is set to 1. This will help us in printing the epoch. At initial sight if there is no improvement, this is not the suitable time for stopping the training as there are some models they get slightly worse or they shows no improvement prior to becoming better. This is performed by "patience" parameter Even if there is a small difference in performance metrics event it is still considered improvement. But if we want to consider it as improvement after a specific improvement that can be done using "min delta" parameter. We can to break the training if metric stays below or above a described verge. This can be done using baseline parameter

4.3.6. ModelCheckpoint. A checkpoint is an intermediary dump of internal state of entire model So when the Early Stopping halts epochs and when metrics are not satisfied even after training the model might not perform with its top accuracy on test data An extra function called Model Check Point () is used to store better ones for the whole observation in training and to use later. This is malleable and it is used in this scenario we will be using it to store better architecture identified throughout the learning. This can be obtained using the pip installer. This function saves model to the respective file. The path and file is mentioned in argument. Monitor argument helps us to monitor the loss function. We mention "mode" as min or max the measurement based on requirement. The mode of standard measure is defaultly set to "auto". We need suitable model throughout the learning. This is achieved by giving save_best_only parameter as True. If we want to know at which epoch model was saved or check the performance measure we set verbose as "1" in callback:

Model Checkpoint('best_lung_unet_densenet121.h5',mode='min', monitor='val_loss', verbose=1).

load_model() is used to load and evaluate the saved model any time.

4.3.7. Proposed model. In addition to providing the final FVC values and their corresponding confidence values, we proposed a Pytorch model for identifying ILD. The results of a spirometer, which measures the amount of air breathed in and out, will be used to assess lung function, and a CT scan of the patient's lungs will be used to estimate the extent of the patient's loss of lung function. All patients' initial confidence intervals are set at 100. The FVC test gauges a patient's capacity to forcefully and swiftly expel air after taking a deep breath. The lung volume is expressed in milliliters. The improvement in FVC over a few weeks is used to forecast the patient's decline in lung function. The results of a spirometer, which measures the amount of air breathed in and out, will be used to assess lung function, and a CT scan of the patient's lungs will be used to estimate the extent of the patient's loss of lung function. All patients' initial confidence intervals are set at 100. The FVC test gauges a patient's capacity to forcefully and swiftly expel air after taking a deep breath. The lung volume is expressed in milliliters. The improvement in FVC over a few weeks is used to forecast the patient's decline in lung function. For a variety of reasons, some individuals may not have improved or not sufficiently responded to treatments. Since their FVC improves steadily and with few peaks and valleys, some people, on either side, are definitely improving. Patients with those traits are incredibly uncommon. The FVC of each patient is assessed once each week to reduce leakage. To evaluate predictions, one modifies the Laplace Log Likelihood. A Confidence measure and an FVC are given to each sample in the test set (standard deviation). For each accurate FVC measurement, we provide a forecast and a level of confidence. As a result, the metric is calculated. Less than 70 confidence points cause a clip.

$$\sigma_{clipped} = \max(\sigma, 70) \tag{4.1}$$

In order to prevent big errors, errors greater than 1000 are also trimmed.

$$\delta = \min(|FVC_{true} - FVC_{predicted}|, 100) \tag{4.2}$$

Metric is defined as:

$$metric = -2\delta - ln(\sqrt{2} * \sigma_{clipped}) \tag{4.3}$$

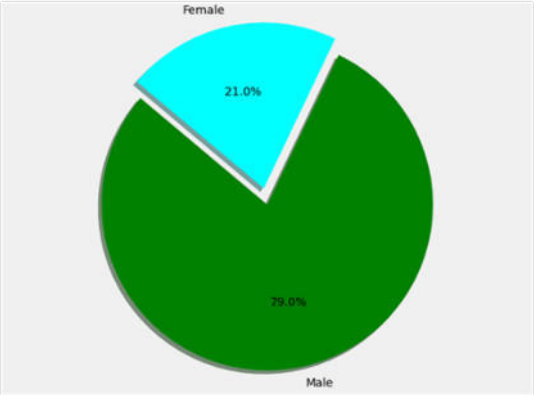


Fig. 5.1: Pie Chart for Gender Distribution

We also used different cmap models to plot CT images of some patients in order to fully understand the extent of their lung function decline. The activation function converts inputs with a given range into outputs. Sigmoid activation functions, for example, take input and map output values between 0 and 1. When enabled, Leaky ReLU provides a non-zero slope to each layer. The activation function's representation is Leaky ReLU: $F(x) = 0.01x, x \geq 0$ and $F(x) = 0, x < 0$. To segment the lung region, a deep learning model using Unet is proposed with provided promising results for the given dataset. For better training of the model, image augmentation were created using the Albumentations and also lung masks are created. Lung Masks are created for 100 CT scans images using the normalization method. The following transformations are applied:

1. Rotate the scan by 30 degrees.
2. Horizontal flip the scan
3. Grid Distortion: It is a method of warping images that relies on mapping between families of equivalent curves that are placed in grid layout
4. Elastic transform: Transform enables us to convert existing Elastic search indices into summarized indices, which provides opportunities for new insights and analytics.
5. Early stopping is also used in the segmentation model to minimize the training time and also to achieve accurate output with sufficient amount of training. The patience value in early stopping class is set to 7. To prevent overfitting during the training phase, early stopping is a type of regularisation. When model performance on a hold out validation dataset stops improving, we can provide an arbitrary large number of training epochs and end training.

Utilizing CUDA, a segmentation model is trained over 80 epochs (Nvidia's CUDA is a parallel computing platform and programming model designed for generic computing on its own GPUs). CUDA enables developers to accelerate computationally demanding programmes by utilising the capacity of GPUs for the parallelizable portion of the calculation. Batch Size (The batch size is a number of samples processed before the model is updated) is 8. We have obtained an accuracy of 92.59 percent for the proposed model after training.

5. Results and Discussion. In this study, two lung segmentation neural network architectures are investigated using two Kaggle datasets. To evaluate the model performance in this work, a dataset of 176 images from these two datasets in Kaggle is used. Fig. 5.1 depicts a pie chart for gender distribution. The chart describes that there are 79 percent of males and 21 percent of females in the given dataset. Fig.5.2 displays a plot of the patients' FVC (measured in millilitres) values using the Laplace distribution. FVC has a maximum of 6399.00, a low of 827.00, mean of 2690.48 and standard deviation of 832.77.

Fig. 5.3 depicts a pie chart for age distribution. According to the pie chart, there are more patients (51.19 percent) in the 60-70 year age group, followed by the 70-80 year age group (31.05 percent). The lowest percentage of smokers (2.78 percent) are over the age of 80. The age group under 60 years old accounts for 14.98 percent of the population. Fig. 5.4 depicts the age of all patients in the dataset using the Laplace distribution.

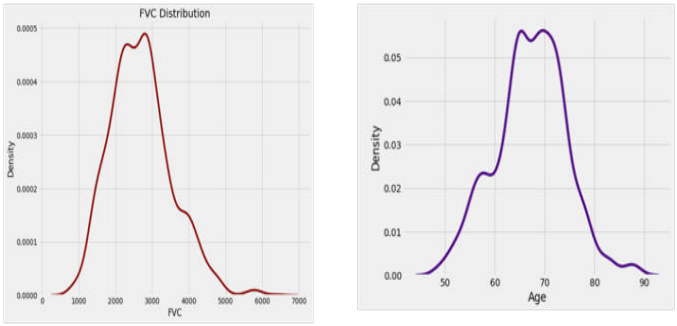


Fig. 5.2: FVC using Laplace distribution

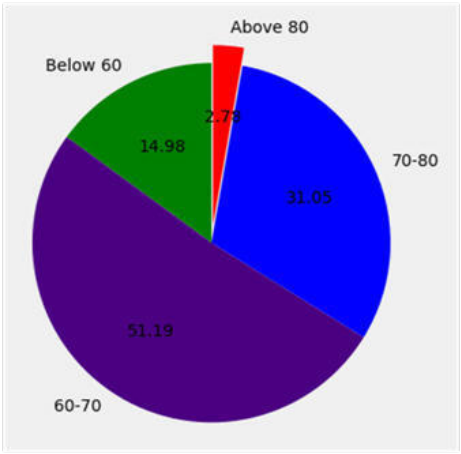


Fig. 5.3: Pie Chart for Age Distribution

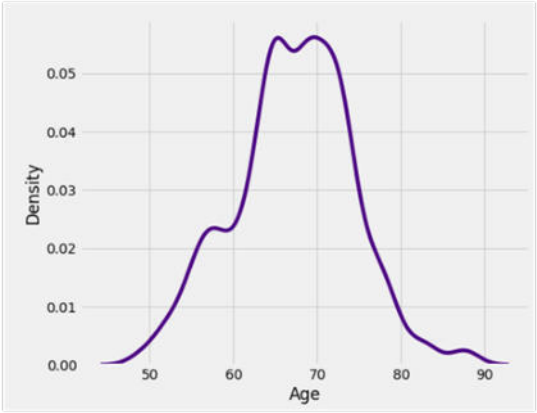


Fig. 5.4: Age based on Smoking Status

The oldest person is 88 years old, the youngest is 49 years old, the average age is 67.19 years, and the standard deviation is 7.06 years.

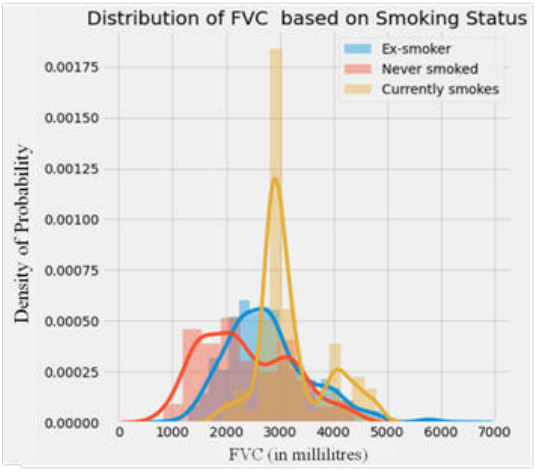


Fig. 5.5: Age plot using Laplace distribution

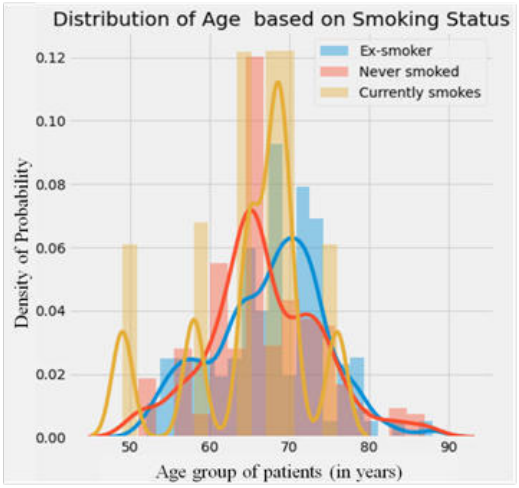


Fig. 5.6: FVC based on Smoking Status

Fig. 5.5 depicts the FVC Laplace distribution of patients and categorises them based on their smoking status (ex-smoker, never smoked, presently smokes). Fig. 5.6 displays the age distribution of patients categorized based on their smoking status (ex-smoker, never smoked, presently smokes) Figures below show the plots of multiple color map models for some patients Fig. 5.7 depicts a color map of patients' CT images in a binary model. It belongs to sequential class of color maps. Sequential color maps are basic color maps that start at a reasonably low lightness value and uniformly increase to a higher value. They are commonly used to represent information that is ordered. Fig.5.8 depicts a color map of patients' CT images in a Hot model. It belongs to sequential class of color maps.

Fig. 5.9 depicts a color map of patients' CT images in a Spectral Model. It belongs to Diverging class of color maps. Unlike sequential color maps, diverging color maps start at their lowest lightness value at both ends and monotonically/linearly increase to the highest lightness value in the center. They could be seen as two sequential color maps combined together, that share the same lightness range and end/begin with the same color. This makes them very useful to represent information that has a critical middle value. Fig. 5.10 depicts a color

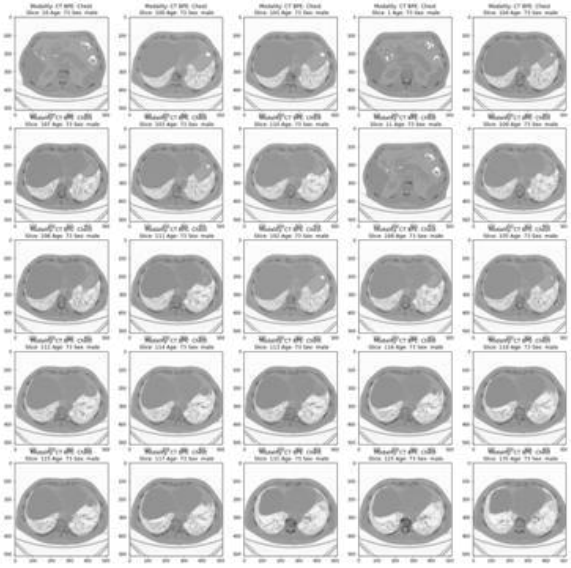


Fig. 5.7: CT scans in Binary Model

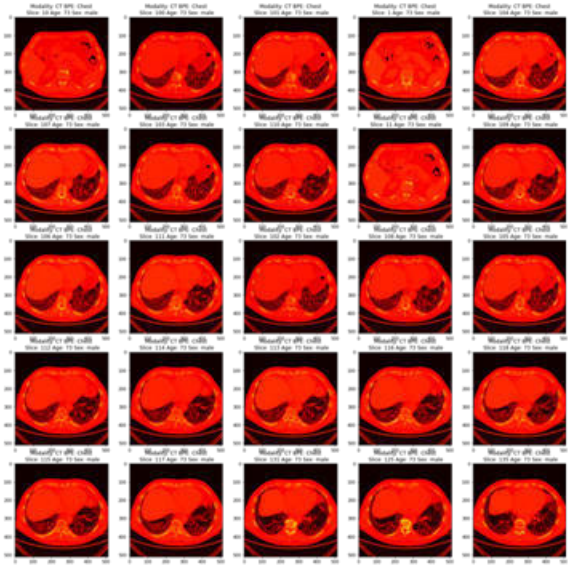


Fig. 5.8: CT scans in Hot Model

map of patients' CT images in a Seismic Model. It belongs to Diverging class of color maps. The correlation matrix is depicted in Fig. 5.10 with the parameters weeks, FVC, percent, and age. Correlation is a statistical technique used by statisticians and data analysts to study the relationships between two numerical variables. A correlation matrix is a collection of correlation coefficients between dataset attributes. A correlation value can be calculated in numerous ways. The correlation matrix is a symmetric matrices with all diagonal equal to one

Fig. 5.11 depicts a colormap of patients' CT images in a Hot model. It belongs to sequential class of colormaps. Fig. 5.12 depicts a colormap of patients' CT images in a Spectral Model. It belongs to Diverging

Fig. 5.9: CT scans in Spectral Model

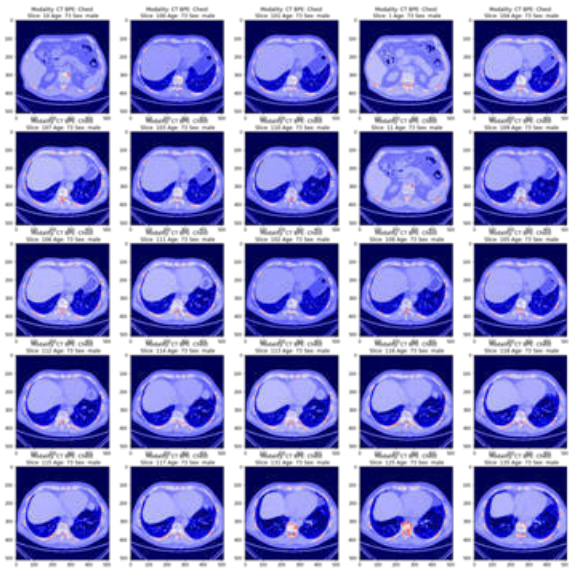


Fig. 5.10: CT scans in Seismic Model

class of colormaps. Unlike sequential colormaps, diverging colormaps start at their lowest lightness value at both ends and monotoni- cally/linearly increase to the highest lightness value in the center. They could be seen as two sequential colormaps combined together, that share the same lightness range and end/begin with the same color. This makes them very useful to represent information that has a critical middle value.

Fig. 5.13 depicts a colormap of CT images from patients in a Seismic Model,belonging to the diverging colormap class. Correlation is used by statisticians and data analysts to investigate the correlations between two numerical variables. A correlation matrix is made up of a collection of correlation coefficients between characteristics in a dataset. A correlation value can be calculated in numerous ways. The correlation matrix is

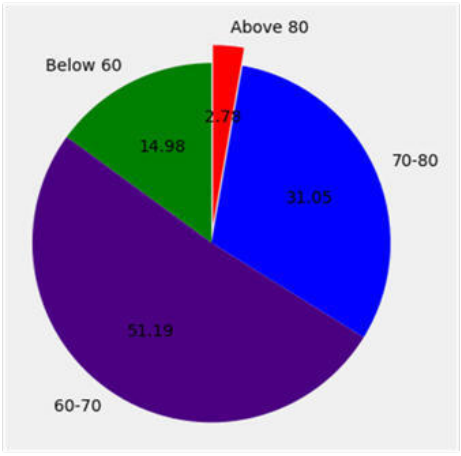


Fig. 5.11: Pie Chart for Age Distribution

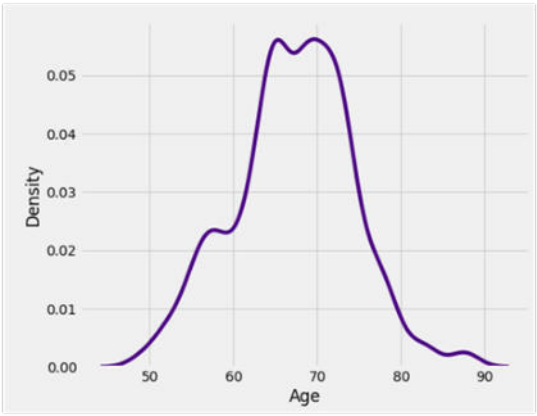


Fig. 5.12: Age plot using Laplace distribution

Table 5.1: FVC final values and patient confidence in the test folder of the presented dataset

Sl.No	Patient Id	FVC	Confidence
1	ID00419637202311204720266	2877.7510	181.3220
2	ID00421637202311550012434	2809.9716	172.7606
3	ID00422637202311677017372	2118.5655	149.2117
4	ID00423637202312137826373	3195.3938	189.5431
5	ID00426637202313170790469	2812.7246	174.9930

a symmetric matrices with all diagonal equal to one.

Fig. 5.14 depicts the correlation matrix with parameters weeks, FVC, percentage and age The Table 5.1 describes the values of FVC and Confidence when the training dataset is fed into the trained model:

The FVC and their corresponding confidence value for the supplied dataset are found using a Python model in this work. Fig. 5.15 depicts the output of the segmentation model.

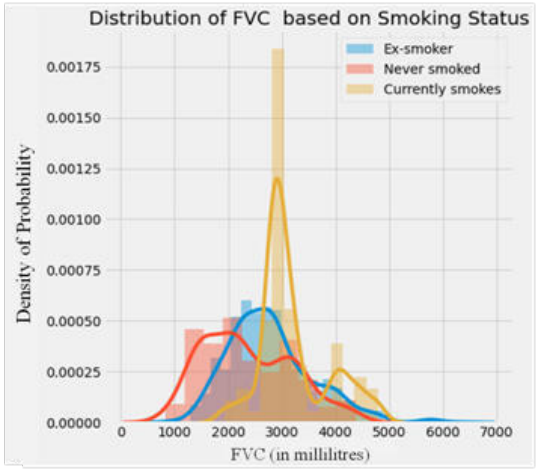


Fig. 5.13: FVC based on Smoking Status

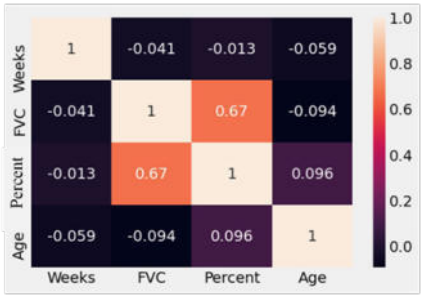


Fig. 5.14: Correlation matrix

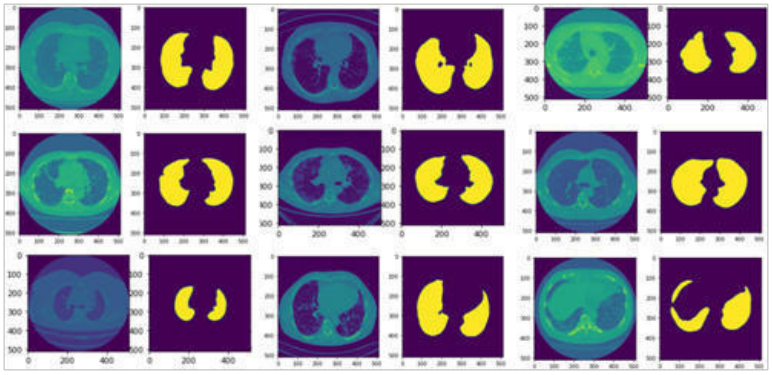


Fig. 5.15: Output of the segmentation model

6. Conclusions and future scope. In this paper, we proposed a systematic approach for ILD segmentation The approach with the cascaded models achieved elegant performance levels in segmentation and also infection quantification, The proposed segmentation technique has proven to be very useful because it can be implemented using machine learning and deep learning techniques for lung segmentation and has garnered

impressive outcomes. The proposed system out performs the existing systems with limited dataset images and producing effective results by attaining 92.5% accuracy.The effectiveness the FVC and their corresponding confidence value for the supplied datasets are determined using a Python model. The stage-wise enhancement of the deep learning algorithm performance has significantly enhanced the performance of the proposed deep learning network-based ILD classifier. Using these findings, we were able to identify whether a person had a condition, allowing for an early diagnosis. When the confidence value is low, the FVC value is measured more effectively. The model network can utilise the crucial data points while suppressing the unimportant ones due to leaky ReLU and activation functions of ReLU. To better evaluate its severity, we plotted the CT image using various color map models. The application was developed using CUDA, a parallel computing environment and programming paradigm that greatly enhances computing performance by utilising the capabilities of the graphics processing unit. To further improve the diagnosis of ILD sickness, the model can be upgraded with a suitable classifier to segment the lung and identify the precise region of the lung that is infected. Deep Learning-based algorithms for fully automatic segmentation have developed as a result of the time-consuming and erroneous nature of manual and semiautomatic segmentation. When lung wall tissue patterns with thick tissue are present, the proposed model performs well.

REFERENCES

[1] Uppaluri, Renuka, et al. "Computer recognition of regional lung disease patterns." American journal of respiratory and critical care medicine 160.2 (1999): 648-654.

[2] Shen, Dinggang, Guorong Wu, and Heung-Il Suk. "Deep learning in medical image analysis." Annual review of biomedical engi- neering 19 (2017): 221-248.

[3] Antoniou, Katerina M., et al. "Interstitial lung disease." Euro- pean Respiratory Review 23.131 (2014): 40-54.

[4] Raj, MR Daniya, and C. Helen Sulochana. "An efficient lung segmentation approach for interstitial lung disease." 2014 International Conference on Circuits, Power and Computing Tech- nologies [ICCPCT-2014]. IEEE, 2014.

[5] Devaki, K., and V. MuraliBhaskaran. "Study of computed tomog- raphy images of the lungs: A survey." 2011 International Con- ference on Recent Trends in Information Technology (ICRTIT). IEEE, 2011.

[6] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." International journal of computer vision 115.3 (2015): 211-252.

[7] LeCun, Yann, et al. "Gradient-based learning applied to docu- ment recognition." Proceedings of the IEEE 86.11 (1998): 2278- 2324.

[8] Greenspan, Hayit, Bram Van Ginneken, and Ronald M. Sum- mers. "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique." IEEE transactions on medical imaging 35.5 (2016): 1153-1159.

[9] Beichel, Reinhard. "Model-Based Segmentation of Pathological Lungs in Volumetric CT Data."

[10] Hamzah, Mohd Faizal Mohd, et al. "Detection of interstitial lung disease using correlation and regression methods on texture mea- sure." 2017 IEEE International Conference on Imaging, Vision and Pattern Recognition (icIVPR). IEEE, 2017.

[11] Mandal, Sampurna, et al. "Prediction analysis of idiopathic pulmonary fibrosis progression from OSIC dataset." 2020 IEEE International conference on computing, power and communica- tion technologies (GUCON). IEEE, 2020.

[12] Hattikatti, Pratiksha. "Texture based interstitial lung disease detection using convolutional neural network." 2017 International Conference on Big Data, IoT and Data Science (BID). IEEE, 2017.

[13] Heitmann, K. R., et al. "Automatic detection of ground glass opacities on lung HRCT using multiple neural networks." European radiology 7.9 (1997): 1463-1472.

[14] Demedts, M., and U. Costabel. "ATS/ERS international multi- disciplinary consensus classification of the idiopathic interstitial pneumonias." (2002): 794-796.

[15] Xu, Ye, et al. "Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM)." Academic radiology 13.8 (2006): 969-978.

[16] Korfiatis, Panayiotis D., et al. "Texture-based identification and characterization of interstitial pneumonia patterns in lung multidetector CT." IEEE transactions on information technology in biomedicine 14.3 (2009): 675-680.

[17] Gangeh, Mehrdad J., et al. "A texton-based approach for the classification of lung parenchyma in CT images." International conference on medical image computing and computer-assisted intervention. Springer, Berlin, Heidelberg, 2010.

[18] Xu, Rui, et al. "Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach." International Conference on Medical Image Com- puting and Computer-Assisted Intervention. Springer, Berlin, Heidelberg, 2011.

[19] Depeursinge, Adrien, et al. "Near-affine-invariant texture learn- ing for lung tissue analysis using isotropic wavelet frames." IEEE Transactions on Information Technology in Biomedicine 16.4 (2012): 665-675.

[20] Loarce-Martos, Jesus, Francisco Xavier Leon-Roman, and San- dra Garrote-Corral. "Recent advances in quantitative computerized tomography and home spirometry for diagnosing and monitoring of interstitial lung disease associated with connective tissue diseases: A narrative review." (2021).

[21] Rehman, Aasia, Muheet Ahmed Butt, and Majid Zaman. "A Survey of Medical Image Analysis Using Deep Learning Ap-

proaches." 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021.
[22] Hua, Panfang. Segmentation of lung tissue in CT images with disease and pathology. Diss. The University of Iowa, 2010.
[23] Tolouee, Azar, et al. "Image based diagnostic aid system for interstitial lung diseases." Expert Systems with Applications 38.6 (2011): 7755-7765.
[24] Gomathi, M., and P. Thangaraj. "A new approach to lung image segmentation using fuzzy possibilistic C-means algorithm." arXiv preprint arXiv:1004.1768 (2010).
[25] Sluimer, Ingrid, Mathias Prokop, and Bram Van Ginneken. "Toward automated segmentation of the pathological lung in CT." IEEE transactions on medical imaging 24.8 (2005): 1025-1038.
[26] Pawar, Swati P., and Sanjay N. Talbar. "Two-Stage Hybrid Approach of Deep Learning Networks for Interstitial Lung Disease Classification." BioMed Research International 2022 (2022).
[27] Reddy, N. Sudhir, and V. Khanaa. "LDDC-Net: Deep Learning Convolutional Neural Network-based lung disease detection and classification." JOURNAL OF ALGEBRAIC STATISTICS 13.1 (2022): 526-542.
[28] Ravi, Vinayakumar, Vasundhara Acharya, and Mamoun Alazab. "A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images." Cluster Computing (2022): 1-23.
[29] Nandy, Sreyankar, et al. "Deep learning-based automated, high-throughput microscopic assessment of interstitial lung disease using endobronchial optical coherence tomography." Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XXVI. SPIE, 2022.
[30] Kihira, Shingo, et al. "U-Net Based Segmentation and Characterization of Gliomas." Cancers 14.18 (2022): 4457.
[31] Agarwal, Mohit, et al. "Eight pruning deep learning models for low storage and high-speed COVID-19 computed tomography lung segmentation and heatmap-based lesion localization: A multicenter study using COVLIAS 2.0." Computers in Biology and Medicine (2022): 105571.
[32] Siddiquee, Sadi Mohammad, et al. "COVID-19 Severity Prediction from Chest X-ray Images using an Anatomy-Aware Deep Learning Model." (2022)
[33] Wang, Risheng, et al. "Medical image segmentation using deep learning: A survey." IET Image Processing 16.5 (2022): 1243-1267.
[34] Walsh, Simon LF, et al. "Deep learning-based outcome prediction in progressive fibrotic lung disease using high-resolution computed tomography." American journal of respiratory and critical care medicine ja (2022).
[35] Ji, Jiarui, et al. "Automated Vein Segmentation from NIR Images Using a Mixer-UNet Model." International Conference on Intelligent Robotics and Applications. Springer, Cham, 2022.
[36] Ghali, Rafik, and Moulay A. Akhloufi. "ARSeg: An Attention RegSeg Architecture for CXR Lung Segmentation." 2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI). IEEE, 2022.
[37] Ahmed, Imran, Abdellah Chehri, and Gwanggil Jeon. "A Sustainable Deep Learning-Based Framework for Automated Segmentation of COVID-19 Infected Regions: Using U-Net with an Attention Mechanism and Boundary Loss Function." Electronics 11.15 (2022): 2296.
[38] Bassi, Pedro RAS. "ISNet: Costless and Implicit Image Segmentation for Deep Classifiers, with Application in COVID-19 Detection." arXiv preprint arXiv:2202.00232 (2022).
[39] Chhabra, Mohit, and Rajneesh Kumar. "A Smart Healthcare System Based on Classifier DenseNet 121 Model to Detect Multiple Diseases." Mobile Radio Communications and 5G Networks. Springer, Singapore, 2022. 297-312.
[40] Inan, Muhammad Sakib Khan, Fahim Irfan Alam, and Rizwan Hasan. "Deep integrated pipeline of segmentation guided classification of breast cancer from ultrasound images." Biomedical Signal Processing and Control 75 (2022): 103553

Edited by: Vinoth Kumar
Received: Aug 23, 2022
Accepted: Nov 7, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 421–438, DOI 10.12694/scpe.v23i4.2010

A DRIVEN MODERN PORTFOLIO THEORY FOR VIRTUAL NETWORK EMBEDDING IN SDN-ENABLED CLOUD

ABDERRAHIM BOUCHAIR,* BELABBAS YAGOUBI † AND SID AHMED MAKHLOUF ‡

Abstract. Network virtualization (NV) has evolved systematically through the urge to share computing resources and improve service deployment in a large-scale environment. Virtual network embedding (VNE) is a well-established technology applied to reinforce the NV process, providing a devoted implementation for a particular case study. In cloud computing, integration of software-defined networking (SDN) has proved to be a practical support to the principal cloud utilities. In return, the SDN-enabled cloud offers innovative deployment techniques for network-based services, which increase the opportunity to efficiently incorporate new network management policies that solve the VNE problem. In this paper, the authors proposed a transition of modern portfolio theory (MPT) into a VNE approach that optimally addresses the selection and ranking of resources in data center networks (DCNs). Results analysis demonstrates the VNE approach's better performance versus similar methods in terms of acceptance ratio, runtime, and substrate resource utilization.

Key words: Cloud computing, Virtual network embedding, Portfolio selection, SDN, Data center network, Allocation policy, Resource importance.

AMS subject classifications. 68T14, 18B15, 90C59, 91G10

1. Introduction. Since its initial launching point, cloud computing kept seeking to proffer as closely as possible the ideal design for high-end resource sharing scalability. Extending on-demand services is the main task of every cloud provider [1]. This is relevant to resource availability placed in physical infrastructure known as the data center (DC). A DCN allows connection between many DC-based servers, guaranteed by offering traffic reliability through efficient network equipment and several virtualization support techniques [2]. The practice of cloud virtualization across DCNs has typically opened new research challenges related to global network performance. Furthermore, various virtualization strategies have been introduced by engaging different types of full or partial virtualization in a collaborative scheme. Thus, VNE derives from finding relationships between virtual networks (VNs) and substrate networks (SNs) that would lead to see the cloud system as a congregate physical structure. The challenge of resource assignment in NV is commonly known as the VNE problem [3]; it primarily consists of mapping virtual resources onto physical devices. Typically, given virtual resources to be embedded are settled as a communication channel of virtual nodes (VNods) interconnected via virtual links (VLins), forming a virtual network request (VNR).

The emergence of SDN back in 2008 has brought outstanding flexibility and robust manageability to traditional networks, resulting from the novel implementation of OpenFlow protocol. Therefore, SDN conception addresses the controllability issue in DCNs by separating different networking stack components into three fundamentally layers: application plane, control plane, and data plane. SDN communications have two divisions crossed at the controller: Firstly, the transfer of information between application and control planes; Afterward, the management servers in the control plane require a pre-set up logic to send instructions to the data plane [4]. The Data plane contains the underlying networking devices like switches and routers to forward the commands.

During every advanced case study, it is inevitable to test future work performances, similar to the prediction routine of stock market behavior in finance. In 1990, Harry Markowitz awarded the Nobel Memorial Prize in

*L.I.O. Laboratory, Department of Computer Science, University of Oran1 Ahmed Ben Bella, P.O. Box 1524 El M'Naouer, Oran, Algeria(bouchair.abderrahim@gmail.com).

†L.I.O. Laboratory, Department of Computer Science, University of Oran1 Ahmed Ben Bella, P.O. Box 1524 El M'Naouer, Oran, Algeria(byagoubi31@gmail.com).

‡L.I.O. Laboratory, Department of Computer Science, University of Oran1 Ahmed Ben Bella, P.O. Box 1524 El M'Naouer, Oran, Algeria(sidahmed.makhlouf@gmail.com).

Economic Science for developing the portfolio choice theory. Markowitz model or portfolio theory (PT) [5] is a mathematical model that considers a portfolio optimization based on two angles: a preferable mean return of a portfolio and an unwanted return denoted as a portfolio risk or variance.

In this work, we attempt to layout a novel node ranking approach for VNE. This challenge adopts MPT analysis as a mapping model for the VNE problem, by resolving these couple significant issues:

(i) What is the optimal selection order of a substrate node (SNod)/link (SLin) for a given VNod/VLin to embed ?

(ii) How to reduce resource utilization rate ?

MPT helps sharpen the supervision quality of the SDN controller and provides a fordable NV running cost by identifying the most suitable substrate resource. In this paper, we explore the most realizable solutions for prior mentioned problematics by introducing the following contributions:

- We propose an analytical solution to solve the VNE problem by projecting the MPT model to adapt to the VNE standards. The MPT model provides a decision support optimization, guaranteed in both SNs and VNRs.
- We suggest using the term *Composite*, a compound element of node and its attached links to facilitate the mapping process in a single stage.
- Employ the obtained composite to implement an effective matching mechanism, which will optimize the resource allocation. This optimization is devoted to overcoming the limitation of MPT liability in dealing with the composite provisioning.
- A set of simulation experiments are conducted to validate the accuracy of the proposed approach.

In the remaining of this paper, we present in Sect. 2, the most associated contributions to our work. Next, an outlined mathematical model of MPT, its projection on our approach, and problem description are presented in Sect. 3. In Sect. 4, we introduce system architecture and the implementation of the proposed VNE policy. Section 5 exhibits the experimental setup and performance evaluation. Finally, a highlighted conclusion is featured along with prospective studies.

2. Related Work. VNE problem remains undoubtedly an NP-hard problem due to the nodes and links constraints [3]. Many studies have drawn much to this matter, specifically those focusing on the issues crossed in cloud computing. The authors in [6] proposed a heuristic VNE algorithm based on defining the topology potential (VNE-TP), by developing a multi-objective optimization approach as an integer linear programming (ILP) model to solve the VNE problem. This novel approach relies on two coordinated stages, regarding node importance calculation and bandwidth function for VLins. The performance evaluations proved an improvement in VNR acceptance ratio and revenue ratio compared with two other similar heuristics. Nevertheless, increments in revenue ratio would generate a lack of VNE accuracy due to the two stages process. In this paper, we consider an ideal VNE process of one coordinated stage mapping, with an enhanced matching mechanism that deals with node and link mapping as one component to be processed.

The work depicted in [7], a metaheuristic-based approach is proposed to solve the VNE problem using a constructive particle swarm optimizer (CPSO-VNE), combined with a step-by-step heuristic scheme dependency. This latter collects the necessary path information from SN, adding more deep processing tasks (matrix calculations with probabilities and discrete vectors). Eventually, additional resource utilization is required to minimize time response, especially for online VNRs. However, in our study, the SDN controller replaces the work of the heuristic solution by offering agile network automation for the MPT measurements, and an ongoing reduced cost while centralizing the necessary operations.

Based on an ILP model, an exact VNE algorithm is proposed in [8] to increase the VNRs acceptance rate through location constraints between the VNRs requirements and the available resources in SN. The result analysis illustrates a better acceptance ratio with 15% more than previous heuristics. Notably, this approach does not support path-splitting, which successfully reduced the harnessing of substrate resources. Alternatively, not adopting a robust network architecture and simply generating random networks using the waxman model [9] does not correspond to the currently updated network technologies. Accordingly, we implemented the most widely-used DCNs to sustain a groundwork reference to a real case scenario in cloud practices.

In [10], the authors got inspired by a biological principle called Yuragi to propose a novel VNE method. Yuragi is a Japanese word which indicates a small disturbance generated extremely and internally. This work's

main idea is to implement a mechanism that adapts organisms formulated as an attractor selection model for software-defined infrastructure (SDI). Performance evaluations illustrate the adaptability of finding a VNE solution under end-to-end delay behaviors, considering several node attributes. The authors have found a lack of access administration and management. Similarly to MPT-VNE, Yuragi-VNE demonstrates a positive application of adopting a solution to the VNE problem from other science fields.

SDN-based techniques promote various solutions to overcome the issues encountered in the rapid scaling infrastructures [11]. In [12], authors proposed a VNE algorithm targeting node selection with the most feasible embedding state, based on clustering coefficient information and node importance for ranking. The authors implemented a non-coordinated two-stage mapping algorithm, which accrues an extensive amount of calculations using the breadth-first search (BFS) algorithm for the node mapping stage and K-shortest algorithm for the link mapping stage. Moreover, the VNE algorithm's final part is based on collecting the outputs of the node-sorting algorithm and the algorithms for each of the two-stages, which occurs a higher time complexity than the Markowitz mean-standard deviation model.

The work accomplished in [13] exemplifies a proper implementation of an abstraction strategy to embed a clustered VNs in edge servers. This approach uses the SDN controller that applies a modified heapsort algorithm to find the pod with the most available resources. Therefore, authors have formulated the resource allocation process as an ILP problem. Experimental results revealed that the proposed work outperformed similar existing algorithms.

Unlike previous research works, our work seeks to maintain a rational background to the real case scenarios in cloud practices. In this work, we try to bypass the high-cost node ranking methods, including massive task implementations and any potential re-embedding process. This is done by employing a simplified node ranking metric inspired by the trading strategy in finance, which will maximize the substrate resource utilization for every mapping request. Thereby, we propose an adapted MPT as a VNE technique related to the investing concept of Markowitz.

3. Modern portfolio theory and problem formulation. This section presents our work's principal sources, including an overview of MPT functionality and problem formulation.

3.1. Modern portfolio theory. The main thrust of MPT is promoting the risk tolerance to build an optimal portfolio of assets with a maximum expected return and low-risk rate. Markowitz first introduces this risk-return relationship published an essay in 1952, and like any portfolio analysis, MPT typically consists of these two steps:

1. **Define the opportunity framework:** A set of particular assets or securities (Portfolio) available to the investor and their properties (mean and standard deviation).
2. **Portfolio selection:** Choose the optimal portfolio that is the most suited to investor requirements.

According to Markowitz, portfolio theory has the following pre-fixed assumptions to make it more applicable:

- Investors' decision (portfolio selection) is rational and based on the mean-variance concepts over a holding period.
- Risk-averse must be the frequent target of all investors.
- Investors focus on maximizing the profit rate for a given level of risk.
- Ability to determine an efficient set of assets or securities in a portfolio to get the most feasible solution (higher return for a given risk).
- All investors are aware beforehand with the sharing probability of return investment.

Based on our perspective setup, we identify the necessary points to achieve a valid projection from the Markowitz model to the VNE problem. Each asset from a given portfolio is characterized by an expected return and expected variance or standard deviation. The portfolio expected return (Ep) in Eq. 3.1 is calculated by multiplying the asset return r with its weight w.

$$Ep = \sum_{i=1}^{N} (w_i \times r_i) \tag{3.1}$$

$$\sum w_i = 1 \tag{3.2}$$

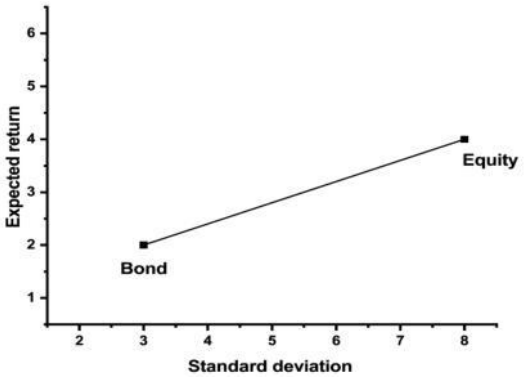


Fig. 3.1: Risk-return trade-off of a portfolio assets

The weight w_i represent the amount of money the investor is willing to spend on the i-th asset. The budget constraint in Eq. 3.2 is the sum of the weights, which must imperatively be equal to 1 or 100% if it is in percentage values (e.g., an investor can divide his capital equally on a portfolio of two assets $a1$ and $a2$ with 50% each).

Throughout the MPT optimization process, the portfolio standard deviation depends on calculating each asset's standard deviation. A high standard deviation portfolio indicates more volatility with the asset returns, and hence it is determined as a highly risky portfolio. Contrariwise, a portfolio with less volatility and more constancy is linked to a low standard deviation rate. The covariance measurement is also included as a diversification tool that can be used to add the assets that reveal negative covariance to reduce the overall volatility of a portfolio and ensure the risk-averse assumption. A positive covariance implies that two securities evolve in tandem; on another side, the covariance is negative when they evolve inversely. The portfolio assets can be displayed across a two-dimensional plane.

The risk-return relationship follows the bond asset's traced line representing low-risk low-return until the Equity asset, which indicates high-risk high-return. From Fig. 3.1, we can deduce that if an investor seeks a high return, he will be more likely to obtain an asset with a higher risk. MPT evaluates the investment impact on the overall portfolio risk and reward. However, our proposed method mainly focused on the assets management.

3.2. Analogy Context. VNE deals with the assignment of interconnected VNods that forms a VNR. Multiple VNods from a single VNR cannot be allocated to the same SNod, while a VLin can be mapped onto several SLins by creating a logical path through SNod, which are considered as intermediaries between the source and the destination. Besides, a single SNod can host several VNods from different VNRs. Based on this VNE standardization, we have found driven indigency to adopt a mutual representation for SNs and VNs elements denoted as Composites. A Composite exemplifies a collected node and its attached links in one element.

From Fig. 3.2, we can observe that the VLin between VNods E and F is passed through a SC that is considered as a hidden hop. Hence, VLin was embedded in two SLins. Composites have the role of simplifying the employment of the coordinated one-stage mapping in our VNE method. The primary concern is to set up MPT as a VNE-oriented strategy. From a VNE perspective, a DCN with its available resources represents a portfolio of N assets or shares (i.e., every substrate or virtual composite (SC or VC) has an asset as an equivalent). In order to apply the Markowitz mean-standard deviation model, we define the importance-standard deviation model as a projection of MPT optimization into a VNE problem, as shown in Table 3.1.

3.2.1. Composite importance. Fundamentally, a node's importance is determined according to its role to satisfy the global objective in a network-based approach. In our VNE method, we define the composite importance (Imp_c) in Eq. 3.3 based on the composite capacity (C_{cap}). This latter corresponds to the return

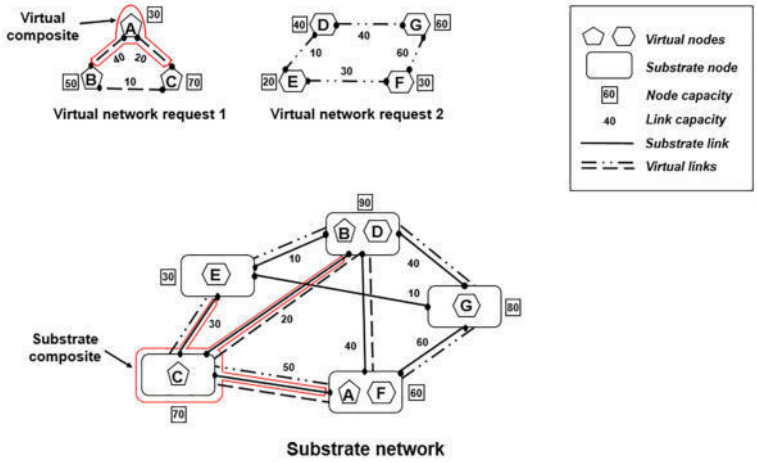


Fig. 3.2: Composites

Table 3.1: Similarities between MPT and VNE

MPT	VNE
Asset	Composite
weight	latency or bandwidth
Portfolio	DCN
Asset capital (money)	Composite capacity
Expected portfolio	Importance
Investing	Embedding or allocation

measurement of a single asset adapted from Eq. 3.1.

$$Imp_c = w \times C_{cap} \tag{3.3}$$

where w is the composite weight depends on the dedicated mean latency or bandwidth for each attached link. In the case of a SC, the weight is related to the latency property of a physical server's attached links.

$$SC_w = \frac{1}{\sum la_i} \tag{3.4}$$

where $\sum la_i$ is the total latency of the attached subtract links in the i-th SC. The reason behind choosing latency as a composite weight returns to its critical impact on the DCN connectivity. Thereby, a lower latency evinces the priority of a given composite. We note VC_w as VC weight related to the attached links' bandwidth of a virtual machine (VM).

$$VC_w = \frac{\sum Bw_j}{\sum Bw_{VN}} \tag{3.5}$$

where $\sum Bw_{VN}$ the total bandwidth of a VN, and $\sum Bw_j$ denotes the sum of the allocated bandwidth for the attached links in the j-th VC. In our approach, the latency is predefined during the creation of a DCN, yet it cannot be specified in case of a VNR until the VN is running, which left us only to consider VLin bandwidth property as its exact weight.

The return factor in MPT is settled on the amount of money granting for each asset, while the expected return is the total money the investor will probably earn after purchasing some assets from a stock market. However, from a resource allocation perspective, VNE deals with the VNR computational requirements on the

one side and with SN's ability to assure the physical resources' availability and reliability on the other side. In light of this, we consider the relative SC capacity to the total network capacity as the main profit for both service providers (SPs) and cloud users.

$$RSC_{cap} = \frac{\sum SC_{cap}}{\sum SN_{cap}} \tag{3.6}$$

where $\sum SC_{cap}$ the total SN capacity. $\sum SN_{cap}$ denotes the sum capacity of the local resources of a physical server:

$$SC_{cap} = \sum(CPU, RAM, Storage, Bw) \tag{3.7}$$

In case of a VNR, the relative VC capacity is calculated in this way:

$$RVC_{cap} = \frac{\sum VC_{cap}}{\sum VN_{cap}} \tag{3.8}$$

where $\sum VN_{cap}$ the total of capacity of VN. $\sum VC_{cap}$ denotes the sum capacity of the local resources of a VM:

$$VC_{cap} = \sum(CPU, RAM, Size, Bw) \tag{3.9}$$

We note that the final value of any composite capacity is normalized using the following feature scaling in Eq. 3.10:

$$FC_{cap} = \frac{C_{cap} - MinC_{cap}}{MaxC_{cap} - MinC_{cap}} \tag{3.10}$$

where $MaxC_{cap}$ and $MinC_{cap}$ denote the maximum and minimum composite resource capacity, respectively, from a set of composites in network. Since C_{cap} features the return aspect from MPT; FC_{cap} values are between 0 and 1, which simplifies the data sorting.

3.2.2. Composite standard deviation. Principally, the standard deviation of a composite σ_c is the square root of the composite variance σ^2.

$$\sigma_c = \sqrt{\sigma_C^2} \tag{3.11}$$

$$\sigma_C^2 = \frac{(FC_{cap} - Imp_C)^2}{K} \tag{3.12}$$

We note that K is related to the VNR size V and SN size S. The variance of VC is calculated in the following way:

$$\sigma_{VC}^2 = \frac{(FC_{cap} - Imp_{VC})^2}{V} \tag{3.13}$$

In a VNE process, a VNR must entirely be mapped (whole population), so in (13), we take all the VNR size. For the SC, when S equals V, it means that every composite in both networks is involved, thereby, K equals the same value as S and V. However, when V is less than S, it implies that only V servers from SN server size are required for the VNE process. Therefore the SC variance is calculated in the following way:

$$\sigma_{SC}^2 = \frac{(FC_{cap} - Imp_{SC})^2}{S - 1} \tag{3.14}$$

K equals to $S - 1$, which indicates that the VNE process is established with just a sample of servers from SN.

3.3. Problem formulation. In finance, MPT is employed to minimize the risk of investment; therefore, the expected return is maximized. However, in this paper, MPT is presented as a heuristic approach to combine resource importance and standard deviations. Specifically, MPT is established by focusing on one of two goals: First, maximizing the expected return is predicated on a certain risk level. Otherwise, an investor sets a desirable expected return and tries to adjust its associated risk by minimizing the mean assets' standard deviation. The second goal is infeasible due to the certainty of dropping the SCs capacity and an evident constant VCs capacity. Besides, in a VNE approach, it is possible to maximize the substrate resource capacity by reinforcing the DCN with additional high-capacity resources; however, achieving such a heavy task is a costly solution.

The first goal is distinguished as a significantly resource-consuming pattern; it is unreasonable to maximize composite importance. Hence, we consider minimizing the substrate resource capacity by adjusting the Markowitz mean-variance model to suit our MPT-VNE approach. Accordingly, we constructed the composite importance model.

$$\begin{aligned} &Find && OSC = (x_1, ..., x_n), OVC = (y_1, ..., y_m) \\ &So\ as\ to \\ &Min\ [F_{Imp}(x)] \\ &Subject\ to \\ &\qquad Imp_{SC}(x_i) \geq Imp_{VC}(y_j);\ x \in R^n;\ y \in L^n \\ &\qquad Imp_{SC}(x_i) \geq 0, Imp_{VC}(y_j) > 0 \\ &\qquad \sum w_{SC} = \sum w_{VC} = 1 \end{aligned} \tag{3.15}$$

where:
OSC: an optimal set of SCs | OVC: an optimal set of VCs
L: a set of available VSs | R: a set of available SCs
Imp_{VC}: importance value of a VC | Imp_{SC}: importance value of a SC
$F_{Imp}(x)$: importance minimization function

The optimality of VC mapping lies in the perfect order of every composite in VNR. Thus, our MPT-VNE method seeks to maximize the utilization of local resources in every SC. This yields to a periodic reduction in SCs importance values.

4. Proposed work. Some research has viewed SDN as part of the network virtualization technology landscape due to its heavy use of advanced software solutions. Hence, it will lead to lower expenditure on expensive equipment and centralized management of resources.

4.1. System architecture. The employment of SDN functionalities into cloud infrastructure has elevated DCN management by deploying an efficient resource virtualization process. This latter creates VNs based on resources capacity, availability, and preferred policy of both SPs and end-users.

Accordingly, a real case scenario to our approach would be a SP receiving renting requests of physical resources. By accepting this request, the SP tries to manage the available substrate resources by creating VNs that need to be embedded in the most optimized way to meet his business planes.

From Fig. 4.1, we can highlight four main entities of our design proposal:

1. **Cloud User:** Represent the external source of VNRs that triggers the internal process of VNE.
2. **Service Provider:** Exemplifies the mediator ground between the cloud user and the available cloud services, where it can determine his requirements and make a decision based on the current offer. In our case, this entity can demonstrate the application plane utility thanks to the SDN broker's role of re-ordering the VNRs received from the least-loaded to the most-loaded and then pass them into a verification function to decide if they are apt for further embedding evaluation.
3. **Control Plane:** It has the primary job of maintaining the running services in DCN and informing the SDN broker about free resources. These tasks are managed by the SDN controller, which is typically a server in charge of featuring cloud monitoring, resource control, VN management, and the overall

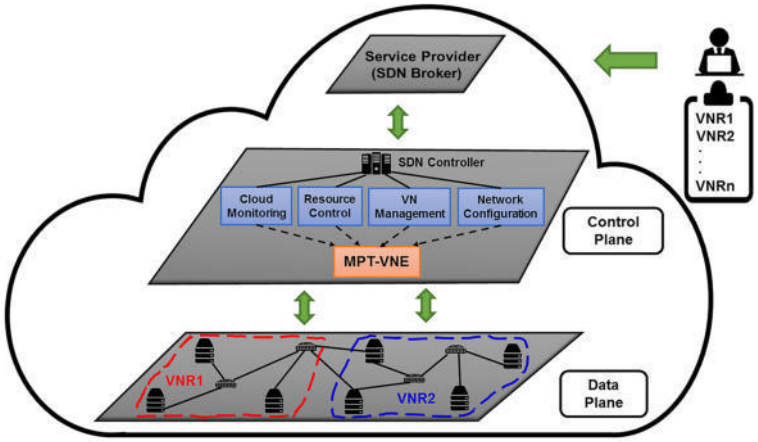


Fig. 4.1: Overview of the proposed SDN-enabled cloud system architecture

network configuration. Furthermore, SDN controller accommodates our core logic designated as MPT-VNE policy by gathering the necessary information in advance like any other heuristic method.

4. **Data Plane (Forwarding Plane):** Contains the underlying types of networking equipment that communicate with the SDN controller by sending network states and forwarding commands to servers. Precisely, to simulate the practicality of this plane, we implemented multiple types of DCNs.

Our system policy applies an offline VNE algorithm (preset VNRs) that depends on the SDN controller to manage the accepted VNRs by the SDN broker in a custom scheduling plan. The initial VNR verification in the application plane does not enable handling the VNRs mapping in real-time, although no remapping procedure is needed afterward.

4.2. MPT-VNE implementation. As VNE techniques rely on VN exigency and the SN's resource availability, our importance-aware policy features the optimal resources in DCNs. Moreover, it determines the most valuable VC in a given VNR (MPT optimizes the correlation pattern between composites). MPT-VNE is originated from two phases, including ranking and mapping.

4.2.1. Ranking of composites. Intuitively, a particular composite rank is calculated based on the node CPU capacity and the combined bandwidth of its attached links. However, since we are dealing with a mapping procedure of VMs, we applied the Eqs. 3.7 and 3.9 to collect the resource requirements that guarantee the running of every VM on a given server.

Algorithm 1 produces four lists of importance and standard deviation for SCs and VSc, which are the essential inputs for ranking and mapping processes. Then, we establish a simple *ranking procedure* to generate two ranked lists $< SC_r >$ and $< VC_r >$. These lists contain the composite ID in conformity with the pre-classified lists of the composites standard deviations.

4.2.2. Mapping Process. In our work, focusing on manipulating the standard deviation of composites capacity is a ruled out objective. Therefore, the standard deviation measurements are only performed for matching purposes between VSs and SCs. We consolidate this matter by engaging the SDN broker to execute a verification function ($VNR\_verify()$) for every VNR. Indeed, this function will return $false$ if the number of VMs is bigger than the number of servers (according to VNE standardization), or the VNR requirements are higher than the DCN capacity.

$$Imp_{SN} = \sum_{i=1}^{N} (Imp_{SC_i}) \tag{4.1}$$

Algorithm 1 Computing of importance and standard deviation

Input: VNR, SN.
Output: $< Imp\_SC >$, $< Std\_SC >$, $< Imp\_VC >$, $< Std\_VC >$.

1: **for** $i \leftarrow < SC > size$ **do**
2: **Compute** $< Imp\_SC > (i)$; ▷ Eqs. 3.4, 3.5, 3.6 and 3.7
3: **if** $< SC > size$ **equal** $< VC > size$ **then**
4: **Compute** $< Var\_SC > (i)$; ▷ Eq. 3.12
5: **Compute** $< Std\_SC > (i)$; ▷ Eq. 3.11
6: **end if**
7: **if** $< SC > size$ **greater than** $< VC > size$ **then**
8: **Compute** $< Var\_SC > (i)$; ▷ Eq. 3.14
9: **Compute** $< Std\_SC > (i)$; ▷ Eq. 3.11
10: **end if**
11: **end for**
12: **for** $i \leftarrow < VC > size$ **do**
13: **Compute** $< Imp\_VC > (i)$; ▷ Eqs. 3.5, 3.8 and 3.9
14: **Compute** $< Var\_VC > (i)$; ▷ Eq. 3.13
15: **Compute** $< Std\_VC > (i)$; ▷ Eq. 3.11
16: **end for**

$$Imp_{VN} = \sum_{j=1}^{N} (Imp_{VC_j}) \tag{4.2}$$

The importance of SN and VN are calculated in Eq. 4.1 and 4.2, respectively. A VNR is accepted for further mapping evaluation when the value of Imp_{SN} is higher than Imp_{VN}. A decisive part of our matching policy exists in the mapping process, emphasizing the important values to reinforce the network management.

For every investment operation, MPT provides a logical assets exhibition beginning with the bond-asset and finishing in the equity-asset. With that in mind, in Algorithm 2, we already sorted the composites from lower to higher in terms of standard deviation values, which keeps a coherence affiliation between composites in the matching procedure. Nevertheless, we need also to review the importance values for every composite and proceed with an accurate bandwidth checking. A VNR is scheduled for further processing if the current available bandwidth for SC cannot satisfy the required bandwidth of the compared VC. If the size of $< SC_r >$ is greater than $< VC_r >$ size, we call the *Search_map*() function at line 20, which is a simple iterative function that iterates through $< Imp_{VC} >$ and $< Imp_{SC} >$ to locate the most suitable SC. An optimal SC list $< OP\_SC >$ is returned when the embedding is accomplished.

4.3. General process of MPT-VNE. Typically, every VNR is accompanied by several preferences related to stationary attributes submitted with the list $< workloads >$ in the inputs.

The overall process of MPT-VNE presented in Algorithm 3 is executed through all the SDN layers. Firstly, it is established at the application plane by verifying the received VNRs on line 2. Secondly, it takes place at the control plane and the data plane by encompassing all the previous algorithms while covering the necessary exchange of information. The proposed MPT-VNE is developed by employing a network virtualization strategy (a greedy heuristic method) as a part of the SDN system design. To formalize this work, we acknowledge the following backgrounds:

- We note that our policy does not consider or apply a fault tolerance procedure. Thus, we assume that all the physical equipment in the DCN performs regularly, and no failure can occur, whether at VN or SN levels.
- The integrated matching mechanism enables the node/link mapping function to be treated as a single composite mapping to reduce the execution time.
- We emphasize that our VNE version deals only with static substrate resources in a DCN with redundant VNRs management and does not establish any virtual resource migration.

Algorithm 2 Mapping process

Input: $< SC_r >$, $< VC_r >$, $< Imp\_SC >$,$< Imp\_VC >$.
Output: $< OP\_SC >$

1: **Sort** $< Imp\_SC >$, $< Imp\_VC >$;
2: **if** $< SC_r >$ **equal** $< VC_r >$ **then**
3: **for** $i \leftarrow < VC_r > size$ **do**
4: **if** $< Imp\_VC > (i)$ **less or equal** $< Imp\_SC > (i)$ **then**
5: **if** $< VC_{r_{(BW)}} > (i)$ **greater than** $< SC_{r_{(BW)}} > (i)$ **then**
6: *Schedule_map*(); ▷ Returns false after the timeout has completed
7: **if** *Schedule_map*() **returns** *false* **then**
8: **Print** "Mapping failed";
9: **end if**
10: **else**
11: **Embed** $< VC_r > (i)$ **into** $< SC_r > (i)$;
12: **Add** $< VC_r > (i)$ **to** $< OP\_SC >$;
13: **end if**
14: **else**
15: **Print** "Mapping failed";
16: **end if**
17: **end for**
18: **end if**
19: **if** $< SC_r > size$ **greater than** $< VC_r > size$ **then**
20: *Search_map*(); ▷ Returns false if no fitted SC is found
21: **if** *Search_map*() **returns** *false* **then**
22: **Print** "Mapping failed";
23: **else**
24: Apply instructions from line 3 to 17;
25: **end if**
26: **end if**
27: **Return** $< OP\_SC >$;

Algorithm 3 An overview of the MPT-VNE process

Input: SN , $< VNRs >$, $< workloads >$.
Output: $< OP\_SC >$

1: **for all** $VNR \in < VNRs >$ **do**
2: **if** $VNR\_verify()$ **equal** $true$ **then**
3: Generate the required Data using Algorithm 1;
4: Rank the SCs and VCs and apply the mapping process using Algorithm 2;
5: **else**
6: **Print** "Mapping failed";
7: **end if**
8: **end for**

5. Performance evaluation. In this section, we investigate the performance of our VNE approach through empirical simulation evaluations. These are performed via two types of DCNs, switch-centric DCN (SwCDCNs) and server-centric DCN (SvCDCNs), to provide a comparison between MPT-VNE and other importance-based VNE approaches. Also, we provide a general behavior analysis between the mentioned DCNs.

Table 5.1: Simulation parameters for VNR

VNR design characteristics	Requirement
VM Size	100MB – 3 GB
VM RAM	256MB – 4 GB
VM Bandwidth	100MBps – 10 GBps
VM number of CPUs	1 – 4
Link Bandwidth	100MBps – 5 GBps
Number of VMs	4 – 120

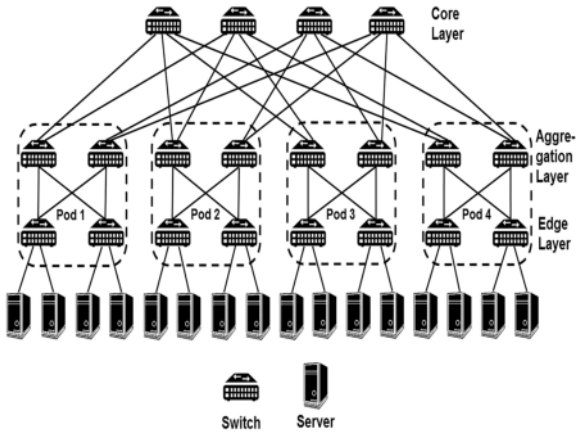


Fig. 5.1: Fat-tree topology with 4 pods

5.1. Simulation setting. In our simulation context, the SDN controller is a logical orchestration entity referred to as a network operating system (NOS) based on the predefined implementation of CloudSimSDN [14]. For modeling SDN-enabled cloud data centers. Throughout the simulations, we used a computer with an Intel Core i5-6200U processor up to 2.8 GHz and 4 GB of RAM.

5.1.1. VNRs setup. Besides the NOS class, CloudSimSDN has two other relevant classes, regarding physical topology generator and virtual topology generator to implement VNRs and DCNs in JSON files. We have generated 23 VNRs where each VNR has been processed for potential mapping in every DCN. A VNR sample contains a list of customized VMs and VLins. Table 5.1 exhibits the possible capacity requirements in the VNRs list, where a VNR can include a minimum of 4 VMs or up to 120 VMs:

We note that once the VNR is generated, an extra output is given regarding its specific workload. The VNR workload will be saved in an excel file and passed in inputs, as shown in Algorithm 3.

5.1.2. DCNs Setup. SwCDCNs rely basically on switches for forwarding the network packets. We focus on the tree-like networks regarding Fat-tree [15], VL2 [16] and Diamond [17], Figs. 5.1, 5.2, and 5.3 exemplify a sample of the implemented SwCDCNs with 16 servers in each.

A SwCDCN instance can contain a minimum of 16 servers or up to 128 servers with a maximum of 0.2 seconds and 12 GB per second of latency and bandwidth, respectively, on the attached links. Further configuration details are available identically in Table 5.2.

In SvCDCNs, servers play the role of a switch (responsible for network routing). They usually consist of programmable servers connected directly with each other and with few low-cost mini-switches. We have selected three SvCDCNs regarding DCell [18], Ficonn [19] and BCube [20]. Figs. 5.4, 5.5, and 5.6 depict a SvCDCNs representation of a level 1 DCell, level 2 Ficonn, and level 1 BCube, respectively. SvCDCNs simulation setup is presented in Table 5.3, where for each instance, we can have up to 20 levels with less maximum latency (0.01 seconds) compared with the SwCDCNs.

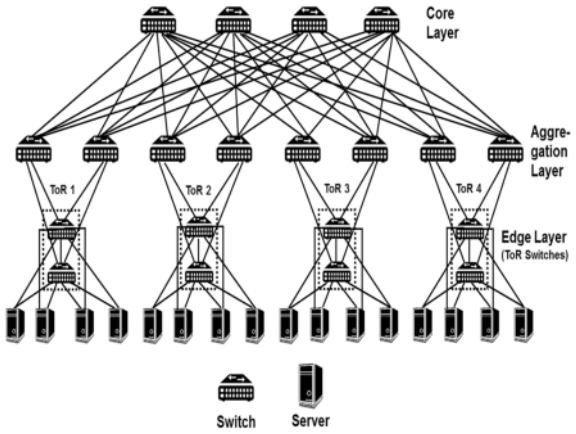


Fig. 5.2: VL2 topology with 4 ToRs

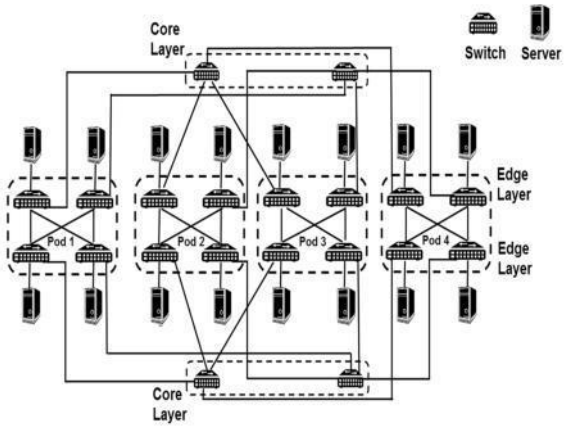


Fig. 5.3: Diamond topology with 4 pods

Table 5.2: Simulation parameters for SwCDCNs

SwCDCN structure characteristics	**Load values**
Number of pods/ ToRs	4 – 8
Number of servers	16 – 128
Number of CPUs	8 – 16
Server RAM	10 – 16 GB
Server storage	10 – 50 GB
Link bandwidth	10 – 12 GB
Link latency	0.008 – 0.2 seconds

We note that depending on the type of DCN; a new DCN instance is implemented of the same DCN based on the number of pods or levels (e.g., we have 2 instances of Fat-tree if we considered to generate it twice, one instance with 4 pods and the second with 8 pods). In our work, the network traffic relies on the forwarding rules generated in the SDN controller, based on the workload size of each VNR. Therefore, we established a logical connection between the SDN controller and the switches in the core layer and all the mini switches in the SwCDCNs and SvCDCNs, respectively.

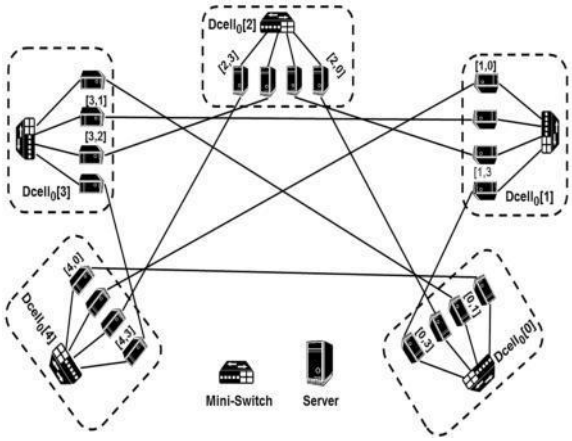


Fig. 5.4: Level 1 DCell topology

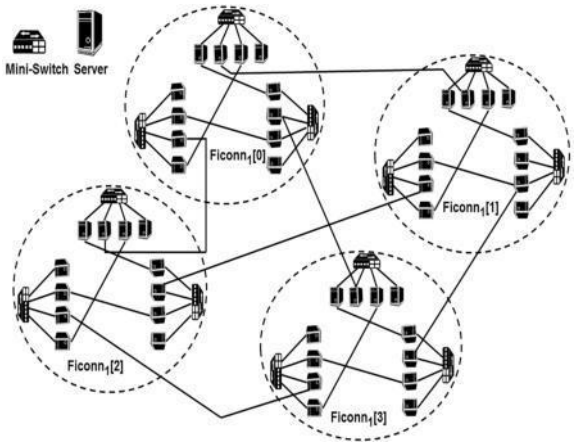


Fig. 5.5: Level 2 Ficonn topology

Table 5.3: Simulation parameters for SvCDCNs

SvCDCN structure characteristics	**Load values**
Number of levels	1 – 20
Number of servers	12 – 128
Number of CPUs	12 – 20
Server RAM	16 – 20 GB
Server storage	10 – 50 GB
Link bandwidth	10 – 16 GB
Link latency	0.001 – 0.01 seconds

5.2. Evaluation methods and metrics. To validate our contribution, we have conducted comparative simulations between the proposed MPT-VNE approach and four similar VNE algorithms, which mainly considers resource capacities as the influencing factor in the mapping process's decisive phase. These approaches are listed in Table 5.4.

In order to keep the comparison in the same context, we modified specific algorithms (NRM, Least-load, and NDC) to assure the embedding is performed in one stage (i.e., mapping VNods and VLins at the same

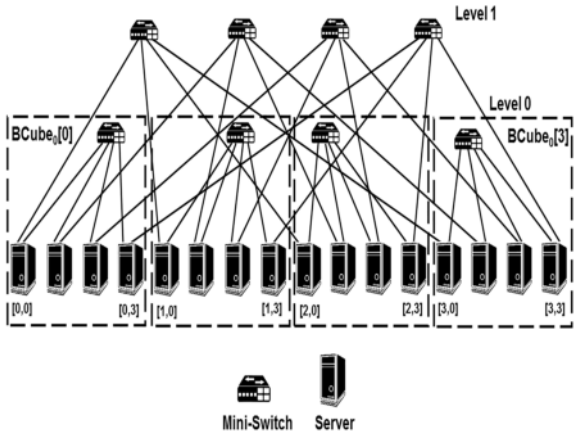


Fig. 5.6: Level 1 BCube topology

Table 5.4: Elected algorithms for results comparison

Importance-based Algorithms	Description
RW-BFS-VNE	A one-stage topology-aware heuristicbased on a breadth-first searching using the random walk model to calculate node ranking values [21].
NRM-SP-VNE	A modified node ranking heuristic to support a one-stage embedding using a short path algorithm [22].
Least-SP-VNE	Network method that treats first the nodes with less stress and connectivity merged with the short path algorithm to perform VNE at one coordinated stage. A similar version is already integrated into CloudSimSDN.
NDC-SP-VNE	An enhanced algorithm to determine the node influence by employing the node degree characteristics. Since it is a two-stage VNE algorithm, we added the short path algorithm to preserve the VNE evaluation at the same stage [23].

time) by associating these algorithms with an algorithm that targets the subtract links with more bandwidth capacity. Thereby, VNods and VLins are mapped simultaneously.

To evaluating our work, we focus on comparing the final results with the existing approaches based on the following performance metrics:

1. **Average acceptance ratio:** We denote AR the acceptance ratio of a VNR in a DCN instance, which indicates the accepted VNRs VNR_{apt} from the total received VNR_{rcv} in a given time range t.

$$AR = \lim_{x \to -\infty} \left(\frac{\sum_{t=0}^{T} VNR_{apt}}{\sum_{t=0}^{T} VNR_{rcv}} \right) \tag{5.1}$$

In our VNE model, the accepted VNRs used in AR are already mapped to distinguish them from the initially accepted VNRs in the application plane. The Average acceptance ratio is defined as the following:

$$AR_{avg} = \frac{\sum_{i=1}^{N} AR_i}{\sum_{j=1}^{N} DCN_j} \tag{5.2}$$

Where DCN_j denotes a single DCN instance of the same SN.

2. **Average runtime:** Handled by CloudSimSDN, it implies the execution lifetime of a given approach, including the algorithm mapping processing added to the traffic transmission period.

$$DRun_{avg} = \frac{\sum_{i=1}^{N} DRun_i}{\sum_{j=1}^{N} DCN_j} \tag{5.3}$$

Where $DRun$ is the runtime of a single DCN instance.

3. **Average rate of SNods and SLins utilization:** This metric relies on $< OP\_SC >$ and the selected switches to identify the rate of local resource utilization for every accepted VNR. The utilization rate of nodes NUR and links LUR in a single DCN instance are calculated as follows:

$$NUR = \left(\frac{\sum_{VNR=1}^{N} SNods_{sel}}{Total\_SNods} \right) \times 100 \tag{5.4}$$

$$LUR = \left(\frac{\sum_{VNR=1}^{N} SLins_{sel}}{Total\_SLins} \right) \times 100 \tag{5.5}$$

where $SLins_{sel}$ and $SLins_{sel}$ respectively denote the sum of selected nodes (hosts and switches) and links, which are used repeatedly in a set of accepted VNRs. The average utilization rates are defined as follows:

$$NUR_{avg} = \frac{\sum_{i=1}^{N} NUR_i}{\sum_{j=1}^{N} DCN_j} \tag{5.6}$$

$$LUR_{avg} = \frac{\sum_{i=1}^{N} LUR_i}{\sum_{j=1}^{N} DCN_j} \tag{5.7}$$

5.3. Results analysis. The initial results review revealed a similar performance between the selected DCNs within the same DCN type, due to the primary objective of the methods in comparison (targeting the most critical resources). Therefore, we preferably present results of different DCNs by only their type.

Figure 5.7 illustrates the average ratio of the accepted VNRs relative to a set of VNE methods. We note that regardless of the DCN type, our VNE approach produces a higher acceptance ratio due to the adapted MPT measurement of the composite importance. The MPT-VNE average acceptance ratio is reduced by 22% within the SvCDCNs, resulting from the additional task of traffic forwarding performed by the servers, while this task is only performed by switches in the case of SwCDCNs. Moreover, SvCDCNs generates more traffic density to the server attached links, which will lower the overall acceptance ratio for all the VNE approaches.

By comparing Figs. 5.8 and 5.9, we note that algorithms in comparison within the SvCDCNs took less time to finish the simulation due to their lower structure scalability compared with SwCDCNs that employ more substrate resources (switches and links) for traffic networking.

In both DCN types, we note a similar running time for small VNR sizes in the case of RW-BFS, NDC, and NRM algorithms. RW-BFS algorithm has the most extended runtime owing that to stochastic matrix manipulations. Contrarily, the MPT model employs a low-cost elementary operations in every procedure, including initial calculations, node ranking, and resource matching.

The algorithms' common goal is to greedily choose the most influential node throughout the mapping process. This selection can be observed in Figs. 5.10 and 5.11, where the average node usage is positively correlated with the corresponding average acceptance ratio in the two types of DCNs (i.e., an algorithm with higher AR_{avg} will induce a higher NUR_{avg} and LUR_{avg} while performed in the same comparative conditions). We emphasize that the obtained average utilization values (i.e., summing the NUR_{avg} and LUR_{avg} values) correspond to the most frequently used substrate resources (e.g., the output results of simulating a fat-tree topology with 8 pods, shows that 64 servers, 59 switches, and 103 links are repeatedly selected while mapping

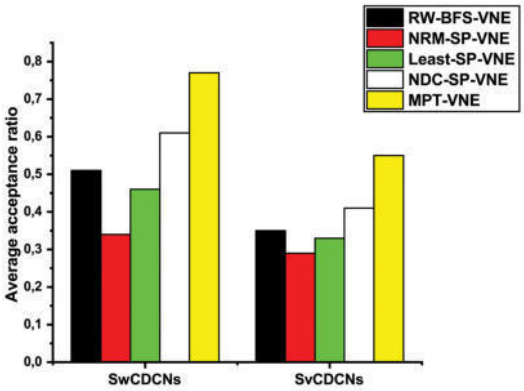


Fig. 5.7: Comparison of accepted VNRs among different DCNs

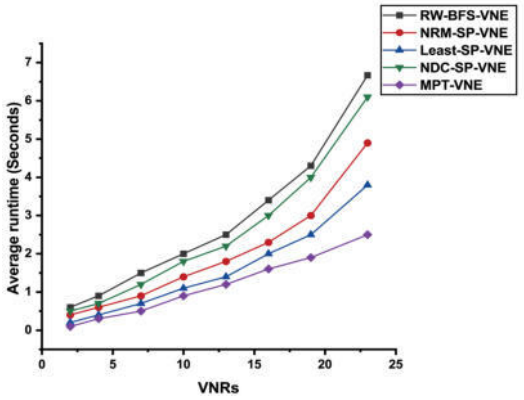


Fig. 5.8: The algorithms runtime for SwCDCNs

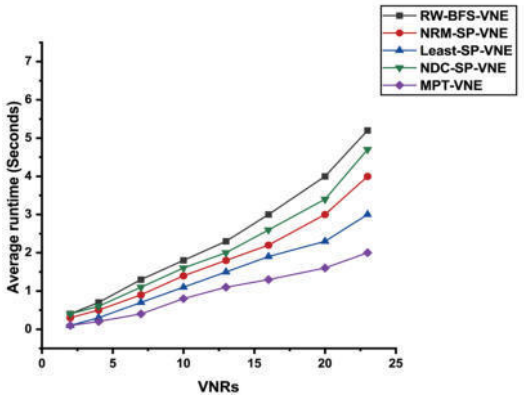


Fig. 5.9: The algorithms runtime for SvCDCNs

12 VNRs from a set of 23 VNRs). The algorithms' performance in SvCDCNs yields a lower average resource utilization than the ones in SwCDCNs. Our proposed algorithm has generated a nearly identical average

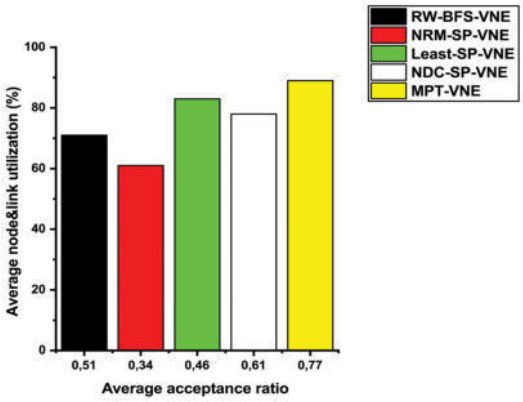


Fig. 5.10: The average node utilization in SwCDCNs

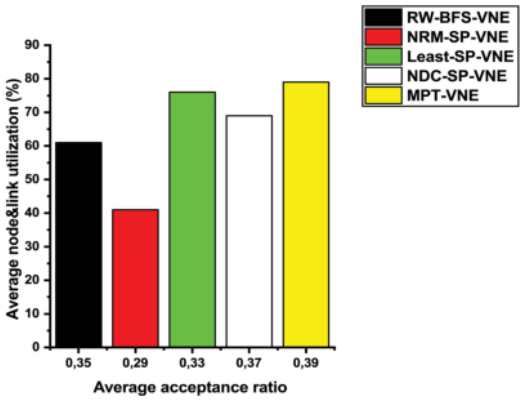


Fig. 5.11: The average node utilization in SvCDCNs

utilization rate with the Least-loaded algorithm (only 3% lower than MPT-VNE depicted in Fig. 5.11). Such an output is generated due to:

- The minimal existence of switches and links.
- The Least-loaded algorithm objective that seeks for the nodes with the lowest capacity and consequently maximizes the overall usage, without forgetting that nodes in SvCDCNs are already loaded nodes (traffic management).

Our approach scored the best average utilization rate thanks to the SC weight factor (latency), which highly increase the same SC selection during the matching process (ranking and mapping). Additionally, higher utilization of the attached links is noted. Nevertheless, for SvCDCNs, we suggest employing enhanced Least-loaded based algorithms in order to get the best possible results.

Notably, Figs. 5.8, 5.9, 5.10 and 5.11 have summarized the aspect of revenue to cost (R/C) relationship. Higher acceptance ratio with less involvement of substrate resources, will increase the SP's profits. Figs 5.10 and 5.11 reveal the average amount of resources repeatedly utilized over the received VNRs. Overall, result analysis proves that SwCDCNs is the most suitable physical structure for importance based VNE techniques.

6. Conclusion. In this paper, we took advantage of this ability to propose an adaptation of financial modeling (Markowitz model) known as modern portfolio theory, intending to provide an improved solution to the VNE problem. By adjusting this mathematical model, MPT proved a sufficient capability to inter-operate with the SDN structure, mainly engaged at the control layer. This work aims to propose a VNE policy that

undertakes decision-making through the SDN controller's supervision functionality. MPT-VNE employs the adapted asset allocation of MPT by treating the node and its attached links as a single composite to embed the VNR optimally.

Simulation assessment has focused on examining the proposed approach on two types of implemented DCNs to deeply explore the potential outcomes. Experimental results revealed that our VNE policy outperforms similar algorithms by scoring a higher acceptance ratio and substrate resource utilization with the lowest execution time in both DCNs types. Perspectively, we aim to propose an efficient energy consumption strategy that enables the enhanced effect of MPT-VNE on cloud resources. Besides, we seek to merge the Markowitz model with a scheduling algorithm based on VNR-priority classification to enforce the mapping process via multi-dimensional resource provisions.

REFERENCES

[1] Sasko Ristov, Kiril Cvetkov, and Marjan Gusev. Implementation of a horizontal scalable balancer for dew computing services. *Scalable Computing: Practice and Experience*, 17(2):79–90, 2016.

[2] Bin Wang, Zhengwei Qi, Ruhui Ma, Haibing Guan, and Athanasios V Vasilakos. A survey on data center networking for cloud computing. *Computer Networks*, 91:528–547, 2015.

[3] Andreas Fischer, Juan Felipe Botero, Michael Till Beck, Hermann De Meer, and Xavier Hesselbach. Virtual network embedding: A survey. *IEEE Communications Surveys & Tutorials*, 15(4):1888–1906, 2013.

[4] Hamza Mutaher Alshameri and Pradeep Kumar. An efficient zero-knowledge proof based identification scheme for securing software defined network. *Scalable Computing: Practice and Experience*, 20(1):181–189, 2019.

[5] Harry Markowitz. Portfolio selection, 1959.

[6] Xinbo Liu, Buhong Wang, and Zhixian Yang. Virtual network embedding based on topology potential. *Entropy*, 20(12):941, 2018.

[7] An Song, Wei-Neng Chen, Tianlong Gu, Huaxiang Zhang, and Jun Zhang. A constructive particle swarm optimizer for virtual network embedding. *IEEE Transactions on Network Science and Engineering*, 7(3):1406–1420, 2019.

[8] Zeheng Yang and Yongan Guo. An exact virtual network embedding algorithm based on integer linear programming for virtual network request with location constraint. *China Communications*, 13(8):177–183, 2016.

[9] Bernard M Waxman. Routing of multipoint connections. *IEEE journal on selected areas in communications*, 6(9):1617–1622, 1988.

[10] Koki Inoue, Shin'ichi Arakawa, Satoshi Imai, Toru Katagiri, and Masayuki Murata. Adaptive vne method based on yuragi principle for software defined infrastructure. In *2016 IEEE 17th International Conference on High Performance Switching and Routing (HPSR)*, pages 188–193. IEEE, 2016.

[11] Tausifa Jan Saleem and Mohammad Ahsan Chishti. Data analytics in the internet of things: a survey. *Scalable Computing: Practice and Experience*, 20(4):607–630, 2019.

[12] Peiying Zhang, Haipeng Yao, and Yunjie Liu. Virtual network embedding based on the degree and clustering coefficient information. *IEEE Access*, 4:8572–8580, 2016.

[13] Bo Li, Songtao Guo, Yan Wu, and Defang Liu. Construction and resource allocation of cost-efficient clustered virtual network in software defined networks. *Journal of Grid Computing*, 15(4):457–473, 2017.

[14] Jungmin Son, Amir Vahid Dastjerdi, Rodrigo N Calheiros, Xiaohui Ji, Young Yoon, and Rajkumar Buyya. Cloudsimsdn: Modeling and simulation of software-defined cloud data centers. In *2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pages 475–484. IEEE, 2015.

[15] Mohammad Al-Fares, Alexander Loukissas, and Amin Vahdat. A scalable, commodity data center network architecture. *ACM SIGCOMM computer communication review*, 38(4):63–74, 2008.

[16] Albert Greenberg, James R Hamilton, Navendu Jain, Srikanth Kandula, Changhoon Kim, Parantap Lahiri, David A Maltz, Parveen Patel, and Sudipta Sengupta. Vl2: A scalable and flexible data center network. In *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*, pages 51–62, 2009.

[17] Yantao Sun, Jing Chen, Q Lu, and Weiwei Fang. Diamond: An improved fat-tree architecture for large-scale data centers. *Journal of Communications*, 9(1):91–98, 2014.

[18] Chuanxiong Guo, Haitao Wu, Kun Tan, Lei Shi, Yongguang Zhang, and Songwu Lu. Dcell: a scalable and fault-tolerant network structure for data centers. In *Proceedings of the ACM SIGCOMM 2008 conference on Data communication*, pages 75–86, 2008.

[19] Dan Li, Chuanxiong Guo, Haitao Wu, Kun Tan, Yongguang Zhang, and Songwu Lu. Ficonn: Using backup port for server interconnection in data centers. In *IEEE INFOCOM 2009*, pages 2276–2285. IEEE, 2009.

[20] Chuanxiong Guo, Guohan Lu, Dan Li, Haitao Wu, Xuan Zhang, Yunfeng Shi, Chen Tian, Yongguang Zhang, and Songwu Lu. Bcube: a high performance, server-centric network architecture for modular data centers. In *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*, pages 63–74, 2009.

[21] Xiang Cheng, Sen Su, Zhongbao Zhang, Hanchi Wang, Fangchun Yang, Yan Luo, and Jie Wang. Virtual network embedding through topology-aware node ranking. *ACM SIGCOMM Computer Communication Review*, 41(2):38–47, 2011.

[22] Peiying Zhang, Haipeng Yao, and Yunjie Liu. Virtual network embedding based on computing, network, and storage resource constraints. *IEEE Internet of Things Journal*, 5(5):3298–3304, 2017.

[23] Min Feng, Jianxin Liao, Jingyu Wang, Sude Qing, and Qi Qi. Topology-aware virtual network embedding based on multiple characteristics. In *2014 IEEE International Conference on Communications (ICC)*, pages 2956–2962. IEEE, 2014.

Edited by: Katarzyna Wasielewska
Received: May 26, 2022
Accepted: Dec 2, 2022


Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 441–455, DOI 10.12694/scpe.v23i4.2027

IMAGE-BASED SEAT BELT FASTNESS DETECTION USING DEEP LEARNING

RUPAL A. KAPDI, PIMAL KHANPARA, ROHAN MODI, AND MANISH GUPTA *

Abstract. The detection of seat belts is an essential aspect of vehicle safety. It is crucial in providing protection in the event of an accident. Seat belt detection devices are installed into many automobiles, although they may be easily manipulated or disregarded. As a result, the existing approaches and algorithms for seat belt detection are insufficient. Using various external methods and algorithms, it is required to determine if the seat belt is fastened or not. This paper proposes an approach to identify seat belt fastness using the concepts of image processing and deep learning. Our proposed approach can be deployed in any organizational setup to aid the concerned authorities in identifying whether or not the drivers of the vehicles passing through the entrance have buckled their seat belts up. If a seat belt is not detected in a vehicle, the number plate recognition module records the vehicle number. The concerned authorities might use this record to take further necessary actions. This way, the organization authorities can keep track of all the vehicles entering the premises and ensure that all drivers/shotgun seat passengers are wearing seat belts.

Key words: Image Processing, Deep Learning, Number Plate Recognition, Seat Belt Detection

AMS subject classifications. 68T05

1. Introduction. When driving, it is critical to adhere to all safety regulations. It not only helps to prevent accidents but also protects drivers in the event of a crash or other adverse incidents. Seat belt detection systems are already installed in the majority of vehicles. However, many drivers just ignore the importance of using them. Many professional organizations make it a point to observe vehicle and traffic safety guidelines to be followed by their employees on their premises. Fastening seat belts is one of the basic and very important traffic and safety rules [17]. Many universities or other corporate organizations make efforts to ensure that everyone who enters their campus/office premises is aware of the fundamental traffic and vehicle safety regulations and also follows the same. However, manually keeping track of all entering vehicles and mandating drivers to wear seat belts is an extremely time-consuming and tedious task that is practically infeasible to execute in large organizations with a large number of employees coming to work in their own vehicles. Therefore, there is a need to have an automated system to keep track of entering vehicles to ensure compliance with the required safety protocols. This kind of system has obvious advantages of reducing manual intervention, improved accuracy, and overall faster execution. In this paper, an automated system to detect the fastened seat belts and number plates of the vehicles is proposed for a university campus. The proposed model is built based on the concepts of image processing and deep learning to achieve a high level of accuracy and performance.

2. Literature Review.

2.1. Existing Solutions for Number/License Plate Detection. Many techniques have been proposed by researchers for identifying the number plates/license plates of vehicles. The identification of the owner of the vehicle via number/license plate is crucial in implementing traffic safety rules either by government authorities or other private/public organizations [20]. The major problem in detecting the registered vehicle number from its number/license plate is the format and language in which the required details are printed. Not every country follows the same format/pattern and/or language for specifying vehicle registration details on the vehicle number plates. Therefore, though many solutions have been proposed in this domain, it is difficult to find a solution that can address all the above-mentioned issues related to extracting vehicle registration details. In the Indian context, it is required to devise a multi-language number plate detection technique as different states allow the

*Computer Science and Engineering Department, Institute of Technology, Nirma University, India (pimal.khanpara@nirmauni.ac.in). This work was supported by Nirma University.

Table 2.1: Summary of existing Number/License Plate Detection Approaches

Ref.	Salient Features	Accuracy	Dataset	Advantages	Limitations
[9]	Detection using image processing and Gaussian filter	Extraction - 84.28% Segmentation - 77.14% Recognition - 71.43%	70 images of Iranian vehicles in different weather conditions	Works well with low resolution noisy and low contrast images	Missing rate is 16% from 70 vehicles
[8]	YOLOv2 sensor with ResNet attribute extractor heart for NP detection; Uses YOLOv2 with ResNet50 model	Indian- 88.33% Test images 4789 (max for India)	DeebNP D2020	Can detect Multi-Language license plates	-
[2]	Multiple Pre-processing stages; License plate region extraction; characters segmentation and recognition; Uses Sobel edge detector, Median filter for noise removal	-	White images with black background with high contrast	Pixel by pixel Segmentation	Blurred images, broken number plates, similarity between some characters like O-D, 5-S, 8-B-E, 0-O
[10]	Input is taken as a video feed, split, segmented and CNN/KNN used in OCR	Accuracy (CNN): 93%	40512 images, (75/25) Test:4800 images + 50 short videos	Segmentation free because of use of RNN	Problem in KNN based prediction due to cropped images
[1]	Epochs 130k, learning rate 0.01; Uses YOLO model	Number plate detection: 100% Recognition: 91% Sorting Output: 100%	Manually collected 6500 images, tested on 640 images (90/10) split	Data augmentation to increase the number of characters extracted from less images training set	Difficulty in detecting similar looking characters such O and 0

usage of different languages for license plates. Moreover, based on our analysis of the existing approaches, it is found that the Convolutional Neural Networks would be the best choice for implementing the detection module for our proposed system. Table 2.1 describes the analysis of various existing techniques of number/license plate detection along with their advantages and limitations.

2.2. Existing Solutions for Seat Belt Recongnition. Many solutions have been proposed by researchers for detecting seat belts. Most of these solutions consider the problem as an object detection task and hence, they suggest YOLO (You Only Look Once) [6] algorithm for implementing the solution [18]. However, in our proposed system, the problem of detecting seat belts is considered an image classification task [5]. The main reason for not using the YOLO algorithm is that it requires a dataset in form of bounding boxes and the same is not available for the given problem domain [14]. Table 2.2 presents the study and analysis of various existing solutions along with their key features and scope of deployment.

As summarized in Table 2.2, most of the existing approaches for seat belt detection either do not provide details about the dataset used for training and testing their models or generate and use their own custom dataset.

3. Overview of the Proposed System. Our proposed system captures the image of the driver sitting in the car via the windshield using the pre-installed camera. The captured image is then processed and provided to a trained deep learning-based model to predict whether or not the driver is wearing a seat belt. In the event that the driver has not secured the seat belt, the system captures the vehicle's number plate. After processing the captured number plate image, the vehicle number is extracted and recorded. This information is then forwarded to the concerned authorities to take necessary measures. The proposed system's working and performance are tested for a university campus in India but the same system can be deployed in any organizational environment by making minimal changes to incorporate the necessary regulations.

The proposed system is divided into two modules – Seat Belt Detection Module and Number Plate Recognition Module. The first step in our proposed system is to capture videos of cars entering the university campus,

Table 2.2: Analysis of the existing Seat Belt Detection Approaches

Reference	Objective(s)	Proposed Methodology	Dataset Used
[7]	Detection of seat belts using monitoring images consisting of full scene information of the moving car	Extracts driver area using vehicle outline; Detects seat belt edges in the HSV color space; Further processing of edges to obtain the final outcome	100 real-world images captured under various conditions
[12]	Classification of driver seat belt status using a camera inside the driver's cabin	Uses a YOLO Neural Network-based model; Classifies outcomes as seat belt fastened correctly, not fastened, and fastened behind the back	Videos of drivers in multiple vehicles over three months in the wintertime in addition to images of unknown drivers taken from the Internet
[15]	To detect whether a driver wears a seat belt or not using a convolutional neural network	Uses a 6-hidden layered convolutional neural network model to classify the input images	A standard dataset containing the 2155 images taken from Yawning detection dataset
[22]	To detect seat belts automatically using salient gradient	Extracts edge features using edge detection; Building the salient gradient feature map and forwarding it to a machine learning model for prediction	Not mentioned
[21]	Seat belt detection using Convolutional Neural Network (CNN)	Use of improved CNN called BN-AlexNet to enhance the classification ability; Analysis of the confidence detection results	Custom dataset
[11]	To build a fully automated seat belt detection system	Use of sensors for weather condition detection to have better accuracy; Use of a single specialized model for each weather condition; Deploys Sensor-based AlexNet for classification;	925 rain images, 1211 night images, 969 rain night images, and 1321 clear images. For standardizing the experiment, 900 images are used in each weather type
[3]	To propose a seat belt detection algorithm for complex road backgrounds based on multi-scale feature extraction using deep learning	Extraction of multi-scale features; Use of CNN and SVM; Use of detection scores to improve accuracy	Not mentioned

using the cameras installed at the main gate of the university. The captured videos are then processed and the required images are extracted to prepare the dataset. Once the image dataset is prepared, data augmentation is performed and the model is trained. In the last step, the model is tested and necessary improvements are made.

Based on the outcomes and understanding of the literature review process, the best approach for implementing the required solution is determined. A review of existing systems also helps to understand the advantages and limitations of the available solutions and build an improved system to overcome the limitations of existing solutions. The proposed approach and implementation steps are explained in detail in the subsequent sections. The derived results and conclusions are also presented at the end of this paper.

4. Design and Deployment of the Proposed System.

4.1. Design and Deployment of Seat Belt Detection Module. Researchers have used different approaches to propose solutions for seat belt detection. However, the design of such solutions highly depends upon the deployment environment and expected end outcomes. As an outcome of our research and literature review process, it is established that CNN-based models can be used to identify whether the driver is buckled up. If the images of vehicles entering the university premises are available, they can be used to train the CNN model. Hence, generating a dataset of images from the video recordings is the first step in the processing as the deployment environment considered for the proposed system does not have images readily available and therefore, it is necessary to extract the images from the video recordings of the main entrance of the university. The dataset generation process also includes resizing and re-scaling of the images that are divided into two groups: with and without seat belts.

A data augmentation process is to next step where the training dataset is populated with additional images generated using various augmentation techniques. Moving further, a transfer learning-based approach [16] is

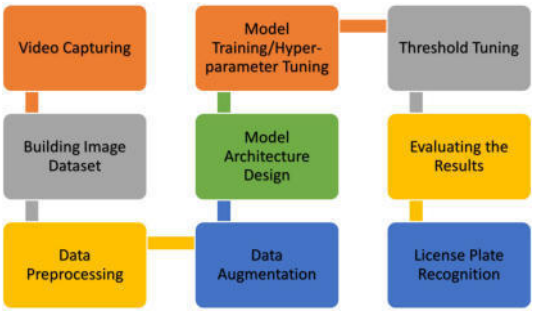


Fig. 4.1: Design Flow of the Proposed System

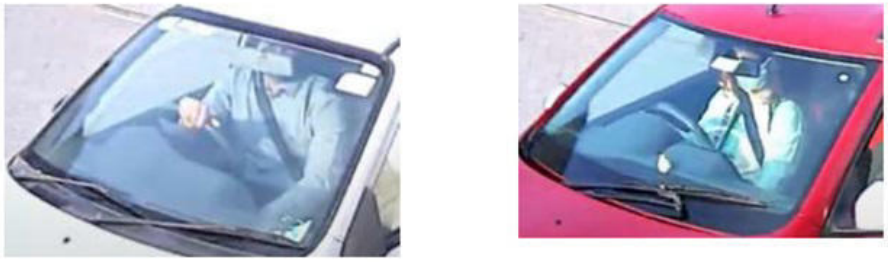


Fig. 4.2: Sample input images with a minimum requirement of pre-processing

employed to generate a CNN-based model. The model uses MobileNetV2 [4] as the base model with a few additional layers for precise classification. Following that, sample testset images are classified and results are generated. Figure 4.1 depicts the functional blocks and the flow of execution for the proposed system. The same is described in detail in the following subsections.

4.1.1. Dataset Generation. As there is no proper dataset available for seat belt detection, it is required to generate one as the first step of the proposed solution. The deployment environment considered in this work is a university organization and the cameras are installed at the entry gates to capture the vehicles entering the university premises. Therefore, it is required to extract images from the captured videos for the preparation of the dataset. There are two categories in which the images are stored: with seat belts and without seat belts. In our experiment, a total of 263 images were stored in the dataset, out of which 233 were included in the "with seat belt" category and the remaining 32 images were included in the "without seat belt" category. However, the angles of the installed cameras and quality of the video recordings lack adequacy for generating the dataset which leads to the further processing of the images.

4.1.2. Data pre-processing. In the data pre-processing phase, the images are processed in such a way that they become well-suited for training the model. There can be different types of images in the dataset and each type requires a different level of pre-processing to resolve specific types of problems. For example, the images shown in Figure 4.2 are easy to use and do not require much processing.

On the other hand, some images are either too bright or too dark and that makes the identification of a seat belt difficult. Moreover, sometimes, the color of the seat belt is also a problem if it blends with the color of the clothes the driver is wearing, making it really difficult to detect. Such erroneous sample images having these problems are shown in Figure 4.3 for which our proposed system could not detect the seat belt.

Another problem with the input images is the presence of some obstacles that block the view of the seat belts. Such obstacles can be any objects such as hands, windshields, etc. Such objects obviously make the detection of seat belts very difficult. An example of such an input image is shown in Figure 4.4.

Fig. 4.3: Sample input images with a maximum requirement of pre-processing

Fig. 4.4: Sample input image with Seat belt view blocked

Before applying any scaling or processing method to the input images, it is required to take care of the above-mentioned practical issues. Many of the above issues can be resolved if the brightness or contrast levels are modified in the images. In our experiments, the following approaches are used for the basic pre-processing of all the images included in the dataset:

- Cropping the image so only the front view of the driver is visible
- Resizing the image to 128 x 128
- Blurring the image for smoothness. Here Gaussian blur is used with the kernel size (1,1)
- Applying the built-in pre-process_input() function from the keras.applications.mobilenet_v2 package. This function directly pre-processes the image according to what MobileNetV2 expects the input to be. It scales input pixels between -1 and 1 and other basic transformations

Following that, the whole dataset is split into a 90% training set and a 10% validation set. Class labels are also kept stratified into both training and validation sets. The model is trained on the training set after applying data augmentation and then validated on the validation set after each epoch.

4.1.3. Data Augmentation. Data augmentation is important to perform when there is a limited number of images available for the data set. The proposed system uses a CNN-based model and it has quite a large number of parameters. Data augmentation is necessary to prevent over-fitting. Another advantage of data augmentation is that it makes the trained model more robust for classifying unseen images.

In the Keras library [13], the ImageDataGenerator class is used for implementing various augmentation transformations. The following transformations are applied with their respective parameters:

- Random rotation (range = 20)
- Random zoom (range = 0.15)
- Random width shift (range = 0.2)
- Random height shift (range = 0.2)
- Random brightness (range = 0.8 to 1.2)

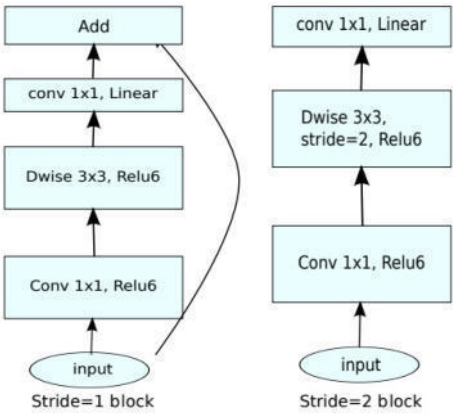


Fig. 4.5: MobileNetV2 convolutional block [19].

Table 4.1: Comparison of base models

Base model	**Total parameters**	**Training accuracy**
Xception	21,386,538	51.85%
VGG16	14,846,530	88.89%
ResNet50	24,112,770	88.89%
InceptionV3	22,327,842	74.37%
MobileNetV2	2,586,434	88.89%

- Horizontal flip

4.1.4. Model Architecture Design. It is required to have a model which performs binary image classification on the dataset. CNN is proven to be best suited for problems related to image classification. Therefore, the proposed system uses a CNN-based model. Since there are many reference models available for image classification problems, they can be directly utilized. Most of these base models are trained on the ImageNet dataset hence they can be utilized for performing preliminary image localization and feature extraction. In addition, the concept of transfer learning is applied in the proposed work. The traditional supervised learning paradigm breaks down in the absence of sufficient labeled data for the domain in which there is a need to train a reliable model. This may lead to a deterioration in performance. Hence, transfer learning helps to deal with this problem by leveraging already existing labeled data of some related domain. Therefore, instead of training a neural network from scratch, the state-of-the-art lightweight MobileNetV2 [19] is used as the base model. Further, instead of training weights, ImageNet weights are directly used. This base model acts as a generic feature detector. The architectural details of MobileNetV2 are as follows:

- There are two types of 3 layers convolutional blocks. One is a residual block with stride 1 and the other is a block with stride 2 used for down-sampling.
- It outperforms both MobileNetV1 and ShuffleNet which have almost similar model size and computational cost.

In the proposed system, as depicted in Figure 4.5, MobileNetV2 is selected as the best choice for the base model as per its comparison with other state-of-the-art CNNs for the given problem of seat belt detection. The training accuracy of various models is shown in Table 4.1.

MobileNetV2 is comparatively faster and has fewer parameters as shown in Table 4.1. These model advantages are best suited for the task which has fewer images for training. However, in some scenarios, MobileNetV2 has lesser accuracy compared to other models but the accuracy is not a major concern in the deployment scenario

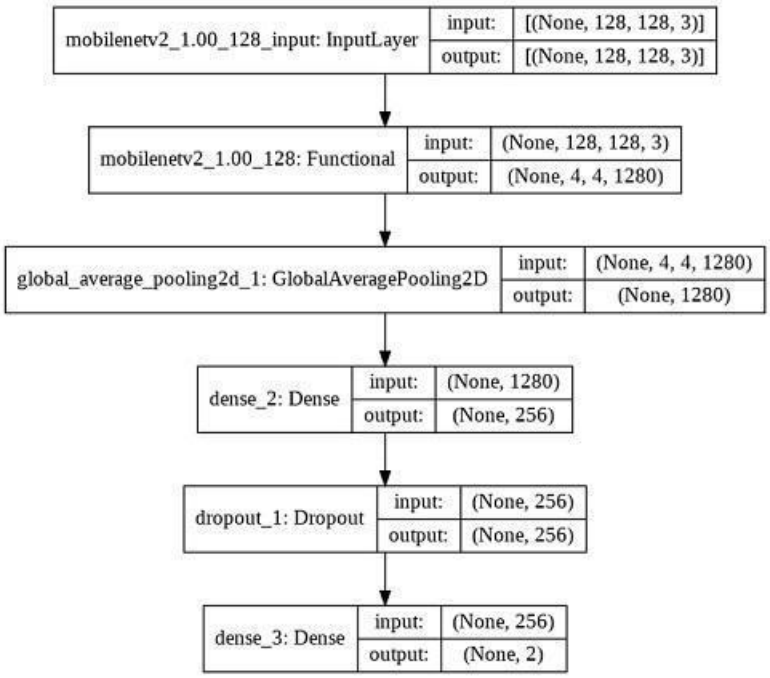


Fig. 4.6: Proposed Model Architecture

of the proposed system. During the deployment phase, some architectural changes are done in MobileNetV2, where the topmost layer is removed from the base model and a few extra layers, as described below, are added to perform the classification:

- Global average pooling layer aggressively summarizes the features from the last convolutional layer
- Fully connected layer with 256 neurons and ReLU as activation function
- Dropout layer with a dropout rate of 0.5 to prevent over-fitting
- Final classification layer with 2 neurons and softmax as the activation function
- Since it is a binary classification problem, Binary Cross Entropy is used as a loss function. Again, Adam is used as an optimizer since it performs the best in this case. Accuracy is chosen as a metric for model training. The model architecture for the proposed system is shown in Figure 4.6.

4.1.5. Model Training. For the model selected for the proposed system, there is a total of 2,586,434 parameters out of which 328,450 are the trainable parameters. The initial learning rate is fixed to 1e-3. The model is trained for 8 epochs with a batch size of 16. Here, the number of epochs is very less. The model starts over-fitting the training data as the number of epochs increases. The graph as shown in Figure 4.7 depicts this fact. The training data is passed after performing all the augmentations as described previously but the validation set is not augmented. The best combination of the parameters is determined after the model is trained multiple times.

4.2. Design and Deployment of Number Plate Recognition Module. Upon identifying a car-driver without a seat belt, the proposed system invokes the number plate recognition module. This module scans the number plate of the identified car so that the owner of the car can be identified and further disciplinary actions can be taken. The process of number plate detection consists of two phases: i) pre-processing the captured image and ii) recognizing characters from the number plate.

4.2.1. Pre-processing.

1. Edge Detection

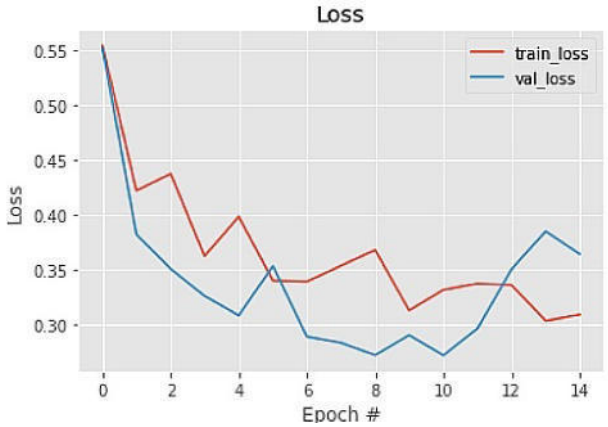


Fig. 4.7: Over-fitting of the model after certain epochs

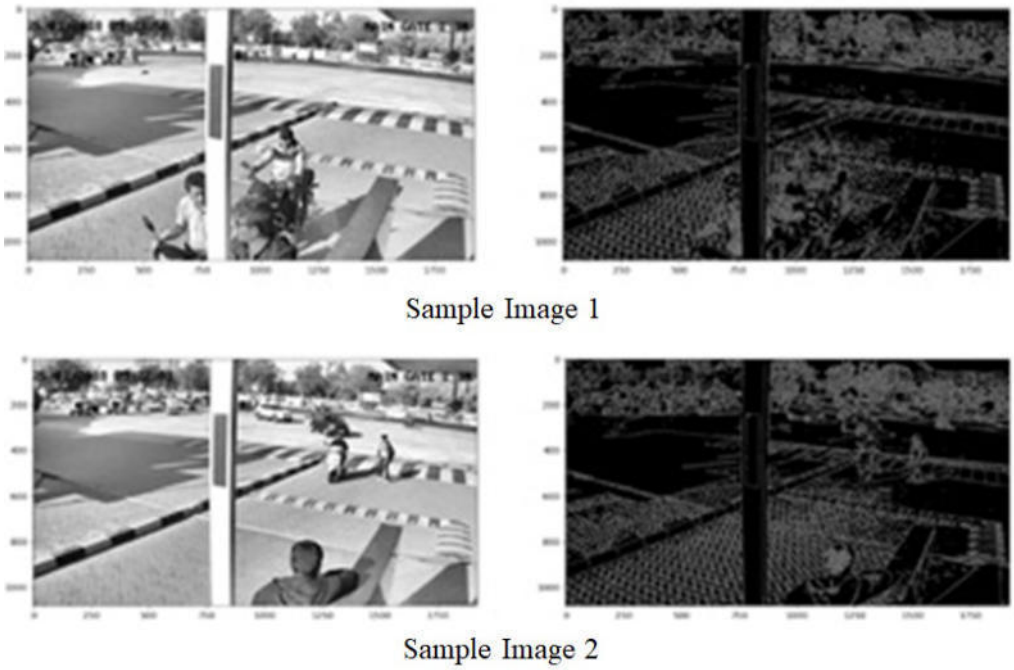
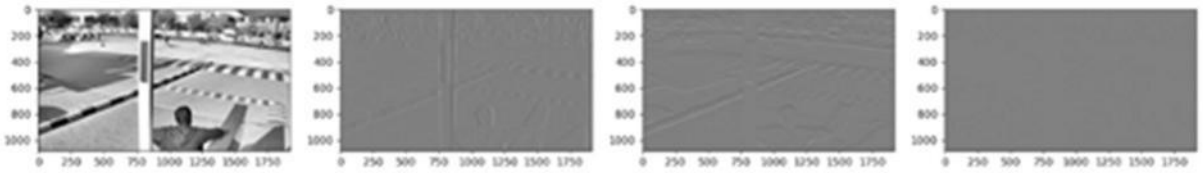


Fig. 4.8: Output of Canny Edge Detection Algorithm for Sample Images

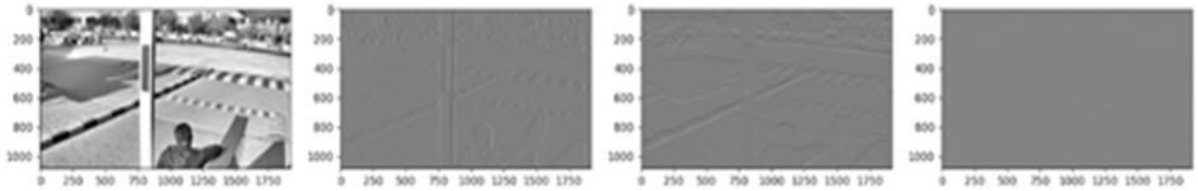
A sudden change of discontinuities forms horizontal edges, vertical edges, and diagonal edges. Most of the shape information of an image is enclosed in the edges. Hence, it is first required to detect these edges in an image using filters and enhance the image area which contains edges to improve the edge sharpness. Canny edge detector and Sobel edge detector are two widely used edge detection algorithms. The outputs of the Canny and Sobel edge detection algorithms on some of the sample images are shown in Figure 4.8 and Figure 4.9 respectively.

2. Color Space Transformations
   A color space is a specific organization of colors. In combination with color profiling supported by various physical devices, it supports reproducible and device-independent color representations - whether

[htb]



(a) Sobel edge Detection using x-direction filter



(b) Sobel edge Detection using y-direction filter

Fig. 4.9: Output of Sobel Edge Detection Algorithm for Sample Images

such representation entails an analog or a digital representation. The input images are tested on three different types of space transformations, i.e. HSV, LAB, and YCrCb as shown in Figure 4.10.

4.2.2. Character Recognition from Number Plate. Number plate recognition is an important application in the area of computer vision. It is already being used in various real-life scenarios to detect and recognize license plates at junction points in urban areas. In our proposed system, the objective of the license plate detection module is to identify the vehicles through their registered license plates which are not following the vehicle driving rules set by the government/organization. The license plate recognition system can be decomposed into two major sub-components which are: i) Detect and localize a license plate in an input image/frame, ii) Extract the characters from the license plate. However, due to the complexity of the deployment scenario, the images are required to be pre-processed further before giving them as input to the proposed model. This pre-processing majorly involves the cropping of images from the scene to maximize the ROI of the license plate.

Step 1: Detect and localize a license plate in an input image

The cropped image is further processed to obtain the exact number plate. Such scanned and cropped image areas are examined to check for existence of alphabets in them. The availability of alphabets is checked with help of contours in the targeted cropped area. K-Nearest Neighbors (KNN) classifier helps to identify contours in the targeted area. The KNN works as follows:

- Prepare the training dataset.
- Find the Euclidean distance for the test dataset record with all the data records in the training dataset.
- Find the top K distance using sorting of the test dataset record with the training dataset.
- Find out the neighbors that happen maximum times.
- Classify the character of that neighbor.

KNN finds all the possible number plates in the scene and stores their dimensions in an array. Further, the dimensions are pixel-wise checked to be a number plate. Upon successful retrieval of the number plate region,

[htb]

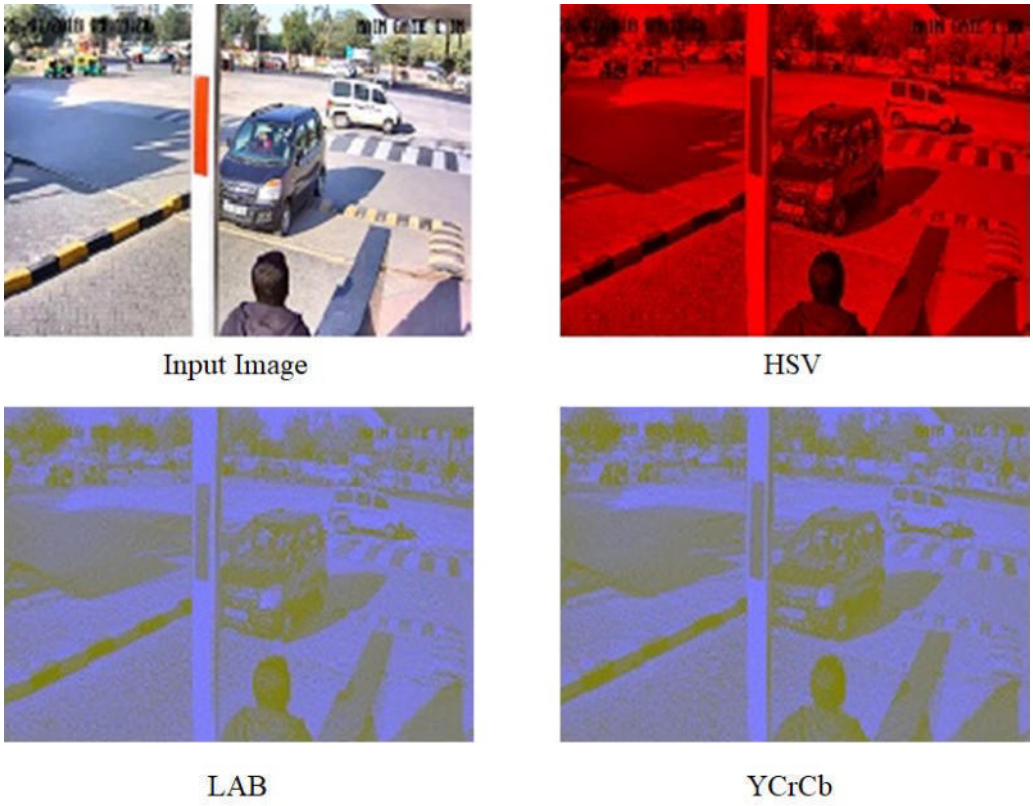


Fig. 4.10: Outcome of Color Space Transformation

it is further processed for character recognition.

Step 2: Extract the characters from the license plate

The ROI is used to extract the characters from the number plate and given to KNN classifier for further classification. The KNN classifies the characters in a similar manner as explained in step 1. The recognized characters generate the sequence of alphabets for number plate output.

5. Result Analysis.

5.1. Testing the trained model. During the training time, the model is found to achieve maximum accuracy of 89.19% and a loss of 0.3314. On the validation set also, the model performs almost similarly. It produces a validation accuracy of 88.89% and the validation loss reduces to 0.2497. The trained model is used to make predictions on an unseen set of data. This data consists of a mixed class of labels. After making predictions on this data, it is compared with the ground truth labels and the classification report is generated, as shown in Table 5.1. Figure 5.1 and Figure 5.2 present the accuracy and loss of the training and validation process after each epoch.

5.2. Thresholding for imbalanced dataset. The dataset generated and used for the proposed system is highly imbalanced. In the dataset, there is around 90% of the data that represents scenarios with seat belts and only 10% of the remaining data depicts the cases without seat belts. Hence, the dataset is said to be highly biased towards making predictions for the category, “with seat belt”. This fact is clearly visible from the classification reports. It is observed that there is 0 recall and F1-score for without seatbelt class. There are many techniques to handle this issue of imbalanced classification. The simplest approach is to change the default

Table 5.1: Classification report

	Precision	**Recall**	**F1-Score**
With Seat Belt	0.89	1.00	0.94
Without Seat Belt	1.00	0.00	0.00
Accuracy	0.89	0.89	0.89
Macro Average	0.94	0.50	0.47
Weighted Average	0.90	0.89	0.84

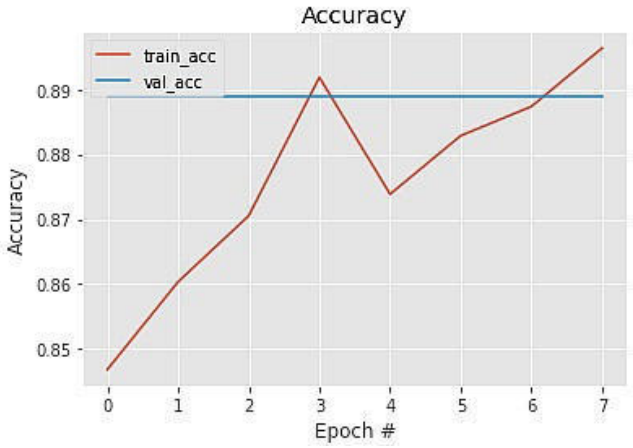


Fig. 5.1: Epoch-wise Training and Validation Accuracy

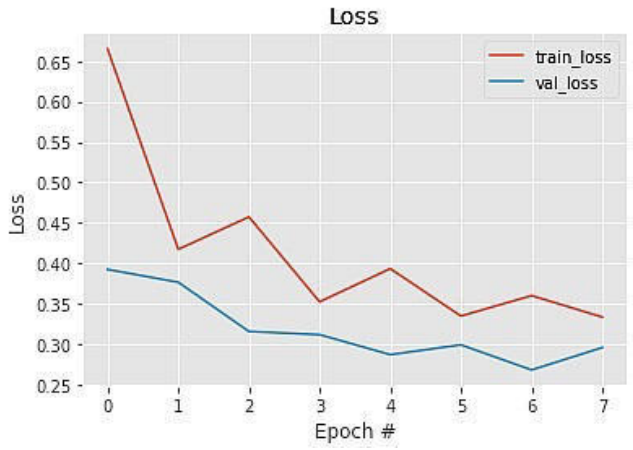


Fig. 5.2: Epoch-wise Training and Validation Loss

decision threshold. Hence, tuning or shifting the decision threshold to accommodate broader requirements of the classification problem is called thresholding.

The sigmoid function used in the experiment produces the final result between 0 and 1. The Keras library takes 0.5 as the default decision threshold and classifies the test instances accordingly. But it is not appropriate to 0.5 as the threshold and there is a need to determine an optimal threshold value. The basic approach is to

Table 5.2: Classification report based on thresholds

Threshold	Precision	Recall	F1-Score
0.001	0.41	0.4	0.40
0.002	0.6	0.5	0.54
0.003	0.6	0.5	0.54
0.004	0.65	0.55	0.60
0.005	0.65	0.55	0.60
0.006	0.51	0.5	0.50

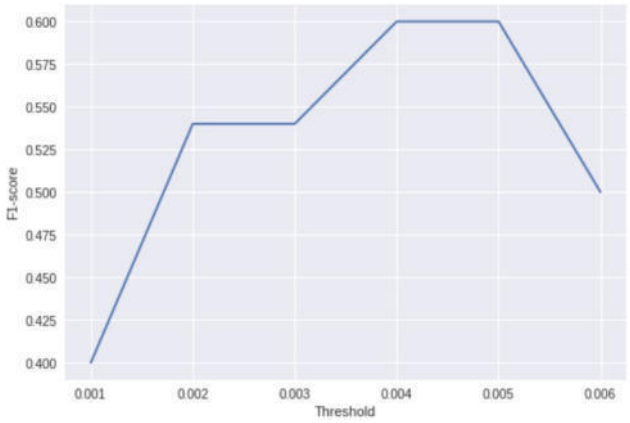


Fig. 5.3: Threshold vs. F1-score

define a set of thresholds and evaluate the predicted probabilities under each case to determine the optimal threshold. This procedure is summarized below:

- It is observed that the model achieved the f1-score of 0.84 on testing data. The same model is used but instead of directly predicting the class labels, class probabilities are predicted.
- Probabilities are required only for the minority class i.e., without a seat belt. Hence, a set of thresholds for evaluating probabilities are defined. For this experiment, thresholds considered are from 0.0 to 1.0 in step size of 0.001 i.e., we will test 0.000, 0.001, 0.002, 0.003 and so on up to 0.999.
- For each threshold, the predicted labels are evaluated using the f1-score. The one which has the largest F1-score is considered the optimal threshold. F1-scores for a few threshold values are shown in Table 5.2. The graphical analysis of the threshold and F1-scores is shown in Figure 5.3.

5.3. Outcome of classification. The example images presented in this section depict the sample output generated by the trained model. The pre-processing steps mentioned in the previous sections are applied to each input image. The model returns whether the seat belt is fastened or not along with the class probability of "with seat belt" cases. Figure 5.4 presents the sample outcome of prediction for positive class (with seat belts). Figure 5.5 shows the results of the prediction where the drivers are not wearing their seat belts.

5.4. Number Plate Recognition result analysis. The correct output for the above steps is shown in Figure 5.6(a). Some of the characters are not properly classified as shown in Figure 5.6(b). The two-step process for number plate output generation gives varied accuracy as shown in Table 5.3.

6. Conclusion. Fastening seat belts is considered one of the most important vehicle safety practices. In the proposed system, seat belt detection is done through the front view of the car by applying deep learning and image processing concepts. A model is designed that can predict whether the seat belt is fastened or not from the input image. The dataset for this model is prepared by extracting the images for the video recordings.

Fig. 5.4: Sample Outcome for Positive Class

Fig. 5.5: Sample Outcome for Negative Class

Table 5.3: Accuracy for two-step number plate output process

Steps	**# input images**	**# output images**	**Accuracy**
Step 1	263	250	95%
Step 2	250	225	90%

Input images are also pre-processed so that they can be used for training the CNN-based model using transfer learning. The results obtained during the testing phase of the proposed model are found to be satisfactory from the accuracy point of view. However, to improve the existing percentage of accuracy which is 89%, the

Fig. 5.6: Outcome of Color Space Transformation

dataset can be made balanced and have more images of higher quality. The proposed system can be used by any organization with minor modifications to ensure the safety of its stakeholders. After some improvements, the proposed model can be made more robust to handle diverse weather conditions and still offer a good level of accuracy.

Acknowledgement. The authors wish to thank Nirma University, Ahmedabad, India for providing all the necessary financial and technical support for carrying out this work.

REFERENCES

[1] R. N. Babu, V. Sowmya, and K. Soman, *Indian car number plate recognition using deep learning*, in 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), vol. 1, IEEE, 2019, pp. 1269–1272.

[2] G. N. Balaji and D. Rajesh, *Smart vehicle number plate detection system for different countries using an improved segmentation method*, vol, 3 (2017), p. 7.

[3] Y. Chen, G. Tao, H. Ren, X. Lin, and L. Zhang, *Accurate seat belt detection in road surveillance images based on cnn and svm*, Neurocomputing, 274 (2018), pp. 80–87.

[4] K. Dong, C. Zhou, Y. Ruan, and Y. Li, *Mobilenetv2 model for image classification*, in 2020 2nd International Conference on Information Technology and Computer Application (ITCA), IEEE, 2020, pp. 476–480.

[5] P. Druzhkov and V. Kustikova, *A survey of deep learning methods and software tools for image classification and object detection*, Pattern Recognition and Image Analysis, 26 (2016), pp. 9–15.

[6] J. Du, *Understanding of object detection based on cnn family and yolo*, in Journal of Physics: Conference Series, vol. 1004, IOP Publishing, 2018, p. 012029.

[7] H. Guo, H. Lin, S. Zhang, and S. Li, *Image-based seat belt detection*, in Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety, IEEE, 2011, pp. 161–164.

[8] J. Gupta, V. Saini, and K. Garg, *Multilanguage number plate detection using convolutional neural networks*, arXiv preprint arXiv:2008.08023, (2020).
[9] P. S. Ha and M. Shakeri, *License plate automatic recognition based on edge detection*, in 2016 Artificial Intelligence and Robotics (IRANOPEN), IEEE, 2016, pp. 170–174.
[10] S. N. Hashmi, K. Kumar, S. Khandelwal, D. Lochan, and S. Mittal, *Real time license plate recognition from video streams using deep learning*, International Journal of Information Retrieval Research (IJIRR), 9 (2019), pp. 65–87.
[11] O. Hosam, *Deep learning-based car seatbelt classifier resilient to weather conditions*, Int. J. Eng. Technol., 9 (2020), pp. 229–237.
[12] A. Kashevnik, A. Ali, I. Lashkov, and N. Shilov, *Seat belt fastness detection based on image analysis from vehicle in-abin camera*, in 2020 26th Conference of Open Innovations Association (FRUCT), IEEE, 2020, pp. 143–150.
[13] N. Ketkar, *Introduction to keras*, in Deep learning with Python, Springer, 2017, pp. 97–111.
[14] M. Maity, S. Banerjee, and S. S. Chaudhuri, *Faster r-cnn and yolo based vehicle detection: A survey*, in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2021, pp. 1442–1447.
[15] D. B. Naik, G. S. Lakshmi, V. R. Sajja, D. Venkatesulu, and J. N. Rao, *Driver's seat belt detection using cnn*, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12 (2021), pp. 776–785.
[16] S. J. Pan and Q. Yang, *A survey on transfer learning*, IEEE Transactions on knowledge and data engineering, 22 (2009), pp. 1345–1359.
[17] D. Saini, V. Arundekar, K. Priya, and D. J. D'Souza, *Identification of helmets on motorcyclists and seatbelt on four-wheeler drivers*, in Recent Advances in Artificial Intelligence and Data Engineering, Springer, 2022, pp. 99–107.
[18] A. Ş. Şener, I. F. Ince, H. B. Baydargil, I. Garip, and O. Ozturk, *Deep learning based automatic vertical height adjustment of incorrectly fastened seat belts for driver and passenger safety in fleet vehicles*, Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 236 (2022), pp. 639–654.
[19] S.-H. Tsang, *Review: Mobilenetv2—light weight model (image classification)*, Towards Data Science, Svibanj, (2019).
[20] Z. Wang and Y. Ma, *Detection and recognition of stationary vehicles and seat belts in intelligent internet of things traffic management system*, Neural Computing and Applications, 34 (2022), pp. 3513–3522.
[21] B. Zhou, D. Chen, and X. Wang, *Seat belt detection using convolutional neural network bn-alexnet*, in International Conference on Intelligent Computing, Springer, 2017, pp. 384–395.
[22] B. Zhou, L. Chen, J. Tian, and Z. Peng, *Learning-based seat belt detection in image using salient gradient*, in 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2017, pp. 547–550.

Edited by: Katarzyna Wasielewska
Received: Jun 28, 2022
Accepted: Nov 2, 2022

Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2022 SCPE. Volume 23, Issues 4, pp. 457–472, DOI 10.12694/scpe.v23i4.2038

BLOCKCHAIN ENABLED ARCHITECTURE WITH SELECTIVE CONSENSUS MECHANISMS FOR IOT BASED SAFFRON-AGRI VALUE CHAIN

JAHANGEER ALI* AND SHABIR A. SOFI †

Abstract. The Internet of Things (IoT) is the backbone behind numerous smart and automated applications in the modern era by providing seamless connectivity and information retrieval among physical and virtual objects. IoT networks are resource constraint platforms hence prone to security and privacy challenges. Blockchain technology comes to the forefront to improvise security, privacy, and less dependency on third-party centralized servers. There exists a rich amount of work with numerous practical applications by fusing IoT and blockchain. In blockchain technology, the consensus mechanisms are considered to be the driving force in its implementation. In this paper, we propose a simplified Blockchain-based Internet of Things (BIoT) architecture for resource-constrained IoT devices with selective consensus mechanisms based on the scale of IoT networks. We have selectively highlighted some of the important consensus algorithms which are favorable for IoT networks. We have tailored the blockchain framework to suit resource-constrained IoT networks. We implemented a prototype leveraging the blockchain and IoT network to evaluate our design. The preliminary results suggest that the proposed system incorporating supply chain management of the Saffron-Agri value chain outperforms the existing systems. Furthermore, we have carried out a detailed case study on the cultivation and marketing strategies for maintaining originality and transparency starting from farmer to consumer in the Saffron-Agri value chain.

Key words: IoT, blockchain, security, BIoT architecture, consensus mechanism, SCM, BIoT applications

AMS subject classifications. 68M14

1. Introduction.

With the advancement in information communication technology, most of the real-world use cases are utilizing its benefits by connecting the smart objects accessible globally with unique addresses to create a smart ecosystem with on-demand services at any place [71]. IoT as a technology is growing enormously in fact having a direct role in the economy of any country [49]. The IoT-based networks mostly possess heterogeneous and inter-operable behavior which becomes a challenging task to maintain all operational standards in the network [21]. The IoT system is dependent on third party like cloud-based centralized services which cannot be trusted all the time [1]. The data generated by IoT devices can be manipulated by unauthorized users which are stored on these centralized servers. And these centralized servers will restrain the adoption of IoT technology due to the huge amount of data generated by IoT end devices. This leads to a single point of failure [5]. The IoT applications are spreading over vast fields like supply chain logistics [30], agriculture [71], healthcare, smart cities [55], smart grids, industries 4.0 [64]. Researchers work in different fields like inter-operable platforms, architectures, standardized protocols, and emerging technologies [71]. Considering the security and privacy of these networks which is of much importance as most of the human-related activities are been processed and shared globally [36]. The last research development is seen to maintain the security and privacy of IoT devices. One of the reasons for the fewer security mechanisms specific to IoT devices is that scarce resources lead to various security and trust-based issues [3].

Blockchain is a distributed ledger that maintains the data in form of blocks in a decentralized manner. The complete chain of blocks is linked together by using a cryptographic hash function in a peer-to-peer communication [5]. The participating nodes maintain identical copies of the complete ledger which makes it more transparent. Once the new block is added to the existing blockchain network after due consensus arrived between the participating parties no change can be made to the data block. At any instance time in history, actual

*Research Scholar, Department of Information Technology, National Institute of Technology Srinagar, J&K, India (jehangir_04phd18@nitsri.ac.in).

†Associate Professor, Department of Information Technology, National Institute of Technology Srinagar, J&K, India (shabir@nitsri.ac.in).

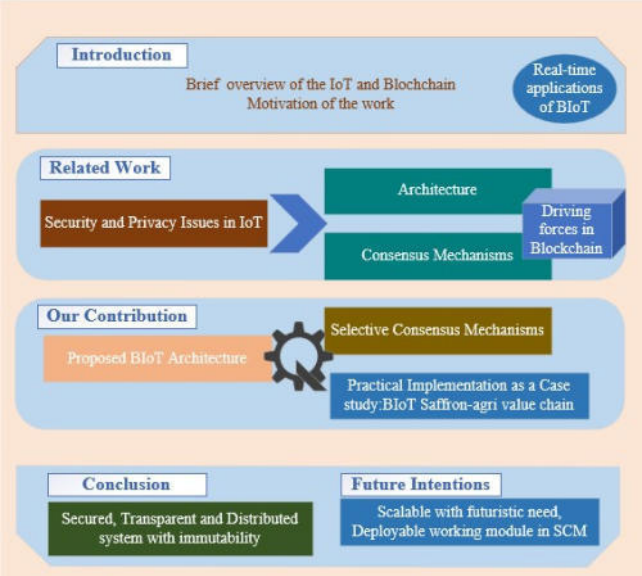


Fig. 1.1: Taxonomy of the article with distinctive highlights.

information can be retrieved from the blockchain network. Blockchain technology can monitor, communicate and perform transactions for large quantities of IoT devices without any intermediary [18]. Blockchain networks come in various variants depending on the complete access and restricted access to the participating peers in the network. In public blockchain networks, anyone can join the network. While in the case of a private blockchain, the owner can restrict limited access to the participating users. The integration of blockchain technology with IoT networks will definitely improve security, transparency, and immutability. Blockchain technology is adding new dimensions to the existing IoT networks with promising applications like supply chain logistics [35], smart agriculture [5], smart land records, System firmware updates [18], smart-grids [49], smart cities [10].

The paper motivates about the importance of integration of blockchain technology with IoT networks in maintaining security and transparency in a distributed architecture. To address the problem of a generalized Blockchain-based Internet of Things (BIoT) architecture, we propose a simplified BIoT architecture with a selective consensus mechanism for large and small-scale IoT applications. The proposed architecture uses inbuilt features of security and privacy to secure the data generated by the IoT devices in encrypted hash codes which are difficult to eavesdrop and all blocks are chained together with the hash codes of every previous block in the chain. And every main node maintains a complete copy of the ledger. Figure 1.1 provides a clear visualization of this article. The rest of the paper is organized as Section 2 discusses the work carried out in areas of interest like security and privacy issues, Real-time applications of BIoT, BIoT architectures, and various consensus mechanisms related to blockchain. Section 3 presents our proposed work like BIoT architecture with a provision for the selection of optimal consensus mechanism with respect to the scale of the IoT application area. Section 4 discusses the detailed use case-based novel study in integrating blockchain with IoT ecosystems. Section 5 concludes the paper and section 6 discusses the possible future perspectives for enhanced BIoT ecosystems.

The highlights of our proposed system are as under:

- We propose a BIoT architecture that comprises vast IoT application domains with customized consensus mechanism applicability.
- Although numerous consensus mechanisms are available in the literature but we have selected some distinctive consensus mechanisms specifically favorable for IoT applications only.
- Practical Implementation as a Case study: BIoT Saffron-Agri value chain.

2. Related Work.

2.1. Security and privacy issues in IoT. With the adoption of more IoT application use cases, millions of devices are getting connected over the internet with huge data collection mostly dependent on centralized cloud services. Most of the IoT devices are resource constrained in nature which makes it difficult for existing security frameworks and protocols to safeguard and preserve the privacy and security of these devices [4]. Almost the least priority is given to security and privacy measures in IoT ecosystems mostly relying on wireless medium in a heterogeneous manner [36, 28]. There are various types of security and privacy attacks in IoT networks which include Denial of services(DoS) and Jamming attacks in which the attacker broadcasts radio signals in the shared wireless channels which badly affects the communication of IoT devices which can discharge the limited battery of IoT nodes and leads to node failure [64, 21, 54]. Distributed DoS leads to blocking of channels, memory and CPU utilization in 96% of IoT devices [31]. In a Sybil attack, the IoT devices are prone to fake identities created by the attacker which deteriorates the functioning of normal IoT nodes in the network [21]. The attacker node with multiple addresses captures most of the data within the compromised network [28]. In sleep denial attack, the IoT devices deployed in remote areas are mostly equipped with sleep-enabled modes in case of no data activity to conserve energy. The attacker can alter the infect the programming code of sleep mode and will make the node active all the time which will decrease the network lifetime of the node [2]. In a sinkhole attack, the sinkhole node compromises the existing IoT network and then acts as a promising sink node for all neighboring nodes. With the result, all nodes forward the data to the sinkhole node [54]. In a wormhole attack, the attacker creates a falsely high resource link which attracts the other nodes to forward their packets via wormhole node [54, 28]. Man-in-the-middle attack (MITM) is the major threat to the security and privacy of the end devices in the IoT ecosystem [31]. The attacker masquerades the communication between two IoT nodes which can monitor or modify the delicate data between comprised IoT nodes [54]. In malicious code injection, various types of malicious programs can be injected in the form of viruses, and trojans into IoT networks which can lead to loss of data or non-availability of resources in the network [2]. And in routing protocol for low power and lossy networks(RPL) attacks in IoT networks can easily be compromised by the attacker nodes which can increase the network latency and restrict the resource usage.

As IoT devices are facing serious challenges in securing the data shared within the IoT networks. There is a dearth of research on new enabling technologies that must be formulated or integrated with IoT networks in order to secure the information stored or transactions carried out in these systems. Thus collaborating blockchain technology with IoT networks in the distributed and decentralized with in-built security and privacy mechanisms[37, 60].

2.2. Real-time applications of blockchain-based IoT systems. IoT has a vast application area, the blockchain involvement as the backbone network will add more dimensions in a new era of application services. S. B. Rane et al.[47] focused on the future commitments of the supply chain towards a green economy. The main contribution of this article actively incorporates the involvement of stakeholders and using BIoT in achieving the goal. The authors have also proposed a BIoT architecture in a very simplified manner by involving the majority of the stakeholders. However this work does not present the practical implementation of the research findings in specific application use cases. Weijun Lin et al. [26] highlighted in detail the importance of BIoT in the field of agriculture systems and practices. The authors have done a thorough survey on the technical aspects like cryptographic security protocols, data structures, consensus mechanisms, smart contracts, blockchain-based agriculture applications, and the possible solutions for the problems faced. Q. Yung et al.[67] presented a smart blockchain scheme for the IoT enabled smart homes specifically focusing on transactive energy management by enhancing the overall performance of smart meters in reducing the power losses, privacy concerns, and safeguards for sudden failures due to centralized systems. Fernández-Caramés et al.[18] stressed applying blockchain technology in many use cases.

Blockchain technologies can be applied in different areas where IoT applications are involved like supply chain management (SCM), smart living, mobile crowdsensing, cyber law, and security in mission-critical scenarios. Blockchain can also be used in IoT agriculture applications. One working example of it is tracking Chinese agri-food supplies, where the main aim is to enhance food safety, and quality and reduce the logistic cost. Blockchain technology has also remarkable imprints in the field of healthcare mainly in clinical trials and precision medicine, smart grids, and smart cities. Aya Reyna et al.[49] mentioned blockchain applications mostly occurring in the field of online financial transactions where many cryptocurrencies are available like Bitcoin,

Litecoin, Namecoin, etc. Furthermore, blockchain technology can be an ideal solution for traceability applications. Many applications that use IoT devices to digitize the sensed data can be improved with blockchain technology. Blockchain can further make the smart city a reality by incorporating security and privacy-based smart contracts without any third-party centralized services [46]. Blockchain can also be incorporated into the Energy sector, Insurance, and Mortgage based IoT applications with the heterogeneous involvement of various stakeholders in a distributed manner [29]. Mainly addressing the important challenges in integrating blockchain technology with IoT applications will pave more directions for its feasible implementation with far better results. M. S. Ali et al. [5] have not restricted the use of blockchain technology only to online financial transactions but significant scientific research is carried out in the field of the Internet of Things.

Blockchain technology has shown promising imprints in several industries like smart insurance, smart grids, healthcare, SCM, smart home applications, and connected vehicle fleet network. The main motive of this survey is to make an attempt by bringing blockchain technology to its maturity in leveraging IoT by considering the challenges faced in its integration. Wang Xu et al.[64] discussed the IoT security challenges faced by centralized service mechanisms, as the IoT devices are data-centric, which generate huge amounts of data that can be easily compromised by potential attackers. Thus blockchain is considered to be an emerging technology that can be inquired with IoT networks to provide security, data integrity, and reliability. Blockchain show promising application areas with IoT in the field of smart grids, nuclear factory, SCM, and smart cities. The decentralized nature of blockchain makes it more feasible for IoT networks but there are underlying technological challenges in its integration, which need to be addressed for future large-scale high capacity IoT applications. Ali Dorri et al.[30] highlighted that enormous research and innovation were carried out in order to implement blockchain in IoT networks. Mostly the use cases take advantage of blockchain features like decentralized control, security, fault tolerance, data integrity, immutability, and the ability to run smart contracts. The applications of blockchain with IoT mainly discussed in this paper are autonomous decentralized peer-to-peer telemetry, secured smart cities, secure firmware updates for the devices to run securely on regular basis, smart home architecture, self-managed vehicle ad-hoc networks, and SCM. As these use cases are in discussion for example in the case of smart cities it's not clear which blockchain platform, consensus protocol, and transaction validation technique are to be implemented. Daniel et al.[35] have given more importance to the cryptographic security and decentralized nature of blockchain technology, for this reason, blockchain technology is resistant to false modifications of underlying information. Various applications have evolved because of their distinguished capabilities like trustworthy contracts between different participating parties, storing information securely, and transferring money safely without any third-party control. One more application mentioned is smart logistic management services in which trigger-based invoices pay themselves when a shipment is received by the user.

Blockchain can also be applicable in IoT environments for device configuration storing sensor data, and performing micro-payments [50]. supply chain provenance [22], software push to remote IoT devices [51]. Muneeb et al.[19] highlighted anonymity as the best to maintain privacy in IoT networks which led to the applications like electronic patient record management. In financial transaction management using blockchain must secure the personal information of the peers which means maintaining anonymity based on two strategies named ZeroCoin [34] and ZeroCash [52]. Application areas involving the trading of energy resources show promising results from the research perspective when leveraging the blockchain with IoT-based networks. In order to maintain the immutable transactional flow transparently throughout the supply chain from producer to consumer via smart contracts which can perform automatic billing invoices and micro-payments as well [23, 58, 41].

The blockchain-based IoT systems are still vulnerable to some privacy issues that need to be resolved before final implementation practically. Thomas B. et al.[8] presented a detailed use case of blockchain in the pharma SCM. Project-based on sensor devices in IoT considers environmental control parameters like temperature and humidity in the most critical use case of the medical supply chain to ensure quality control and company regulatory norms as per international standards. the sensor nodes attached continuously monitor the temperature of each shipment item, then the data is transferred to the blockchain based on smart contact's assessment decisions made to ensure the safety of the patients using the medicines. Ali Dorri et al.[15] considered the security and privacy of IoT networks as the prime attributes thus integrating the blockchain with the IoT systems mainly in the application area of smart homes. As blockchain is a decentralized network and uses

cryptographic hash functions to secure the data shared in the form of chained blocks to maintain security and privacy. As IoT devices are mostly resource-constrained in nature thus it will be impracticable to use the energy-consuming proof of work(PoW). Thus the blockchain-based smart home consists of three main entities: cloud storage, overlay, and smart home. The smart home is embedded with the high resource device called a miner which is responsible for handling the block's authentication and authorization into the blockchain. It has shown improving results. But there are still some overheads in terms of traffic, processing time, and energy consumption which needs proper justification to compromise a certain extent of these overheads with this proposed technology.

The above study summarizes the diversified application domains of both technologies and their integration will have a significant impact on real-time applications considering some challenges for further improvements. The blockchain approach needs to be augmented by robust and scalable architectures.

2.3. Blockchain Architectures. The commencing notability of blockchain technology in its inception with the existing technologies and applications was relying on its distributed architecture with security, transparency, immutability, and traceability. The IoT ecosystems face serious challenges in security and distributed architectural framework needs, which are mostly based on traditional centralized servers. Therefore, the need for the combination of IoT with blockchain technology will have satisfying results [63]. O. Nova [40] incorporated blockchain framework for IoT-based networks. The access control mechanism in the architecture is fully distributed which removes any dependence on single centralized servers. Syed T. Ali et al. [59] carried out a comparative analysis of blockchain technology to be incorporated into IoT. The inbuilt features like security, decentralization, immutability, transparency, and privacy in blockchain technology manage various IoT-based critical applications as a backbone network. H. Bai et al. [6] proposed a two-layer consensus-based architecture for IoT considering the resource constraints of the IoT devices. The paper lacks details about the selection of top-layer nodes as servers and storage of a complete chain of blocks in the network which further limits the decentralization. C. K. Pyoung et al. [44] proposed LiTiChain, a scalable and lightweight blockchain-based IoT architecture that is based on a finite lifetime of blocks in the network. Thus reducing the storage and scalable problem in massively deployed IoT networks. P.K. Sharma et al. [56] proposed a blockchain-based integration of cloud and fog services architecture for IoT applications. It comprises of three layers: the device layer of IoT devices, Fog-based distributed blockchain subnetworks, and the blockchain-based cloud layer. Wattana Viriyasitavat et al. [62] proposed an interoperable and trust-based blockchain framework for service-oriented IoT networks. Ali Dorri et al. [15] proposed a blockchain-based smart home architectural framework for secured communication and management of information remotely in device-to-device interactions. M. Pourvahab et al. [43] enhanced the security mechanisms of existing software-defined networking based IoT systems by using blockchain technology as the most promising peer-to-peer distributed framework. The blockchain-based layered architecture possesses various rules for the control and validation of blocks in the network. The Chain of Custody (CoC) consensus mechanism is specifically utilized for digital forensics as a use case. S. K. Singh et al. [57] proposed a state of art framework which includes technologies like blockchain, and artificial intelligence integrated with IoT systems. Blockchain enhances the security, trust, and transparency of existing intelligent IoT systems via decentralized architectures and removes the single point of failure in big data analysis. Y. Qian et al. [45] proposed a blockchain-based security management architecture for IoT devices. Blockchain helps in maintaining security services, operations, and software firmware updates throughout the life cycle of all IoT devices. Any type of threat carried in the system can be easily detected and traced. K. Prescilla et al. [24] focused on the resource-constrained nature of the IoT devices and centralized servers are prone to single-point failures in existing IoT networks. Considering the decentralized manner of blockchain technology proposed sliding window-based blockchain architecture for IoT applications uses the most complex PoW consensus mechanism for resource-constrained IoT devices.

Besides having numerous developments in the integration of blockchain-based IoT architectures. There is still a need for the most generalized BIoT architecture which will cater to security, privacy, and interoperability with the scale of the IoT devices. BIoT architectures are extended even further to the foundation of blockchain technology, and consensus procedures, which are covered in the following subsection.

2.4. Consensus Mechanisms. The core executing environment in blockchain technology is consensus mechanisms. They are predefined computational logics that come into the execution in peer-to-peer distributed

Table 2.1: Comparative analysis of distinctive areas in leveraging Blockchain and IoT

Research Domain	Paper [Year]	Contribution	Advantages	Limitations	Future Intentions
Real-Time BIoT Applications	[12] [2022]	BIoT	security and autonomy of smart object	Authentication of nodes	Improving computation of IoT devices using ML
	[48] [2021]	BIoT	storing, retrieving, managing Healthcare data	actual implementation missing	implementation of proposed model
	[20] [2020]	Blockchain based QR cheque system	Secure authentication scheme	Limited to cheque only feature in banks	complete bank solution
	[35] [2018]	BIoT	Security highlights	Implementation	Optimal deployment
	[27] [2021]	BIoT	Security solution areas	Practical implementation	smart contract driven solutions
	[68] [2021]	Fruit and vegetable SCM	on-off chain mechanisms	limited use-case approach	consensus mechanisms
BIoT Architectures	[40] [2018]	First BIoT architecture	constrained IoT specific, secured, and better scalability	public access	Application use case
	[44] [2020]	LiTiChain finite state based BIoT	better scalability	Cost overhead	permissionless version of erasable blockchains
	[57] [2019]	AI enabled BIoT	efficient mechanism in terms of security, Bigdata analytics, and low latency	limited use	wider dimensions given
	[16] [2021]	BIoT	Architecture with elastic smart contract, Smart city	Addressing security concerns on deployment	machine learning and transaction cost
Consensus Mechanisms(CM)	[69] [2020]	CMs comparative analysis	CMs categorizes in probabilistic-finality, absolute-finality	no idea proposed	Fault tolerant based CM with applications
	[39] [2018]	Types of CMs	Proof, voting criteria	limited comparison	Not available
	[65] [2020]	CMs detailed survey with vulnerability analysis	Analysis of Performance metrics	application specific	Futuristic CM design
	[66] [2021]	Enhanced CM for D2D communication	Reliability and performance	Issue with limited Lightweight device	Further enhancing for generalized BC system

communication as an agreement between the participating nodes while maintaining security, privacy, transparency, and immutability throughout the life cycle of the network. S. Nakamoto [38] initially proposed the concept of using blockchain technology for creating virtual currency such as Bitcoin. They utilized Proof of Work (PoW) as the main consensus mechanism for the fulfillment of agreement and validation of peers. PoW uses most of the computational power and high latency thus is not feasible for IoT systems. Proof of Stake(PoS) [9] is an improvement over PoW with less computational cost [49, 5]. Practical Byzantine Fault Tolerance (PBFT) [11, 5, 13] is more efficient and applicable than PoW and PoS. Proof of Activity (PoA) [7] is another consensus protocol based on PoW and PoS. Delegated Proof-of-Stake (DPoS) is representative democratic while PoS was directly democratic. Delegates are selected for the creation and validation of new blocks. DPoS is not totally decentralized but block finality is faster as compared to PoS [70]. Stellar consensus protocol (SCP) [33] is a variant of Byzantine agreement protocol. SCP gives the participants a valid choice to select the participants from a group of trusted nodes[70]. The participating node takes a decision based on the consensus of

Table 2.2: Selective IoT-based Consensus mechanisms

Consensus Mechanism	Scalability(IoT Networks)	Latency	Throughput	Computational Cost	Network Overhead
PBFT	Small Scale IoT	Low	High	Low	High
DPoS	Large Scale IoT	Low	High	Low	Low
PoET	Large Scale IoT	Low	High	Low	Low
PoI	Large Scale IoT	Low	High	Low	Low
SCP	Large Scale IoT	Low	High	Low	High
Ripple	Large Scale IoT	Low	High	Low	High
Tendermint	Large Scale IoT	Low	High	Low	Low
SDTE	Small Scale IoT	High	Low	High	High
PLEDGE	Small Scale IoT	Low	High	Low	Low

its trusted circle. Delegated Byzantine Fault Tolerance (DBFT) is a proxy voting-based consensus mechanism. Some nodes are chosen as delegates or bookkeepers, which maintain the digital ledger after a valid agreement is achieved to add a new block [30, 42]. Ripple [53] is a consensus mechanism that organizes the nodes in the form of trusted clusters within the larger network. There are two types of nodes in the network, a server for carrying the consensus process and a client for only transferring funds. Every cluster server stores a Unique Node List (UNL). In order to write a new block in the ledger, the nodes in UNL with almost 80 % majority agreement are mandatory for adding a new block in the ledger which reduces the chances of attacker nodes. Ripple is more scalable, thus favorable for IoT devices that only store the previous and new balance with no monetary rewards. Tendermint [25, 30, 5, 70] a variant of BFT consensus mechanism. It utilizes some features of PoS and PBFT to enhance high throughput, security, and less block finality time. Block is initiated by a proposer. The proposer node is in turn selected by round robin procedure by dedicated validators without any mining process. Nodes need to deposit their coins to act as validators which increases the security by punishing the dishonest nodes. Secure blockchain-based Data Trading Ecosystem (SDTE) [14] is a new consensus mechanism for data trading e-markets. In SDTE protocol, the broker and the buyer can interact with limited analysis mostly restricted to the personal data of the seller. It runs a three-way trusted secured communication protocol between the seller, buyer, and the trusted nodes. The SDTE is more secure against malicious contracts and prevents data theft and fraud. Proof of Honesty (PLEDGE) [32] is a secured consensus protocol having low transaction latency and computational cost. PLEDGE reduces the reputations and contribution of malicious and non-performing nodes in the consensus mechanism.

In Table 2.2, we have selected some of the promising consensus protocols feasible for IoT-based networks considering the scalable nature of the devices with highlights of performance metrics in terms of latency, throughput, computational cost, and network overhead of these protocols.

From the above discussion, we concluded that there are different numbers of consensus mechanisms in the literature applied to blockchain applications. Most of the consensus mechanisms are more economical for financial services. Relation to the usage of blockchain in IoT applications should fulfill some additional constraints already present in IoT systems. IoT devices are mostly connected in a distributed way and have limited commuting and communicating capabilities. The most promising consensus algorithms which are presently used in blockchain-based applications are impractical for IoT networks. However, in relation to blockchain architectures for IoT systems, some consensus algorithms may be effective in comparison to others. The reason stems from the intrinsic characteristics of blockchain architectures. In Table 2.1, we have highlighted some recent research papers based on the criteria of contribution, advantages, limitations, and future intentions in the research domains of applications, architectures, and consensus mechanisms. The next section presents our proposed blockchain architecture and also discusses the effectiveness of some consensus mechanisms in the proposed architecture.

3. Proposed Work. In this section, we will discuss about our detailed work including various dimensional attributes in leveraging blockchain with IoT-based system applications. Initially, we proposed a BIoT architecture for generalized real-time application use cases. Secondly, we summarized customized consensus mechanisms

specifically feasible for IoT-based applications. And lastly, we have discussed the real-time application use in focusing on SCM(India).

3.1. Proposed BIoT architecture. Our proposed novel blockchain-based IoT architecture (BIoT) addresses the security, failure, resilience, and many more issues associated with existing state-of-the-art IoT systems. The proposed architecture is a fusion of blockchain technology with IoT systems. In this novel architecture, the blockchain acts as the main backbone of the IoT systems, implementing security, transparency, immutability, auditability, and fault tolerance in a distributed fashion. Our architecture also incorporates the provision for the optimized consensus mechanisms for the different IoT-based applications depending on the number of IoT devices deployed in the whole network. Let us discuss the proposed architecture in detail. Figure 3.1 shows our proposed architecture in detail. The 3-tier architecture comprises of IoT device layer, the blockchain (BC) layer, and the application layer. These three layers are briefly discussed as follows.

The IoT device layer comprises the actual IoT devices which are deployed for the application-specific operation. IoT devices are mostly limited in resources. As these devices have to deal with the BC framework, a cluster of resource-efficient edge gateways will handle the computation and communication workloads of the distributed peer-to-peer network. IoT end devices are incapable of retaining the complete block of information from the blockchain ledger. However, IoT devices will store the block header information in order to maintain transparency and traceability throughout the communication network. The edge gateways are responsible enough to act as participating peers in the blockchain network. These edge gateways directly take part in the consensus mechanisms as miners or coordinators for achieving the agreement between the participating peer nodes in the distributed blockchain network. After the selection by the consensus mechanism and authorization by the certificate agency mostly government-based (eGov). The edge gateway acting as a miner also will broadcast it to the neighboring IoT nodes in the cluster. The IoT nodes will acknowledge the corresponding edge gateway as their cluster head and forward all the data directly to acknowledged cluster heads.

The blockchain layer logically comprises the main peer-to-peer BC framework and the BC access control layer. Since there is no central authority in the blockchain network, all the nodes are eligible in taking part in joining the network. The edge gateways from the IoT layer are capable of handling the computation and communication costs in the distributed BC network. The edge gateways are considered as main peer nodes in the making of a blockchain network. We have also proposed a consensus selector mechanism to select the optimal consensus mechanism which is explored in the next section. In the case of the private blockchain application security and privacy is maintained as only the specified parties are allowed to participate and carry transactions with the existing network. In our approach, some portion of the information excluding personal data in the form of analytical patterns, graphs, etc. can be uploaded to a cloud data store for business research analysis. The BC access control layer manages the creation and deployment of smart contracts, government approved certificate authority which maintains the authentication and authorization of all the participating parties in a decentralized manner. The certificate authority generates tokens as digital signatures which are further associated with the nodes for their validity in adding blocks to the distributed ledger. This layer also manages role-based policy for nodes on the basis of computing and communication capabilities and is differentiated as light IoT nodes and full nodes.

The application layer directly represents the information retrieval mechanism via different interfaces like web portals, dashboards, and decentralized applications (DApps). These are numerous IoT application uses cases with blockchain technology where transparency, security, and immutability are in a distributed manner without any third party.

The key features of the proposed BIoT architecture:

- Our model extends the existing 3-tier generalized IoT architecture with blockchain integration in a simplified manner highlighting some promising application use cases for its early adoption.
- The IoT device layer takes care of the constrained-IoT devices which simply capture the data and forward it to their corresponding Intelligent edge gateway. The information is stored in the blockchain network via full nodes by preserving the security and privacy of all nodes in a distributed and transparent manner.
- The BC layer controls the main logic of the entire system model by invoking the pluggable consensus mechanism with autonomous smart contracts among the stakeholders in agreed predefined business pro-

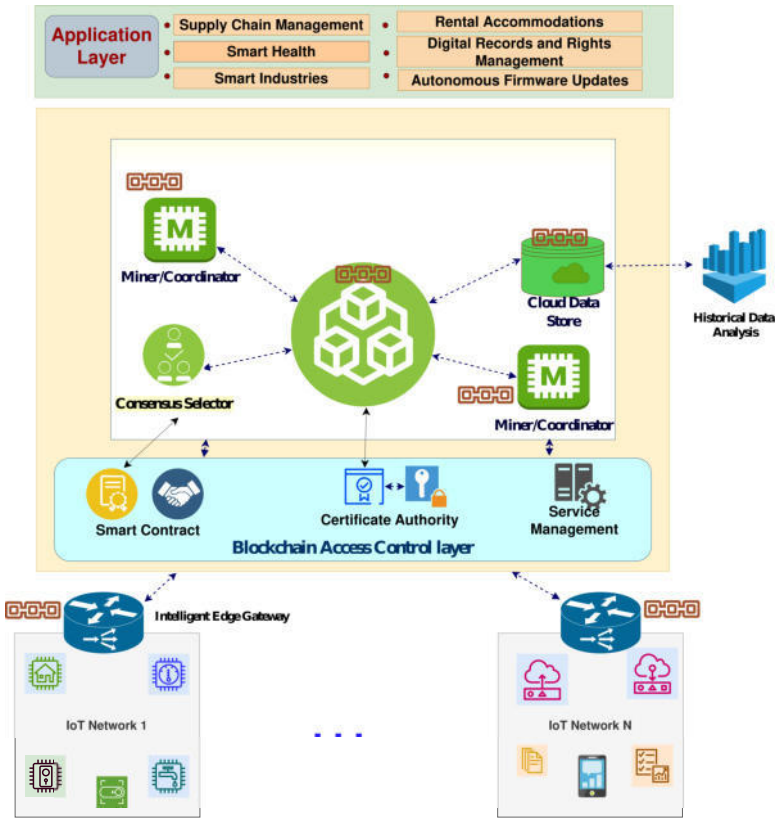


Fig. 3.1: Proposed BIoT architecture.

cedures. The distinctive feature of this layer includes the certified agency for verifying the authenticity of the registrants in the network.

3.2. Selective Consensus Mechanisms for IoT. The consensus mechanism is the core of blockchain technology in its integration of IoT ecosystems. Consensus mechanisms make the blockchain a resilient distributed and transparent mode of data storage and communication in real-time. As from the literature, there is a number of consensus mechanisms followed at different blockchain networks in varied applications. In the case of IoT-based applications, the limit of using consensus mechanisms is restricted because of resource constraints in terms of computation and communication capabilities.

Our idea of designing a customized container of optimal consensus mechanisms for the integration of blockchain-based IoT applications is motivated by the complexity of choosing among numerous consensus protocols available in the field of blockchain technology for finance, cryptocurrency, and other application services. Since IoT networks are densely deployed with IoT end devices. The scale of the number of devices varies in IoT-based applications. Thus there must be a simplified mechanism for selecting among the feasible consensus mechanisms in the framework specifically meant for IoT networks with optimal parameters like scalability, latency, throughput, computational cost, and network overhead.

The key attributes in our algorithm comprise of Z = (threshold value of the max range for small IoT network), IoT subsystems = IoT_n, I_n = number of IoT nodes in the IoT subsystem, S[x] = contains small scale consensus mechanisms, L[y] = contains large scale consensus mechanisms and selected-CM = final consensus mechanism selected. In a blockchain network, the participating peers engage in distributed peer-to-peer communication fashion. The edge gateways will act as participating peers in order to be part of the network and add their corresponding blocks in the immutable distributed ledger of blockchain after successful computational

operations. Depending on the feasible parameters required by the particular IoT-based application, the most favorable consensus mechanism will be selected and finally implemented in order to maintain the prerequisite properties of the main blockchain network. Then accordingly the edge gateways will act as miners or coordinators in different consensus mechanisms. The IoT nodes will simply follow a hierarchical tree form topology, where the edge gate will act as a cluster head for the particular IoT subsystem. Since all the IoT devices rely on cluster head gateway for their further data communication. Most of the IoT nodes will be reluctant in taking part in the mining process of the blocks, thus local copies of the block ledger will be maintained only at the level of edge gateways. With the help of this enabled procedure in the BIoT architectural framework, the work will become easier in distributed communication with resilient features of security, transparency, and high throughput in the majority of IoT applications.

Algorithm 1: Selective Consensus mechanism for IoT network

Input: Z, S[x], L[y]
Output: Selected CM
initialization;
while $IoT_n! = 0$ **do**
  check for the scale of IoT network;
  if $n^* \sum I_n >= Z$ **then**
    L[y] = [DPos, PoET, PoI, SCP, Ripple, Tendermint];
    selected-CM = L[y];
  else
    S[x] = L[y] $\bigcup$ [PBFT, SDTE, PLEDGE] ;
    selected-CM = S[x] ;
  end
end

4. Case study: Blockchain based Saffron Agri-Value Chain. Since agriculture sector is the backbone of the Indian economy, as it comprises the main pillar of India's GDP (Gross Domestic Product). We focused on ways to improvise agriculture and its allied service sectors. Supply Chain Management (SCM) is the streamlined flow of products and their data exchanges throughout the communication link between various parties in the network. The main objective of the enhancement in terms of innovation and technology in the field of SCM in the Indian agricultural sector is to maintain a transparent processing and distribution system of generating value for all starting from farmers to consumers. BC technology's integration with SCM has sufficed to maintain prerequisite features like security without any third party, transparency, data integrity, minimal risk, cost reduction, immutability, and traceability. Blockchain technology improvises the social sustainable development in SCM [17, 61]. The COVID-19 pandemic drastically highlighted the loopholes in existing SCM throughout the world and even the World Economic Forum report highlighted the need for integration of distributed and transparent technology like blockchain technology in SCM for smooth transparency and traceability with negligible reliance on third-party servers.

To understand the importance of BC in SCM in the agriculture sector. We tried to integrate the various stakeholders involved in the agriculture SCM. The main motive for integrating IoT; which will monitor real-time data via IoT devices and BC technology; acting as the backbone network which will preserve the transparency, provenance, and security of the data received by the IoT device in a permanent distributed leader. Let us take the case of saffron marketing from production, processing, and distribution. Saffron is considered to be the costliest spice grown in the world and cultivated in the northern region of India, particularly in Pampore, Jammu and Kashmir. The real challenge which the cultivators, as well as the consumers, are facing in its distribution and consumption is the availability of fake and adulterated variants of saffron in the market. Integrating the BC technology with the existing saffron marketing strategies helps in maintaining transparency and traceability throughout the saffron distribution life cycle in the whole ecosystem.

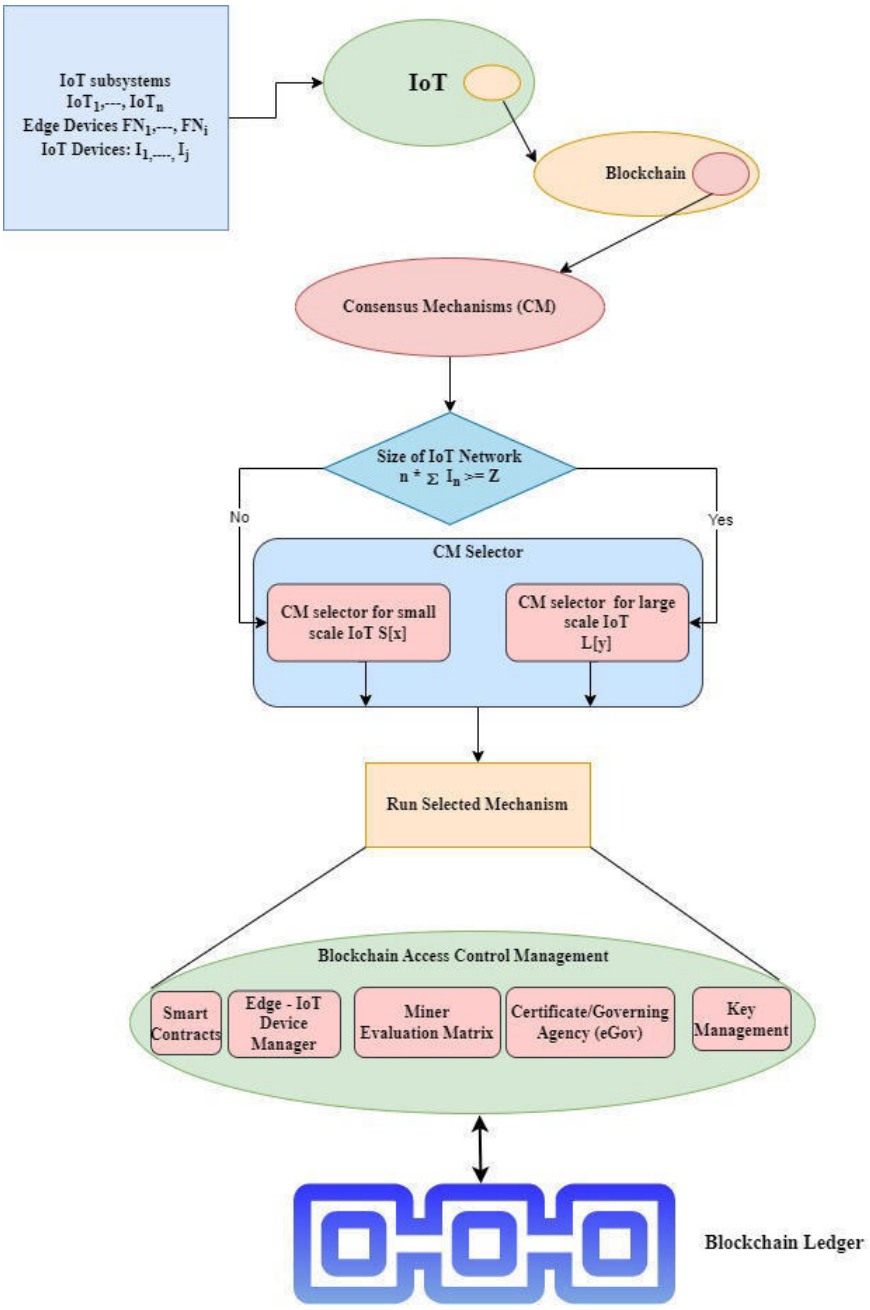


Fig. 3.2: Proposed Selective Consensus Mechanisms for BIoT Application.

Assumptions in the system model

1. BC type: Permissioned BC.
2. Light-weight IoT objects like sensors, Arduino, and Raspberry pi.
3. Full nodes- Edge gateways, Laptop, PC, Workstations.

4.1. Work Flow in Saffron-Agri Value Chain. A detailed graphical model is proposed as depicted in figure 4.1. The system model comprises real-world entities: raw material suppliers, saffron growers, government

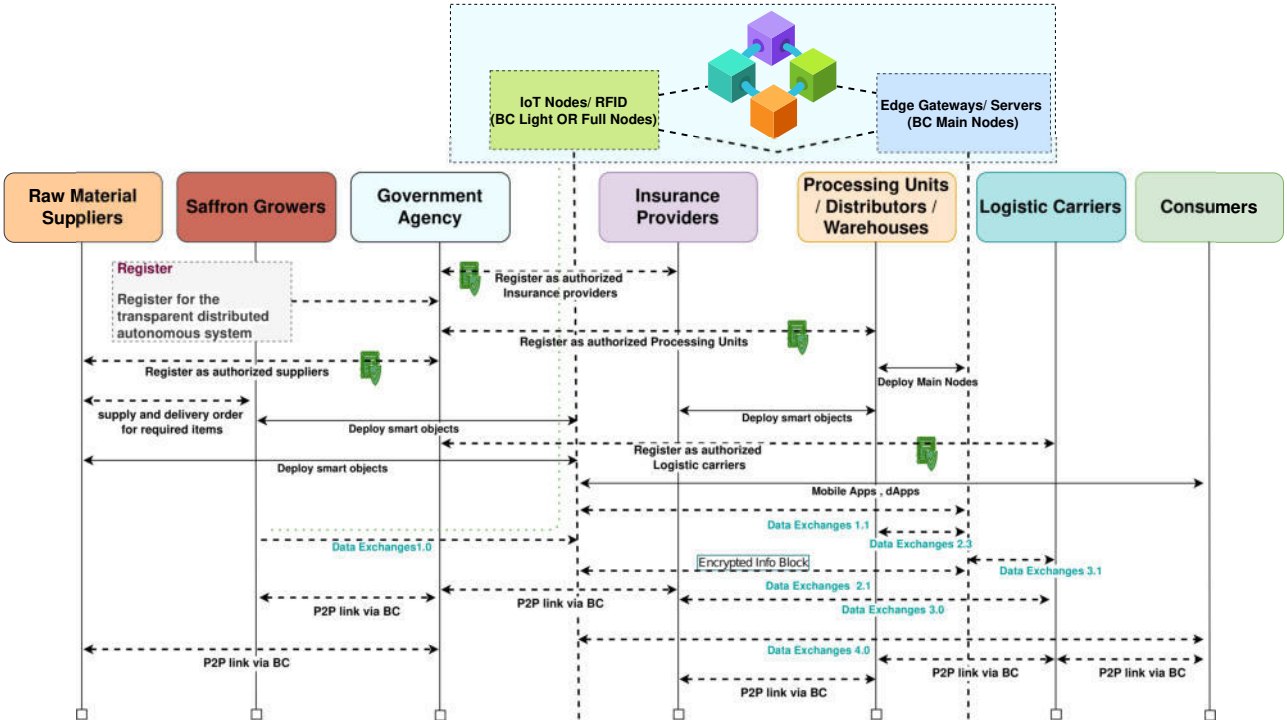


Fig. 4.1: Graphical representation of BIoT enabled Saffron-Agri Value Supply Chain in J&K, India.

market agencies, logistic carriers, insurance companies, processing units, warehouses, distributors, and end-user consumers. And the virtual objects from a technological perspective include Lightweight IoT nodes, RFID tags, and full nodes comprised of desktops, laptops, Raspberry Pi nodes, routers, and switches. The IoT networks seamlessly monitor the SCM from the farmer to the end consumer but with less security and transparency due to the main dependency on third-party services. The blockchain transaction process is based on mutual consensus in a peer-to-peer fashion and data blocks are cryptographically hashed with the previous data exchanges within the network. BC acts as the main source of proof for actual status and information throughout the network life-cycle in a transparent way. The transactional flow of data exchanges between different subsidiaries in the SCM of saffron farming initiates from the cultivating field of saffron. The saffron grower initially registers with the government agriculture department for authentication and future communication. The farmer will arrange the required raw material from government-approved suppliers. The data interactions at the actual saffron field will be captured by IoT objects and forwarded to the edge gateways. The edge gateways will act as main nodes to store the complete copy of the blockchain in their local memory. The participating parties in the system will perform autonomous transactions based on the access control management roles assigned to different peers in the network. The peers perform transactions based on the consensus mechanism implemented in the blockchain framework network. The autonomous programmable codes are created and deployed as smart contracts in BC. We focus on using the PBFT algorithm for the mutual consensus to be achieved between all participating peers in a decentralized manner. PBFT algorithm is highly efficient than PoW, PoS, and its other variants. PBFT algorithm was successfully implemented in the Hyperledger Fabric BC framework.

After authentication and authorization, any peer node can add its block to the existing ledger of transactional blocks in the blockchain network. Once the blocks are added to the chain of blocks, it is impossible to change the data block even not by the creator of the data block. As a result, all the data exchanges that are to be carried from the farmer to the consumer will be preserved efficiently and transparently.

Fig. 4.2: Prototype of test-bed network of BIoT setup

Table 4.1: Comparative analysis of existing SCM and proposed BIoT-based SCM

Index	Existing SCM	BIoT based SCM	Impact
Network Service	Centralized	Distributed	No third party
Fault Tolerance	Single point of failure	High	Replicas at multiple peers
Transparency	Low	High	Data Availability in the SC
Government agency	Least involvement	Active involvement	Report for any fraud etc
Reliability	Low	High	P2P interactions via smart contracts
Immutability	Low	High	Immutable transactions
Security	Low priority	Inbuilt hashing	Trust-less behaviour
Scalability	Low	High	IoT specific CMs
Time Saving	Low	High	Stakeholders are connected in P2P way
Cost Reduction	Low	High	Negligible monopoly of third party

4.2. Analogy of result metrics considering proposed system. The performance of the proposed BIoT architecture with selective consensus protocols in the Saffron-Agri value chain is based on our laboratory experimental simulation with limited IoT devices as light nodes and edge nodes as the full nodes. We have chosen Hyperledger Fabric, a permissioned private blockchain most favorable for our application use case. There are no transaction fees involved on any block to be added to the ledger. Any stakeholder entity needs to get registered with a government agency (eGov), which acts as the main certifying authority in the network.

The configuration of the main node comprises (CPU:i7 @3.2GHz, RAM:16GB, Storage:512GB) and one main distributed network kit with five nodes (Main Node: Quad-core @1.2GHz, RAM:2GB, OS: Ubuntu MATE 16.04 and 4 Sub-Nodes Quad-core @1.6GHz, RAM:1GB, OS: Debian 8). Light nodes comprise Arduino Uno integrated with limited sensors. Repeated tests were carried out in which the light nodes only capture data and forward it to the nearby edge nodes. The edge nodes take active participation in the block generation and storing in its database. The prototype of the test-bed network is shown in Figure 4.2. In Table 4.1 we have evaluated the comparison of our proposed scheme against the traditional approaches for different IoT-specific parameters. Our approach outperforms the traditional approach in almost all the comparison indices.

5. Conclusion. This paper focuses on enhancing security and transparency in the much-highlighted technology of the present industrial revolution i.e IoT by integrating with blockchain technology. We proposed a

BIoT architecture to incorporate BC and IoT together with a simplified access control mechanism. We have also designed a strategy for the selection of consensus mechanisms for the various IoT network applications based on determining factors like scalability, latency, throughput, computational cost, and network overhead. Our proposed model is based on a permissioned blockchain network. The advantage of using permissioned blockchain maintains limited access controls for unknown users and provides certain checks for the usage of information for different participating peers in the network. We have also discussed the case study of the proposed model in the area of the agriculture sector specifically for maintaining the transparency and traceability of the Saffron crops throughout the supply chain from farmer to the end-user. Integrating BC technology in the agricultural sector of the Saffron value chain cultivated in J&K, India, will certainly enhance provenance for the end consumers throughout the chain. It also improvises updating autonomously in various enabled services like licensing, products, and firmwares. We have implemented an initial prototype integrating blockchain with IoT devices. The preliminary results suggest that the proposed scheme implemented on SCM in the Saffron-Agri value chain outperforms the existing systems in SCM.

6. Future Perspectives. The work can be extended to other valued Agri- chain systems with modifications in their production and processing. The work can be enhanced with certain modifications for enhanced real-time track and trace other than Agri-value chain systems like handicrafts, etc.

REFERENCES

[1] R. Agrawal, P. Verma, R. Sonanis, U. Goel, A. De, S. A. Kondaveeti, and S. Shekhar, *Continuous security in iot using blockchain*, in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 6423–6427.

[2] M. M. Ahemd, M. A. Shah, and A. Wahid, *Iot security: A layered approach for attacks & defenses*, in 2017 International Conference on Communication Technologies (ComTech), IEEE, 2017, pp. 104–110.

[3] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, *Internet of things: A survey on enabling technologies, protocols, and applications*, IEEE communications surveys & tutorials, 17 (2015), pp. 2347–2376.

[4] J. Ali and S. Sofi, *Ensuring security and transparency in distributed communication in iot ecosystems using blockchain technology: Protocols, applications and challenges*, International Journal of Computing and Digital System, (2021).

[5] M. S. Ali, M. Vecchio, M. Pincheira, K. Dolui, F. Antonelli, and M. H. Rehmani, *Applications of blockchains in the internet of things: A comprehensive survey*, IEEE Communications Surveys Tutorials, 21 (2019), pp. 1676–1717.

[6] H. Bai, G. Xia, and S. Fu, *A two-layer-consensus based blockchain architecture for iot*, in 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), 2019, pp. 1–6.

[7] I. Bentov, C. Lee, A. Mizrahi, and M. Rosenfeld, *Proof of activity: Extending bitcoin's proof of work via proof of stake [extended abstract] y*, ACM SIGMETRICS Performance Evaluation Review, 42 (2014), pp. 34–37.

[8] T. Bocek, B. B. Rodrigues, T. Strasser, and B. Stiller, *Blockchains everywhere-a use-case of blockchains in the pharma supply-chain*, in 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), IEEE, 2017, pp. 772–777.

[9] V. Buterin, *What proof of stake is and why it matters*, Bitcoin Magazine, 26 (2013).

[10] R. Casado-Vara, J. Prieto, F. De la Prieta, and J. M. Corchado, *How blockchain improves the supply chain: Case study alimentary supply chain*, Procedia computer science, 134 (2018), pp. 393–398.

[11] M. Castro, B. Liskov, et al., *Practical byzantine fault tolerance*, in OSDI, vol. 99, 1999, pp. 173–186.

[12] E. Corradini, S. Nicolazzo, A. Nocera, D. Ursino, and L. Virgili, *A two-tier blockchain framework to increase protection and autonomy of smart objects in the iot*, Computer Communications, 181 (2022), pp. 338–356.

[13] L. Da Xu and W. Viriyasitavat, *Application of blockchain in collaborative internet-of-things services*, IEEE Transactions on Computational Social Systems, 6 (2019), pp. 1295–1305.

[14] W. Dai, C. Dai, K.-K. R. Choo, C. Cui, D. Zou, and H. Jin, *Sdte: A secure blockchain-based data trading ecosystem*, IEEE Transactions on Information Forensics and Security, 15 (2020), pp. 725–737.

[15] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, *Blockchain for iot security and privacy: The case study of a smart home*, in 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops), IEEE, 2017, pp. 618–623.

[16] S. Dustdar, P. Fernández, J. M. García, and A. Ruiz-Cortés, *Elastic smart contracts in blockchains*, IEEE/CAA Journal of Automatica Sinica, 8 (2021), pp. 1901–1912.

[17] B. Esmaeilian, J. Sarkis, K. Lewis, and S. Behdad, *Blockchain for the future of sustainable supply chain management in industry 4.0*, Resources, Conservation and Recycling, 163 (2020), p. 105064.

[18] T. M. Fernández-Caramés and P. Fraga-Lamas, *A review on the use of blockchain for the internet of things*, IEEE Access, 6 (2018), pp. 32979–33001.

[19] M. U. Hassan, M. H. Rehmani, and J. Chen, *Privacy preservation in blockchain based iot systems: Integration issues, prospects, challenges, and future research directions*, Future Generation Computer Systems, (2019).

[20] N. Kabra, P. Bhattacharya, S. Tanwar, and S. Tyagi, *Mudrachain: Blockchain-based framework for automated cheque clearance in financial institutions*, Future Generation Computer Systems, 102 (2020), pp. 574–587.
[21] M. A. Khan and K. Salah, *Iot security: Review, blockchain solutions, and open challenges*, Future Generation Computer Systems, 82 (2018), pp. 395–411.
[22] H. M. Kim and M. Laskowski, *Toward an ontology-driven blockchain design for supply-chain provenance*, Intelligent Systems in Accounting, Finance and Management, 25 (2018), pp. 18–27.
[23] D. Kirli, B. Couraud, V. Robu, M. Salgado-Bravo, S. Norbu, M. Andoni, I. Antonopoulos, M. Negrete-Pincetic, D. Flynn, and A. Kiprakis, *Smart contracts in energy systems: A systematic review of fundamental approaches and implementations*, Renewable and Sustainable Energy Reviews, 158 (2022), p. 112013.
[24] P. Koshy, S. Babu, and B. Manoj, *Sliding window blockchain architecture for internet of things*, IEEE Internet of Things Journal, 7 (2020), pp. 3338–3348.
[25] J. Kwon, *Tendermint: Consensus without mining*, Draft v. 0.6, fall, 1 (2014).
[26] W. Lin, X. Huang, H. Fang, V. Wang, Y. Hua, J. Wang, H. Yin, D. Yi, and L. Yau, *Blockchain technology in current agricultural systems: from techniques to applications*, IEEE Access, 8 (2020), pp. 143920–143937.
[27] A. H. Lone and R. Naaz, *Applicability of blockchain smart contracts in securing internet and iot: a systematic literature review*, Computer Science Review, 39 (2021), p. 100360.
[28] K. Lounis and M. Zulkernine, *Attacks and defenses in short-range wireless technologies for iot*, IEEE Access, 8 (2020), pp. 88892–88932.
[29] Q. Luo, R. Liao, J. Li, X. Ye, and S. Chen, *Blockchain enabled credibility applications: Extant issues, frameworks and cases*, IEEE Access, 10 (2022), pp. 45759–45771.
[30] I. Makhdoom, M. Abolhasan, H. Abbas, and W. Ni, *Blockchain's adoption in iot: The challenges, and a way forward*, Journal of Network and Computer Applications, (2018).
[31] I. Makhdoom, M. Abolhasan, J. Lipman, R. P. Liu, and W. Ni, *Anatomy of threats to the internet of things*, IEEE Communications Surveys Tutorials, 21 (2019), pp. 1636–1675.
[32] I. Makhdoom, F. Tofigh, I. Zhou, M. Abolhasan, and J. Lipman, *Pledge: A proof-of-honesty based consensus protocol for blockchain-based iot systems*, in 2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 2020, pp. 1–3.
[33] D. Mazieres, *The stellar consensus protocol (whitepaper2015)," 11 2015.*
[34] I. Miers, C. Garman, M. Green, and A. D. Rubin, *"zerocoin: Anony- mous distributed e-cash from bitcoin, " in IEEE Symposium on Security and Privacy (SP)*, (2013), p. 397–411.
[35] D. Minoli and B. Occhiogrosso, *Blockchain mechanisms for iot security*, Internet of Things, 1 (2018), pp. 1–13.
[36] N. Mishra and S. Pandya, *Internet of things applications, security challenges, attacks, intrusion detection, and future visions: A systematic review*, IEEE Access, 9 (2021), pp. 59353–59377.
[37] B. K. Mohanta, D. Jena, S. Ramasubbareddy, M. Daneshmand, and A. H. Gandomi, *Addressing security and privacy issues of iot using blockchain technology*, IEEE Internet of Things Journal, 8 (2020), pp. 881–888.
[38] S. Nakamoto et al., *Bitcoin: A peer-to-peer electronic cash system*, (2008).
[39] G.-T. Nguyen and K. Kim, *A survey about consensus algorithms used in blockchain.*, Journal of Information processing systems, 14 (2018).
[40] O. Novo, *Blockchain meets iot: An architecture for scalable access management in iot*, IEEE Internet of Things Journal, 5 (2018), pp. 1184–1195.
[41] S.-V. Oprea, A. Bâra, and A. I. Andreescu, *Two novel blockchain-based market settlement mechanisms embedded into smart contracts for securely trading renewable energy*, IEEE access, 8 (2020), pp. 212548–212556.
[42] A. Panarello, N. Tapas, G. Merlino, F. Longo, and A. Puliafito, *Blockchain and iot integration: A systematic survey*, Sensors, 18 (2018), p. 2575.
[43] M. Pourvahab and G. Ekbatanifard, *An efficient forensics architecture in software-defined networking-iot using blockchain technology*, IEEE Access, 7 (2019), pp. 99573–99588.
[44] C. K. Pyoung and S. J. Baek, *Blockchain of finite-lifetime blocks with applications to edge-based iot*, IEEE Internet of Things Journal, 7 (2020), pp. 2102–2116.
[45] Y. Qian, Y. Jiang, J. Chen, Y. Zhang, J. Song, M. Zhou, and M. Pustišek, *Towards decentralized iot security enhancement: A blockchain approach*, Computers & Electrical Engineering, 72 (2018), pp. 266 – 273.
[46] M. A. Rahman, M. M. Rashid, M. S. Hossain, E. Hassanain, M. F. Alhamid, and M. Guizani, *Blockchain and iot-based cognitive edge framework for sharing economy services in a smart city*, IEEE Access, 7 (2019), pp. 18611–18621.
[47] S. B. Rane, S. V. Thakker, and R. Kant, *Stakeholders' involvement in green supply chain: a perspective of blockchain iot-integrated architecture*, Management of Environmental Quality: An International Journal, (2020).
[48] P. P. Ray, D. Dash, K. Salah, and N. Kumar, *Blockchain for iot-based healthcare: Background, consensus, platforms, and use cases*, IEEE Systems Journal, 15 (2021), pp. 85–94.
[49] A. Reyna, C. Martín, J. Chen, E. Soler, and M. Díaz, *On blockchain and its integration with iot. challenges and opportunities*, Future Generation Computer Systems, 88 (2018), pp. 173–190.
[50] M. Samaniego and R. Deters, *Blockchain as a service for iot*, in 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2016, pp. 433–436.
[51] M. Samaniego and R. Deters, *Using blockchain to push software-defined IoT components onto edge hosts*, in: Proceedings of the International Conference on Big Data and Advanced Wireless Technologies BDAW '16, Blagoevgrad, Bulgaria, 58 (2016).
[52] E. B. Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, *"zerocash: Decentralized anonymous*

payments from bitcoin, " in IEEE Symposium on Security and Privacy (SP), (2014), p. 459–474.
[53] D. Schwartz, N. Youngs, A. Britto, et al., *The ripple protocol consensus algorithm*, Ripple Labs Inc White Paper, 5 (2014).
[54] J. Sengupta, S. Ruj, and S. D. Bit, *A comprehensive survey on attacks, security issues and blockchain solutions for iot and iiot*, Journal of Network and Computer Applications, (2019), p. 102481.
[55] P. Sethi and S. R. Sarangi, *Internet of things: architectures, protocols, and applications*, Journal of Electrical and Computer Engineering, 2017 (2017).
[56] P. K. Sharma, M.-Y. Chen, and J. H. Park, *A software defined fog node based distributed blockchain cloud architecture for iot*, IEEE Access, 6 (2018), pp. 115–124.
[57] S. K. Singh, S. Rathore, and J. H. Park, *Blockiotintelligence: A blockchain-enabled intelligent iot architecture with artificial intelligence*, Future Generation Computer Systems, 110 (2020), pp. 721–743.
[58] J. G. Song, E. S. Kang, H. W. Shin, and J. W. Jang, *A smart contract-based p2p energy trading system with dynamic pricing on ethereum blockchain*, Sensors, 21 (2021), p. 1985.
[59] T. A. Syed, A. Alzahrani, S. Jan, M. S. Siddiqui, A. Nadeem, and T. Alghamdi, *A comparative analysis of blockchain architecture and its applications: Problems and recommendations*, IEEE Access, 7 (2019), pp. 176838–176869.
[60] P. Urien, *Blockchain iot (biot): A new direction for solving internet of things security and trust issues*, in 2018 3rd Cloudification of the Internet of Things (CIoT), IEEE, 2018, pp. 1–4.
[61] V. Venkatesh, K. Kang, B. Wang, R. Y. Zhong, and A. Zhang, *System architecture for blockchain based transparency of supply chain social sustainability*, Robotics and Computer-Integrated Manufacturing, 63 (2020), p. 101896.
[62] W. Viriyasitavat, L. Da Xu, Z. Bi, and A. Sapsomboon, *New blockchain-based architecture for service interoperations in internet of things*, IEEE Transactions on Computational Social Systems, 6 (2019), pp. 739–748.
[63] S. Wang, L. Ouyang, Y. Yuan, X. Ni, X. Han, and F.-Y. Wang, *Blockchain-enabled smart contracts: Architecture, applications, and future trends*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, (2019).
[64] X. Wang, X. Zha, W. Ni, R. P. Liu, Y. J. Guo, X. Niu, and K. Zheng, *Survey on blockchain for internet of things*, Computer Communications, (2019).
[65] Y. Xiao, N. Zhang, W. Lou, and Y. T. Hou, *A survey of distributed consensus protocols for blockchain networks*, IEEE Communications Surveys & Tutorials, 22 (2020), pp. 1432–1465.
[66] D. Yang, S. Yoo, I. Doh, and K. Chae, *Selective blockchain system for secure and efficient d2d communication*, Journal of Network and Computer Applications, 173 (2021), p. 102817.
[67] Q. Yang and H. Wang, *Privacy-preserving transactive energy management for iot-aided smart homes via blockchain*, IEEE Internet of Things Journal, 8 (2021), pp. 11463–11475.
[68] X. Yang, M. Li, H. Yu, M. Wang, D. Xu, and C. Sun, *A trusted blockchain-based traceability system for fruit and vegetable agricultural products*, IEEE Access, 9 (2021), pp. 36282–36293.
[69] S. Zhang and J.-H. Lee, *Analysis of the main consensus protocols of blockchain*, ICT express, 6 (2020), pp. 93–97.
[70] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, *Blockchain challenges and opportunities: A survey*, International Journal of Web and Grid Services, 14 (2018), pp. 352–375.
[71] A. Čolaković and M. Hadžialić, *Internet of things (iot): A review of enabling technologies, challenges, and open research issues*, Computer Networks, 144 (2018), pp. 17 – 39.

Edited by: Katarzyna Wasielewska
Received: Jul 27, 2022
Accepted: Nov 2, 2022

AIMS AND SCOPE

The area of scalable computing has matured and reached a point where new issues and trends require a professional forum. SCPE will provide this avenue by publishing original refereed papers that address the present as well as the future of parallel and distributed computing. The journal will focus on algorithm development, implementation and execution on real-world parallel architectures, and application of parallel and distributed computing to the solution of real-life problems. Of particular interest are:

Expressiveness:

- high level languages,
- object oriented techniques,
- compiler technology for parallel computing,
- implementation techniques and their efficiency.

System engineering:

- programming environments,
- debugging tools,
- software libraries.

Performance:

- performance measurement: metrics, evaluation, visualization,
- performance improvement: resource allocation and scheduling, I/O, network throughput.

Applications:

- database,
- control systems,
- embedded systems,
- fault tolerance,
- industrial and business,
- real-time,
- scientific computing,
- visualization.

Future:

- limitations of current approaches,
- engineering trends and their consequences,
- novel parallel architectures.

Taking into account the extremely rapid pace of changes in the field SCPE is committed to fast turnaround of papers and a short publication time of accepted papers.

INSTRUCTIONS FOR CONTRIBUTORS

Proposals of Special Issues should be submitted to the editor-in-chief.

The language of the journal is English. SCPE publishes three categories of papers: overview papers, research papers and short communications. Electronic submissions are preferred. Overview papers and short communications should be submitted to the editor-in-chief. Research papers should be submitted to the editor whose research interests match the subject of the paper most closely. The list of editors' research interests can be found at the journal WWW site (`http://www.scpe.org`). Each paper appropriate to the journal will be refereed by a minimum of two referees.

There is no a priori limit on the length of overview papers. Research papers should be limited to approximately 20 pages, while short communications should not exceed 5 pages. A 50–100 word abstract should be included.

Upon acceptance the authors will be asked to transfer copyright of the article to the publisher. The authors will be required to prepare the text in LaTeX 2$_\varepsilon$ using the journal document class file (based on the SIAM's `siamltex.clo` document class, available at the journal WWW site). Figures must be prepared in encapsulated PostScript and appropriately incorporated into the text. The bibliography should be formatted using the SIAM convention. Detailed instructions for the Authors are available on the SCPE WWW site at `http://www.scpe.org`.

Contributions are accepted for review on the understanding that the same work has not been published and that it is not being considered for publication elsewhere. Technical reports can be submitted. Substantially revised versions of papers published in not easily accessible conference proceedings can also be submitted. The editor-in-chief should be notified at the time of submission and the author is responsible for obtaining the necessary copyright releases for all copyrighted material.