A hybrid deep learning-improved BAT optimization algorithm for soil classification using remote sensing hyperspectral features

Abstract

Now a days, Remote Sensing (RS) techniques are used for earth observation and for detection of soil types with high accuracy and better reliability. This technique provides perspective view of spatial and aids in instantaneous measurement of soil's minerals and its characteristics. There are a few challenges that is present in soil classification using image enhancement such as, locating and plotting soil boundaries, slopes, hazardous areas and drainage condition, land use, vegetation etc… There are some traditional approaches which involves few drawbacks such as, manual involvement which results in inaccuracy due to human interference, time consuming, inconsistent prediction etc. To overcome these draw backs and to improve the predictive analysis of soil characteristics, we propose a Hybrid Deep Learning improved BAT optimization algorithm (HDIB) for soil classification using remote sensing hyperspectral features. In HDIB, we propose a spontaneous BAT optimization algorithm for feature extraction of both spectral-spatial features by choosing pure pixels from the Hyper Spectral (HS) image. Spectral-spatial vector as training illustrations is attained by merging spatial and spectral vector by means of priority stacking methodology. Then, a recurring Deep Learning (DL) Neural Network (NN) is used for classifying the HS images, considering the datasets of Pavia University, Salinas and Tamilnadu Hill Scene, which in turn improves the reliability of classification. Finally, the performance of the proposed HDIB based soil classifier is compared and analyzed with existing methodologies like Single Layer Perceptron (SLP), Convolutional Neural Networks (CNN) and Deep Metric Learning (DML) and it shows an improved classification accuracy of 99.87 %, 98.34 % and 99.9 % for Tamilnadu Hills dataset, Pavia University and Salinas scene datasets respectively.

Keywords: HDIB, BAT optimization algorithm, priority stacking approach, recurrent deep learning neural network, Convolutional Neural Network, Single Layer Perceptron

1. Introduction

Over the past few years, Brazil has become a pioneer in horticulture because of its venture into new agrarian crops, upgradation & automation of machinery and excellent endeavors to address

the problems faced by the agriculturists [1][2]. To proceed with this development, Brazil must practice new methodologies for utilizing and overseeing uncultivated land zones. Most changes occurring in the soil are moderate and subtle especially when seen in the time span of human life. Be that as it may, due to natural calamities, it results in disintegration which brings quantifiable changes to the land mass [3]. The progressions are principally in the structure and organization of the material and such changes are referred to as 'basic changes'. Soil is the base for each generation framework and information on their properties, degree and spatial dispersion is critical before we start cultivating any products in that land mass [4]-[7]. More than 90% of world's nourishment issues are related to soil. Soil asset stock gives an understanding into the possibilities and constraint of soil for its viable use. In soil mapping we can experience stock of various soils, their type and nature and it additionally gives data regarding land structure, vegetation. Suitable evidence on the type of soils is essential in creating a plan for land use in farming, water system, waste management and so on. Standard soil study collates data about soil in an efficient way with respect to their degree, constraints and to group them. Satellite remote sensing and GIS methods are the procedures which are adequately utilized as of late in deciding the quantitative portrayal of the landscape highlights and geomorphologic mapping [8][9].

Over the years remote sensing [10] has grown as a fundamental tool in soil asset study which helps to increase the ideal land utilization plan for economic development at scale running from provincial to smaller scale levels [11]. The variables associated with physiographic forms pretty much compare to the elements of soil development. Hyperspectral information provides procedures such as collinear indicator factors, which is, autonomous from the independent variable, and these images also has a disadvantage. To solve this problem, dimensionality reduction technique is the standard procedure [12][13]. Partial Least Squares Regression (PLSR) [14][15], for instance, extends the information into a low-dimensional space shaped by a lot of symmetrical factors with an intention to increase the covariance among indicators and target variable(s). PLSR is the most widely recognized multivariate data analysis, albeit different methodologies have additionally been applied effectively. Customary techniques for distinguishing and checking of salt influenced soils can outline saltiness just up to a limited degree. These techniques are costly, tedious and require large number of tests from a region to describe spatial changeability, which limits the practice of these techniques for contemplating huge and non-uniform territories [16][17]. Different RS images are by and large generally utilized in portrayal and mapping of salt influenced

soils including ethereal photos, Multi Spectral (MS) and HS remote sensing images. Soil saltiness mapping is a troublesome assignment to perform, due to the high impact of different soil physical and synthetic properties (for example dampness, surface harshness, natural issue) on soil reflectance. Prior to the RS techniques, observing and mapping of salt influenced soils over large regions were broadly and effectively attempted utilizing broadband multispectral information [18].

Contributions of this paper are that, we propose a Hybrid Deep learning-Improved BAT optimization algorithm (HDIB) for soil classification using RS hyperspectral features. Initially, we introduce an improved BAT optimization algorithm for feature extraction of both spectral-spatial features and removed by choosing physically pure pixels from hyperspectral image, after which we are using a priority stacking approach for combining the spectral vector with spatial vector as training samples. Thirdly, we have introduced a recurrent deep learning neural network which is used to acquire the discerning metric aimed at the cataloguing of hyperspectral images, which improves the reliability of classification. Finally, the HDIB algorithm is compared with the other existing techniques in terms of classification accuracy.

The paper is arranged in the following sequence: In part I, we explain overview of soil classification using RS. In part II, we discuss the issues of soil classification and RS based on some illustrative examples. The analysis of the various issues prevalent in the previous methods and the system model design are briefed in section III and it briefs about the proposed algorithm. In part IV we have discussed about the performance analysis and the description of the dataset and in part V of the paper is used to discuss about the implementation aspects and the results. Finally, section VI outlines the major conclusions of this research.

2. Related works

Enormous research has been presented in the paper for the soil classification using RS hyperspectral features. Some of the latest work has been reviewed below.

Neto et al. [19] have proposed Artificial Neural Network (ANN) for analysis of soil which have been degraded over the times and its recuperation due to gypsum & lime deposits. The examined soil was named Oxisol & the parameters which was taken into consideration were: soil porosity, soil thickness and soil resistance. The Backpropagation method is the type of ANN which is

applied in this investigation, which is made out of 2 layers, the center & the output layer, considering the supervised training. The system designed has 4 inputs, which are the characteristics of the soil. In the center layer, the system consists of 10 neurons and the final layer which is the output layer consists of a single neuron, having the capability of analyzing if the soil has Recuperated (R), Partially Recuperated (PR) or Not Recuperated (NR). After analysing the performance of the ANN, the network was verified to have acquired satisfactory training demonstrating low Mean Square Error (MSE) which is used to classify the degraded soils.

Wu et al. [20] have used the data of Normalized Difference Vegetation Index (NDVI) which was considered over an uneven terrain using the LANDSAT images which was devoid of clouds. The Landscape indicators like Wetness Index of the Topography, elevation and slope were obtained from a computerized map with resolution of upto 30 m. Models with various criteria like pure NDVI, pure topography & stratum, NDVI + Stratum and Topography was created. By and large Receiver Operating Characteristics (ROC), kappa measurement, and the Area under ROC curve (AUC) was analysed to assess classification accuracy. The results were very encouraging and it had great impact on the outputs. Amongst all the models which were considered, the model with NDVI + Stratum and Topography had the best performance with respect to AUC, Kappa Measurement & accuracy of 0.907, 0.918, and 0.975 respectively. The findings will give important data to evaluate the nature of biological condition utilizing RS images.

Bingbo Gao et al. [21] have proposed environmental quality grade techniques that are used for mapping or assessing global means, which is basically centered on assessing the connections between unsampled areas & the edges that arrange the earth quality grades. These groupings must utilize a testing format enhancement technique to disperse extra inspecting units hooked on zones with an elevated danger of misclassification. To determine these issues, this exertion gives an extra specimen format improvement technique that at first progresses an inaccuracy index by constructing a multi-Gaussian prototypical for the unsampled locations which gives the error and variances of it which is then used to calculate the threshold assessment probability in the Gaussian curve. The regular error records of entire areas in the investigation territory are then customary as the detachment capacity of the extra testing format streamlining, and the spatial reenacted strengthening is received to acquire the upgraded examining design by limiting the objectivity work.

Ghosh et al. [22] have proposed Neuro-Fuzzy (NF) technique which is a classification-based algorithm, that helps in detecting the different soil classes from a large soil database. This classification technique is a Fuzzification method that is based on detecting the different soil classes by looking into the feature-wise characteristics of the image dataset under investigation. The Fuzzification method creates a Membership Matrix whose elements are the inputs to the neural network. This technique has been applied to the Forset Covertype, Wilt for soil and Statlog Landsat Satellite database. This paper is used to find classify the types of soils using Fuzzification method and comparing the results with Support Vector Machine (SVM), K-Nearest Neighbour, Adaptive NF Inference system and Radial Basis Function Network (RBFN). Using the NF technique, the Accuracy, Recall, Precision, Kappa Statistic, False Positive Rate (FPR), Area Under Curve (AUC), F measure, Root Mean Square Error (RMSE), True Positive Rate (TPR) were calculated and the proposed NF method has effectively proved its supremacy over the above mentioned 4 algorithms.

Yang et al. [23] have proposed Minimum Noise Fractions (MNFs) method which is integrated with a Fast and Adaptive Bidimensional Empirical Mode Decomposition (FABEMD) to resolve the scattered pixel issue while working with the HS images which is created by the atmospheric scattering and noise. The paper is based on the Indiana Pine Dataset to increase the characterization precision of airborne infrared spectrometer HS images. The MNF + FABEMD breaks down a HS image into a few Bidimensional Intrinsic Mode Functions (BIMF) & remnant picture. The initial four BIMFs are eliminated & the rest of BIMF's are used to integrate to create an informative image which is then classified using SVM.

Xue et al. [24] The image classification on HS images has its own challenges because of the usage of different types of HS data. As of late, spectral-spatial methodologies were created by together dealing with the spectra and spatial data. This paper shows a totally unique methodology from a subpixel target location's viewpoint. It executes in 4 phases - preprocessing phase, a recognition stage, that generates subpixel target maps, followed by an iterative phase, that builds up an Iterative Constrained Energy Minimization (ICEM) by applying the Gaussian filters to detect the spatial data. Finally, the Gaussian filtered images are fed into the BSNE band pictures to reclaim CEM in an iterative way.

Pike et al. [25] have proposed a classification method that is used for detecting the cancerous cells from the healthy ones by combining the spectral and spatial data in the HS images. This paper is based on the Minimum Spanning Forest (MSF) which in turn uses a band selection method for distinguishing the cancerous cells and the healthy cells. A SVM Classifier has been trained to create a pixel wise probability map for the cancerous and healthy tissues which is used later to calculate for the various band range in the HS data & finally selecting the optimal band after calculation. The MSF method is finally used to segment the HS images based on the spatial and spectral information.

Guo et al. [26] proposed a multichannel optical imaging sensor that has expanded usage of HS data for RS. For HS data, a training dataset is important for guaranteeing exceptional accuracy. Nonetheless, in RS, labelled sample data are difficult to obtain and its often costly. This makes Active Learning (AL) a significant method in image analysis. AL symbolizes to effectively construct an efficient library dataset that provides more information for classifying the images. A spatial-spectral AL technique is created that coordinates spectral and spatial highlights extricated from super pixels in an AL system.

Yu et al. [27] have introduced Hyper Spectral Image Classification (HSIC), in which the Background (BKG) is by and large rejected because of the way that getting total information on BKG is almost impossible. Tragically, BKG has huge effect on Band Selection (BS) and classification. This paper examines the above mentioned two issues and it presents an innovative method called Class Signature-Constrained BKG Suppression (CSCBS) to deal with BS in HS images, where class signatures might be acquired from the earlier or posteriori information or training datasets, besides BKG concealment might be achieved by calculating the inverse of the correlation matrix R. The concept of the Linearly Constrained Minimum Variance (LCMV) which is developed by constraining the signatures of the class, thereby reducing the effect due to BKG thereby increasing the performance of classification.

Li et al. [28] Due to the huge demand in the Big Data analysis of RS images it is proposed to have Large-Scale Remote Sensing Image Retrieval (LSRSIR) which is increasingly used by data analysts and researchers. LSRSIR might be separated into 2 classes: Uni-source LSRSIR (US-LSRSIR) & cross-source LSRSIR (CS-LSRSIR). Moreover, US-LSRSIR implies the RS images

originated from the same RS information source, whereas, CS-LSRSIR is intended to recover RS images from different RS data source. In this paper, a Source-Invariant Deep Hashing Convolutional Neural Networks (SIDHCNN) is proposed using various optimization methods and this model is constructed using double source RS dataset consisting of around 10, 000 double samples in each of these categories.

3. Problem methodology and system model

This part explains about the identification of the problem of existing soil classification using RS hyperspectral features, followed by the proposed system model.

3.1 Problem methodology

Cao et al. [29] have proposed a Deep Metric Learning (DML) neural system which is basically used for assigning small distances for the samples which are present in the same class and large distances for the samples present in the different classes. The non-linear relationships between the samples cannot be captured by the various old metric learning techniques. This paper uses the Deep Neural Network (DNN) which is used to learn about metrics which are discriminating for classifying HS images. The heaviness regularization is likewise acquainted through break from the over-fitting. To improve the classification of the system further, the spectral and spatial data are both consolidated together. The most famous learning techniques for example DML neural system can viably get familiar with the picture which is mutually unaided and administered route in state of marked information is rare. In view of the attribute of HS images, DML could give another plan to separating a more grounded enlightening capacity of HS image classification with constrained information [30]-[33].

Owing to its elevated discriminative capability to categorize besides extricate diverse materials; HS imaging has stood in RS to scrutinize the earth surface for various applications such as military, mining, atmosphere monitoring, agriculture, etc. Moreover, HS imaging can also be used in applications like biomedicine, food security, biometrics, quality control [34]. Soil cataloguing is problematic to envisage the soil reflectance by means of physical prototypes besides theories due to the probability of quantitative translation of the reflectance spectrum of the multi-mineral is quite cumbersome. In addition, the theoretical consequences do not habitually approve with reality

besides are not lawful for the impost of soil possessions. Hence, we necessitate instituting a new technique that is capable to disclose the multifaceted relationships amongst reflectance from various minerals present in the soil and exclusively in very large areas [35].

3.2 System model - A hybrid deep learning-improved BAT optimization algorithm for soil classification using remote sensing hyperspectral features (HDIB)

To overcome this draw backs, we propose a hybrid deep learning-improved BAT optimization algorithm (HDIB) for soil classification using remote sensing hyperspectral features.

1. In HDIB, first we propose an improved BAT optimization algorithm for feature extraction of both spectral-spatial features and removed by picking physically pure pixels from HS image.
2. Second, Spectral-spatial vector as training illustrations is attained by way of merging spatial vector with spectral vector by means of precedence assembling methodology. Then, a recurrent deep learning neural network is applied to acquire the discerning metric aimed at classifying the HS images.
3. Finally, the performance of proposed HDIB based soil classifier is compared with the existing state-of-art techniques with respect to classification precision.

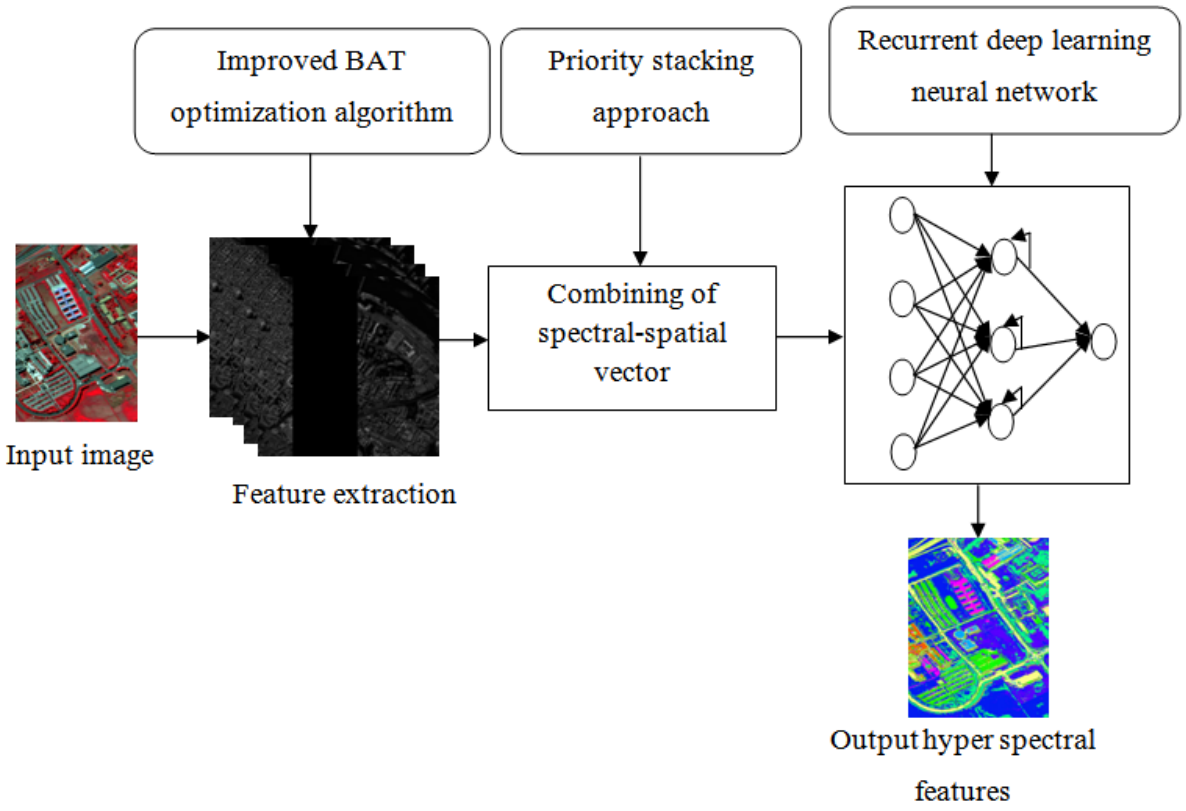


Fig.1 Proposed HDIB scheme

3.3 Feature extraction process using Improved Bat Optimization

Bat Algorithm (BA) is a metaheuristic methodology which is based on the reverberation vacillating activities of bats. Normally the bats analyze and identify their prey using their echolocation attributes. Likewise, the hypersensor is used to sense the HS features which is first analyzed and later senses the object. So here we are using improved bat optimization technique. For the sake of straightforwardness, the subsequent unrealistic guidelines are demarcated:

1) All bats exploit echo oscillation, besides the dissimilarity amongst prey as well as circumstantial barricades are recognized by means of their uniqueness in several enchanted mode;

2) Bats move arbitrarily by way of velocity V_i at position P_i, additionally all bats have a solid frequency $f_{\min}$ fluctuating wavelength λ in addition to loudness A_0 to prey. They might perceptively precise the wavelength (otherwise frequency) of their emanated pulses. Moreover,

bats remain capable to amend the pulse emission rate $r \in [0,1]$ based on the type and distance of the prey.

3) Owing to the loudness, here we consider that the loudness variation from a large progressive value A_0 to a minutest perpetual value $A_{\min}$.

For effortlessness, the characteristic plus rate arrays for bats remain customary amongst 0 in addition to 1, where 0 stances for no pulses at all and 1 stand for the pulse emanation.

3.3.1 Movement of virtual bats

Here, for extracting the HS images and to get the pure pixels, we initialize the following parameters. To deal with the glitches, the simulated bats remain castoff besides the guidelines that appraise new clarifications P_i^t besides velocities V_t^i at time phase t rationalized in a multidimensional space which are demarcated by the succeeding formulations.

$$f_t = f_{\min} + \left(f_{\max} - f_{\min}\right)\alpha \,(1) \tag{1}$$

$$V_j^i(t) = v_j^i(t-1) + \left|P^i - P_j^j(t-1)\right| f_i \tag{2}$$

$$P_j^i(t) = P_j^i(t-1) + V_j^i(t) \tag{3}$$

3.3.2 Pulse emission and loudness

The HS image is sensed by the sensor and it is extracted to get the pure pixels. Likewise, the bat is searching their prey with help of echolocation attributes, in the course of the penetrating progression; *bi* would emanate pulse with huge loudness besides trivial frequency. Formerly BA_i has found its prey, the loudness besides frequency could be modernized as trails:

$$Pu_i^{t+1} = \beta \times A_i^t \tag{4}$$

$$L_i^{t+1} = r_l^0\left[1 - \exp(-\gamma \times t)\right] \tag{5}$$

Where, BA_i is Bat and pulse emission of bats is Pu_i^{t+1} and loudness at time step t is represented in L_i^{t+1}. There is same value as β and γ.

3.3.3 Feature selection and extraction

The customary BA just might be used to resolve various optimization problems like function optimization. Nevertheless, in the case of feature selection, some strategies would be familiarized to elucidate the crusade of bats in the n-dimensional Boolean framework. In the Bat optimization algorithm, we project a binary description of BA for feature selection and the Improved Bat Optimization (IBO) algorithm practices a sigmoid utility to constrain the location of bats to only binary assessment.

$$S_i\left(v_j^i\right)=\frac{1}{1+\exp\left(-v_j^i\right)} \tag{6}$$

Instead of equation (3) the bats position can be updated as following eqn. (7),

$$P_j^i=\begin{cases}1 & if\ s_i\left(v_j^i\right)>\sigma\\ 0 & otherwise\end{cases} \tag{7}$$

Where σ ~ U (0, 1). Consequently, each bit of bats might be constrained to only binary values, which might designate the occurrence or nonappearance of the structures.

3.3.4 Fitness function

The BA exploits a fitness function to replicate the convergence of various classes (signified by D_1), of the consistency feature, besides the dispersion of inter classes (symbolized by D_2) of dissimilar features, which is demarcated as (8).

$$Fitness=\frac{\min\left(D_2(P)\right)}{1+\max\left(D_1(Q)\right)} \tag{8}$$

Where, $Q=1,2,....,NC$ and $P=1,2,....,C_{NC}^2$ are represented number of texture modules. Nc is the number of interclass. $D_2\left(Q\right)$ is the intra-class square inaccuracy of the Q^{th} texture class, $D_2\left(P\right)$ is an inter-class square inaccuracy, mentioned in (9) and (10)

$$D_1(Q)=\frac{\sum_{k=1}^{n}\left(Te_k-u_q\right)^2}{sn} \tag{9}$$

$$D_2(P)=W_a\times W_b\times\left(u_a(p)-u_b(p)\right)^2 \tag{10}$$

Where the number of samples of Q^{th} texture class is represented as sn. Te_k is Texture energy of k^{th} Sample of Q^{th} texture class and u_q its average value sample $u_a(P)$ and $u_b(P)$ is mean value of a^{th} and b^{th} samples of texture class and w_a and w_b is total number of samples of a^{th} and b^{th} texture class.

The process of improved bat optimization algorithm using feature extraction is represented in figure 2.

3.4 Merging spectral vector through spatial vector by means of priority stacking methodology

For improving the quality of images and combining the multiple classification models, we use the priority stacking approach.

3.4.1 Priority scheduling of spatial vector

In this technique, the processes are implemented, with the end objective, that the practice having utmost noteworthy needed to be executed first. Centered on implementation of each technique, the allotment time besides turnaround time is unwavering, for example First Come First Serve (FCFS). For each point p on the limit δ, we fixed a square pixel φ_p with the inside p. Specifically, we observe a set of spatial vectors (preparing tests) pixel size with 9 × 9 pixels.

Calculating the precedence $P(p)$ for each patch,

$$P(p) = C(p)D(p) \tag{11}$$

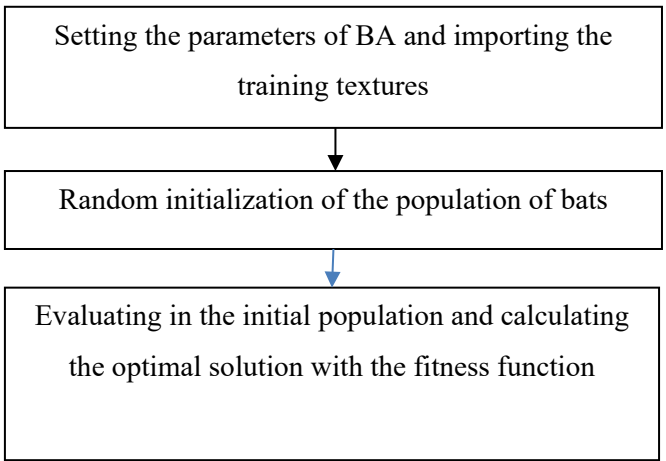


Fig.2 Feature extraction process of HS images using improved bat optimization

Where, $C(p)$ and $D(p)$ are confidence term and data term, which given below:

$$C(p) = \frac{\sum_{q \in \varphi} \cap \bar{\rho} C(p)}{\left|\varphi_p\right|} \tag{12}$$

$$D(p) = \frac{\left|\nabla_p . n_p\right|}{\alpha} \tag{13}$$

Where $\bar{\rho}$ is the harmonizing set of objective region ρ, $\left|\varphi_p\right|$ is patch area, ψ_p, n_p is an unit vector orthogonal to borderline δ and it is a spatial vector besides α is regularization parameter.

If case there is an additional training of the image adjoining the pixel p, $C(p)$ would attain an updated value. In precise, the initialization is that $C(p) = 0$, $\forall p \in \rho$, in addition $D(p) = -1$, $\forall p \in \bar{\rho}$, which ρ is the objective region.

Choosing training samples φ_p with the uppermost priority, in addition filling the patch by penetrating the utmost comparable patch after basic image ϕ. The succeeding equivalence is castoff to denote the resemblance amongst two training illustrations,

$$\varphi_q = \arg\min_{\varphi q \in \phi} d\left(\varphi_p, \varphi_q\right) \tag{14}$$

Every pixel p′, p′ $\in$ $\varphi_p \cap \rho$, is occupied by conforming pixel in φ_p besides apprising the assurance value with the subsequent eqn (15),

$$C(q) = C(p), \forall_q \in \psi_p \cap \rho \tag{15}$$

3.4.2 Stacking approach

Stacking is another grouping strategy that is used to link the enormities of a few characterization strategies utilizing a meta-classifier. The proposed methodology is exploited, at the principal level of grouping. Subsequently, the order of potentials with various component congregations, and the consequences of the finest classifier results are stacked using the requisite calculation. In the wake of assembling the potential results, basic classifiers are applied to the blend of these results and the picture includes in the investigations, stacking techniques are independently applied to each

level fig.3 and Three stacking methodologies (stacking on the main level, stacking on the subsequent level, and stacking on the two levels) and characterization without stacking are looked at. We utilized three stacking techniques to join the characterization yields of the single grouping models.

3.5 Classification of HS images using recurrent deep learning neural network

We propose a novel and a repetitive deep learning neural system architecture for classifying the HS images. The outline of the deep learning neural system is given in Fig. 4. It comprises of two sections: The intermittent portion removes semantic portrayals from pictures; the deep neural part mockups picture/mark affiliation.

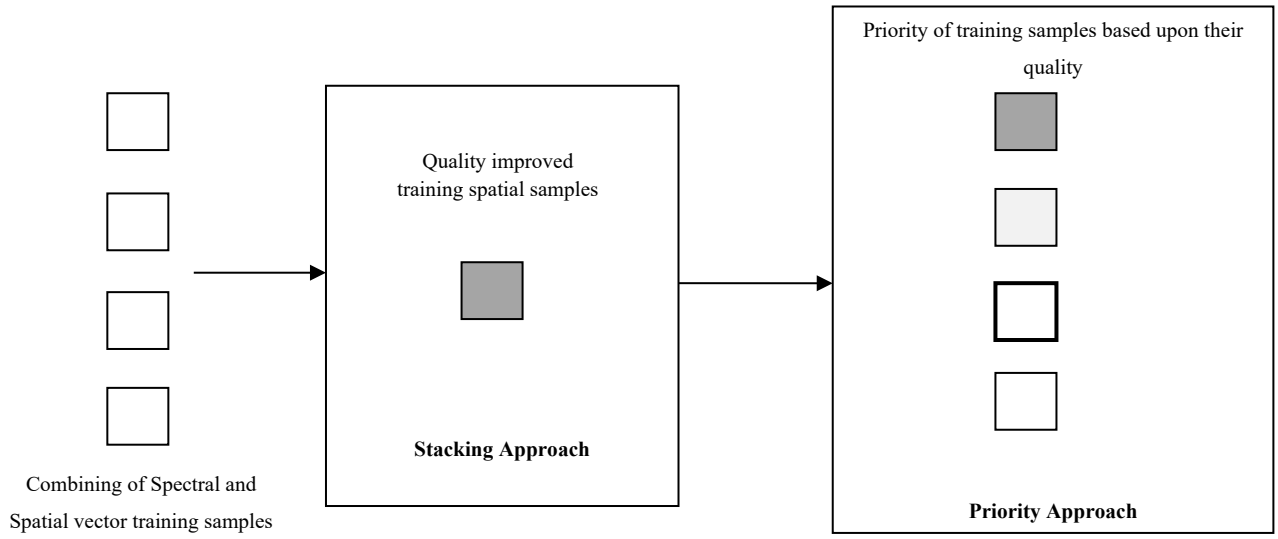


Fig.3 Merging spatial vector with spectral vector by means of priority stacking method

The likelihood of prediction can be registered by the intermittent deep learning neural system. The picture, mark and intermittent portrayals are anticipated to a similar low dimensional space to demonstrate the picture content relationship. The repetitive deep learning neural system model is utilized; at this point ground-breaking portrayal marks the co-event dependence. It grosses the embedding of the anticipated mark at every time besides keeps up a masked state to display the name on the co-event data. From the probability of the name given, the newly anticipated marks can be figured by their advert frameworks with the aggregate of the image and repetitive

embedding. The likelihood of a prediction can be acquired as the result of an earlier likelihood of each mark given the past names in the expectation way.

In a characteristic recurring network as shown in Figure 4, categorization by means of length S, hidden layer aspect $h^{(s)}$ and the output label $Y^{(s)}$ at the state s ∈ [1, ⋯ , S] might be described as:

$$h^{(s)} = fh\left(W_{hh}\, h^{(S-1)} + W_{ih} x^{(S)} + bh\right) \tag{16}$$

$$Y^{(s)} = f_0\left(W_{h_0 h}\, h^{(S)} + b_0\right) \tag{17}$$

Where, $x^{(s)}$ is the s-th input data, $h^{(s)}$, is the hidden layer unit, $Y^{(s)}$ embodies the output, W_{ih}, W_{hh}, W_{ho} are the revolution matrices amongst $x^{(s)}$ in addition $h^{(s)}$, $h^{(s-1)}$ as well as $h^{(s)}$, besides $Y^{(s)}$. b_h, b_0 are the persistent bias terms, whereas f_h, f_0 are the non-linear stimulation functions.

The label k is signified as a one-hot vector e_k = [0, . . . 0, 1, 0, . . . , 0], which is 1 at the k-th position, and 0 at some other location. The label implanting might be attained by burgeoning the one-hot vector with a label entrenching matrix u_l.

$$W_k = u_l . e_k \tag{18}$$

The intermittent layer grosses the label embedding of the prophesied label, by nonlinear utilities:

$$o(t) = h_0\left(r(t-1), W_k(t),\right), r(t) = h_r\left(r(t-1), w(t)\right) \tag{19}$$

Where r (t) besides o (t) are the hidden states in addition to the output of the contemporary layer at the time phase t, correspondingly, wk (t) is the label implanting of the t-th label in the prophecy track, and $h_0(.)$ and $h_r(.)$ are the non-linear RNN functions.

$$x_t = h\left(u_0^y\, o(t) + u_I^y\, I\right) \tag{20}$$

The matrices for repeated deep layer output besides pixel depiction is $\left(u_0^y, u_I^y\right)$ amount of columns of u_0^y and u_I^y remain the similar as the label embedding matrix u_l. I is the DNN training set, which

is from priority stack approach representation. Finally, the label marks might be calculated by proliferating the reverse of u_l beside x_t to total the distances amongst x_t and every label implanting.

$$S(t) = u_l^T \, x_t \tag{21}$$

The label possibility might be computed by means of Softmax normalization.

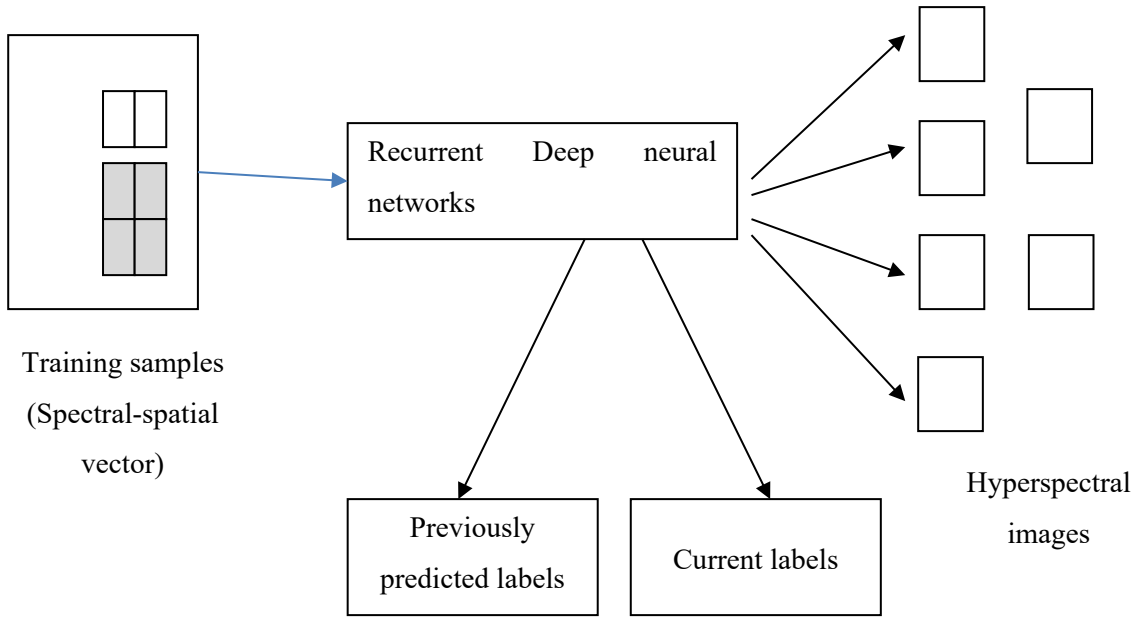


Fig.4 Recurrent deep neural networks classifies the hyperspectral images with help of comparing the previous predicted labels and current labels

4. Performance Analysis

Our proposed HDIB design is synthesized and simulated in TensorFlow. Initially, we differentiate the pure pixels of HS images from spectral-spatial features using the improved bat optimization algorithm and further improve the quality of those pixels using spatial vectors as training samples, with help of priority stacking approach. We improve the accuracy and classify the HS images with help of recurrent DNN. We can use TensorFlow for simulating and compressing the HS images, the details of which are shown in Table- 1. The simulation results are compared with the existing methods like DML, CNN and S+ LP classifiers. The CNN technique encompasses five stages: the full connection layer, the input layer, the max pooling layer, the convolutional layer and the output layer. The simulation and synthesis are done by TensorFlow, the performance of the proposed HDIB algorithm is also shown in Table-1.

4.1 Data set and testing details

4.1.1 Tamilnadu hills scenes

The dataset comprises of 145 × 145 pixels with 224 spectral replication bands with wavelength of $0.4{-}2.5{\times}10^{(-6)}$ m. The investigational groupings accessible for the scrutiny are classified into sixteen groupings. The dataset encompasses 10,249 labeled pixels, out of which 9 classes have been considered. The training and testing statistics are enumerated in Table 1. The pseudo-color map and the RS image are illustrated in Fig. 4.

4.1.2 Pavia University scene

The Pavia University dataset is captured by the Reflective Optical System Imaging Spectrometer (ROSIS-03) optical radar in a municipal region nearby the University of Pavia in Italy & it consists of 115 spectral bands, besides infrared spectra remain enclosed, with wavelength in the range of 430–860 nm. This HS image comprises of 610 × 340 pixels having a spatial perseverance of 1.3 m. A total of 42,776 labels, as well as 9 groupings are considered. The training and testing statistics are enumerated in Table 2. The pseudo-color map and the RS image are shown in Fig. 5.

Table.1 Number of Training and Testing Samples for Tamilnadu hills scenes Data Set

No.	**Class**	**Training**	**Testing**
1	Grass-pasture	200	793
2	Grass-trees	200	1,043
3	Woods	200	1,083
4	Trapezoidal areas	200	433
5	Trees	200	20,445
6	Bare soil	200	4,523
7	Soybean	200	1,230
8	Corn	200	2,578
9	Buildings-Grass-Trees-Drives	200	1,458

Table.2 Number of Training and Testing Samples for Pavia University scene

No.	**Class**	**Training**	**Testing**
1	Asphalt	200	565

2	Bricks	200	13,900
3	Metal sheets	200	345
4	Bare soil	200	7,421
5	Trees	200	13,445
6	Gravel	200	8,523
7	Meadows	200	14,342
8	Bitumen	200	1,000
9	Shadows	200	548

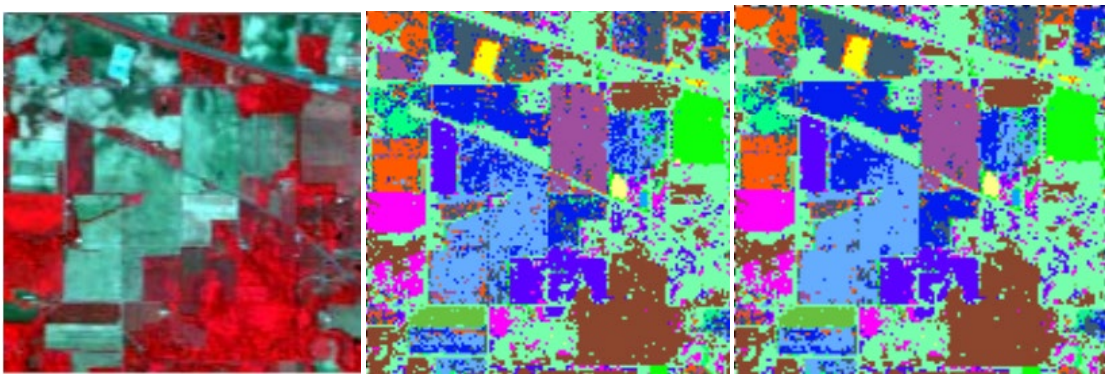
4.1.3. Salinas scene

The Salinas dataset is captured using the AVIRIS radar, apprehending an extent over Salinas Valley, California, having a spatial resolution of 3.7 m and the dataset consists of 204 bands. It primarily covers soils, vegetation etc…. and there are 16 diverse land varieties. The training and testing statistics are itemized in Table 3. The RS image and the outputs with various techniques are illustrated in Fig. 5

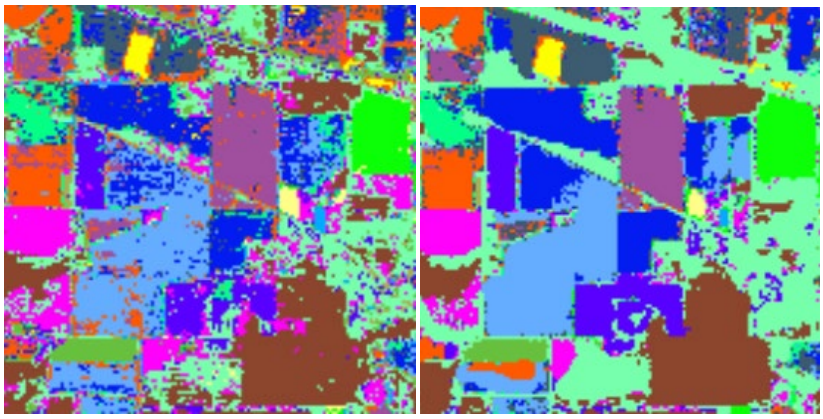
Table.3 Number of Training and Testing Samples for salinas dataset

No.	**Class**	**Training**	**Testing**
1	Brocoli green weeds 1	200	1809
2	Brocoli green weeds 2	200	3521
3	Fallow rough plow	200	1194
4	Corn senesced green weeds	200	3807
5	Lettuce romaine 4wk 200 868	200	868
6	Lettuce romaine 5w	200	1727
7	Vinyard untrained	200	3379
8	Grapes untrained	200	11,701
9	Fallow smooth	200	2478
10	Fallow	200	1976
11	Stubble	200	3959

12	Celery	200	3579
13	Corn_senesced_green_weeds	200	3278
14	Lettuce_romaine_6wk	200	916
15	Lettuce_romaine_7wk	200	1070
16	Vinyard_vertical_trellis	200	1807



(a) (b) (c)



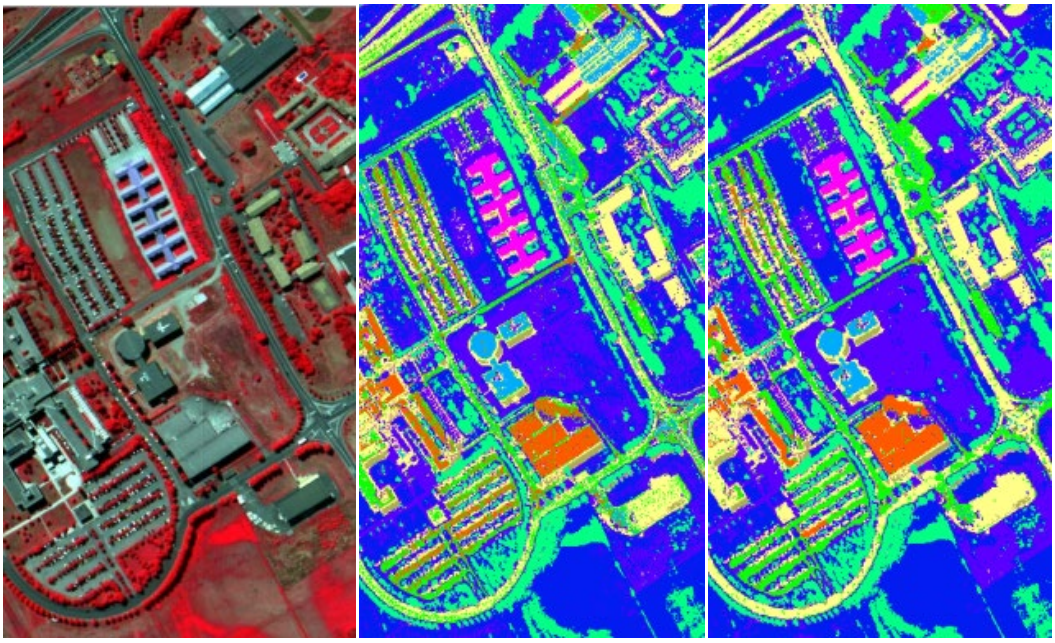
(d) (e)

Fig. 4 Tamilnadu hills scene (a) input hyper spectral image (b) output of SLP classifier (c) CNN classifier (d) DML classifier (e) proposed HDIB classifier

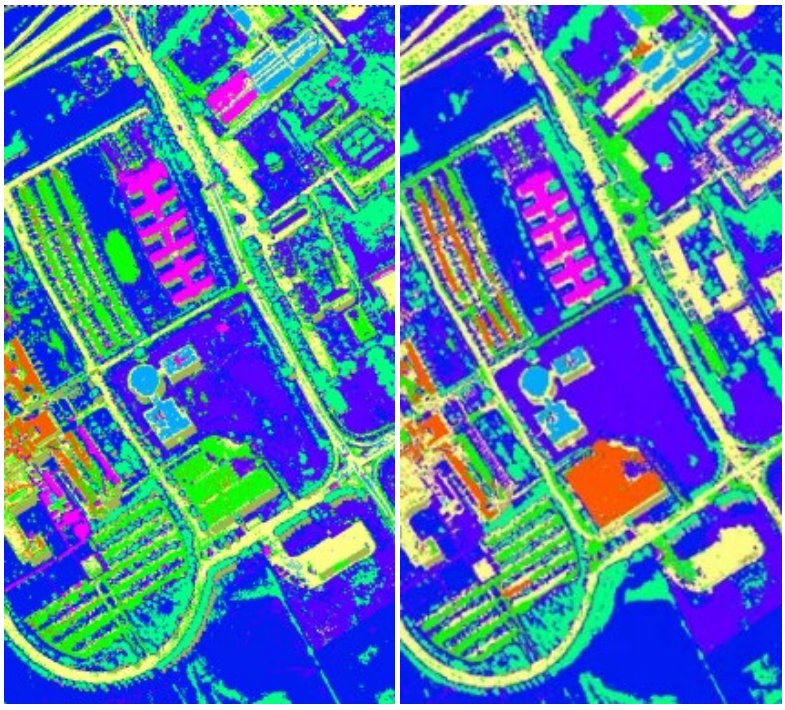
5. Results and Discussion

The proposed HDIB is compared with existing methods and it uses three data sets. Fig. 4a shows the input HS image of the Tamilnadu hill scene. Fig. 4b-d shows the classification result of existing classifiers like SLP, CNN and DML. Fig. 4e shows the classification output of proposed HDIB

method. Considering the results, one might perceive that SLP did slightly better than the DML besides both approaches have comparable difficulties of misclassification of built-up land and agronomic area. The usage of spatial information in addition to homogenization of solitary pixels in much enhanced outcomes for the HDIB classifier.



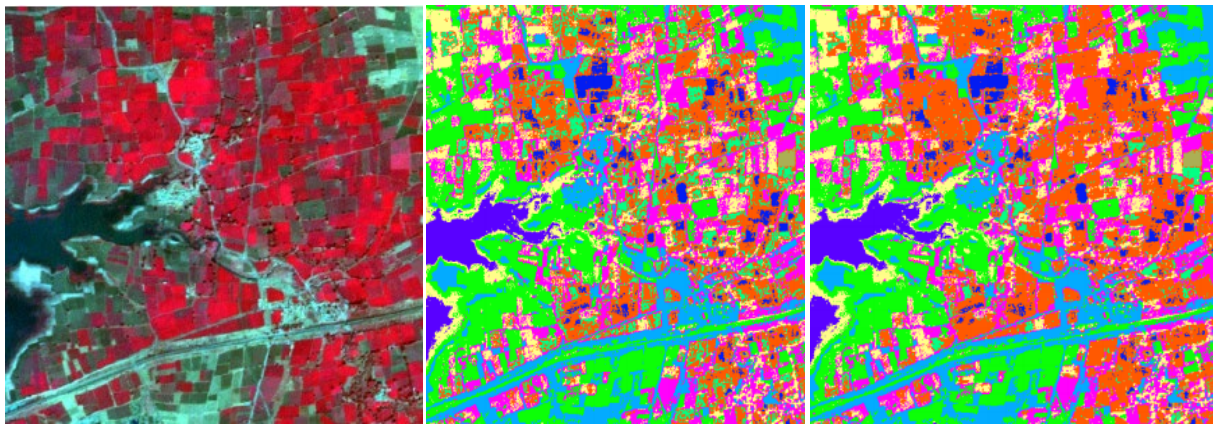
(a) (b) (c)



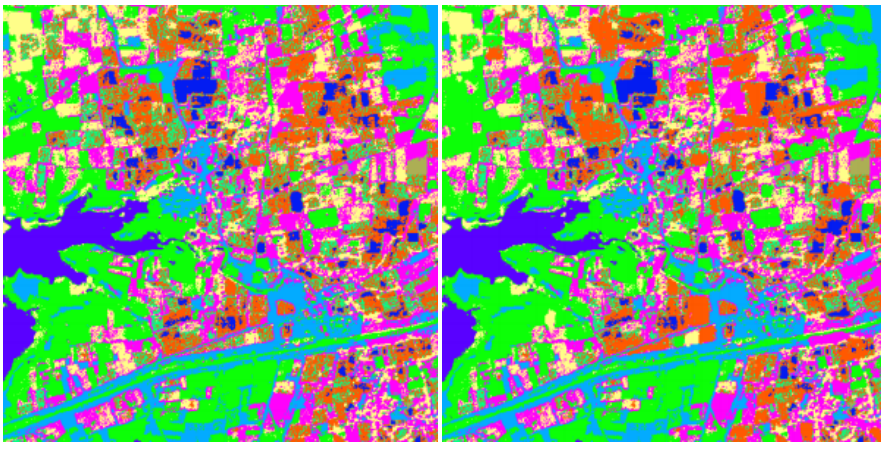
(d) (e)

Fig. 5 Pavia University scene (a) Three-band color composite image (b) output of SLP classifier (c) CNN classifier (d) DML classifier (e) proposed HDIB classifier

Fig. 5a refers to the HS image from the Pavia University scene data set. Fig. 5b-d shows the classification result of existing classifiers like SLP, CNN and DML. Fig. 5e shows the classification output of proposed HDIB method. The results validate that the HDIB process is more appropriate for the classification of data sets with elevated dimensions.



(a) (b) (c)



(d) (e)

Fig. 6 Salinas scene. (a) Three-band color composite image (b) output of SLP classifier (c) CNN classifier (d) DML classifier (e) proposed HDIB classifier

Dropping of the dimensionality by a profligate besides effectual band reduction process ensued in massive computational expense investments. Also, in precise, as majority of the images are unruffled of homogenous provinces, restraining the application of spatial information solitary for pixels recognized as miscellaneous lessened the computational necessities of the hybrid classification. On an average, the projected approach ensued in 5-10% upsurge in classification precisions deprived of incurring several momentous computational above.

Table.4 Accuracy comparison for Tamilnadu hills scene

Class	**S+LP**	**CNN**	**DML**	**HDIB (proposed)**
C1	91.21	87.89	95.54	96.34
C2	91.84	98.3	93.43	94.23
C3	94.15	99.70	99.76	99.87
C4	99.70	95.72	96.43	97.55
C5	92.02	90.54	91.34	92.34

Table 5 Accuracy comparison for Pavia University scene

Class	**S+LP**	**CNN**	**DML**	**HDIB (proposed)**
C1	91.23	78.89	94.23	97.23
C2	92.33	89.33	97.43	98.34
C3	92.15	89.70	91.76	93.87
C4	92.11	95.56	96.34	97.56
C5	78.02	90.23	93.34	91.45

Table 6 Accuracy comparison for Salinas scene

Class	**S+LP**	**CNN**	**DML**	**HDIB (proposed)**
C1	99.23	99.72	99.78	99.913
C2	93.84	78.46	90.12	98.34
C3	99.61	99.45	99.56	99.89
C4	86.98	89.63	90.l4	97.98
C5	88.90	90.33	91.45	94.39

From the above criteria, the proposed HDIB provides high accuracy images and better reliability of classification. Compared with the existing state-of-art techniques such as CNN, S+LP and DML, the proposed HDIB based soil classifier provides higher accuracy. We create a comprehensive investigation of two constraints: learning rate μ in addition regularization limitation σ with the 3 HS data sets. When μ = 10, the accuracy is 39.07%, 67.52% and 62.89% for three datasets. The intention is that the huge learning frequency creates the DML network tough to congregate to an unwavering state. By way of diminution of μ, the classification accurateness upsurges progressively. When μ = 0.1, the uppermost precisions are acquired for the 3 datasets. The motive is that a too trivial learning proportion might create the network drop hooked on a limited least point.

6. Conclusion

We have proposed a hybrid deep learning-improved BAT optimization algorithm (HDIB) for soil classification using remote sensing hyperspectral features and we extracted the spectral-spatial features by choosing manually pure pixels from HS image and we combined Spectral-spatial vector by the help of priority stacking approach. We classified the HS images by using of recurrent deep learning neural network also we improved the reliability of classification. Finally, the proposed HDIB classifier results are compared with existing classifiers are S+LP, CNN and DML for three different standard datasets. The comparative analysis shows the effectiveness of proposed

HDIB classifier in terms of accuracy. The accuracy of proposed HDIB classifier is very high compared to existing classifiers. Hence, in the future, we will focus on developing multi-class soil features classification methods for HS images.

References

[1] Francois Tavin F, Audrey Roman, Sandrine Mathieu, Frederic Baret, Weidong Liu, Pierre Gouton, *"Comparison of metrics for the classification of soils under variable geometrical conditions using hyperspectral data",* IEEE Geoscience and Remote Sensing Letters, Volume 5, Issue 4, Pp. 755-759, 2008

[2] Maxim Shoshany, Ophir Almog, Victor Alchanatis, *"Wavelet decomposition for reducing flux density effects on hyperspectral classification",* IEEE Geoscience and Remote Sensing Letters, Volume 6, Issue 1, Pp. 38-41, 2008

[3] Rishi Prakash, Dharmendra Singh, Nagendra P Pathak, *"A fusion approach to retrieve soil moisture with SAR and optical data",* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume 5, Issue 1, Pp. 196-206, 2012

[4] Sajjad Ahmad, Ajay Kalra, Haroon Stephen, *"Estimating soil moisture using remote sensing data: A machine learning approach"*, Advances in Water Resources, Volume 33, Issue 1, Pp.69-80, 2010

[5] Nilda Sanchez N, Jose Martínez-Fernández, Alfonso Calera, Enrique Torres, Carlos Pérez-Gutiérrez, *"Combining remote sensing and in situ soil moisture data for the application and validation of a distributed water balance model (HIDROMORE)",* Agricultural water management, Volume 98, Issue 1, Pp.69-78, 2010

[6] Mulder V.L, S De Bruin, Schaepman M.E, Mayr T.R, *"The use of remote sensing in soil and terrain mapping—A review",* Geoderma, Volume 162, Issue 1-2, Pp.1-19, 2011

[7] Wang Hong-wei, Fan Yong-hong, Tashpolat Tiyip, *"The research of soil salinization human impact based on remote sensing classification in oasis irrigation area",* Procedia Environmental Sciences, Volume 10, Pp 2399-2405, 2011

[8] Alexis Comber, Peter Fisher, Chris Brunsdon, Abdulhakim Khmag, *"Spatial analysis of remote sensing image classification accuracy",* Remote Sensing of Environment, Volume 127, Pp.237-246, 2012

[9] Emilio Chuvieco, Louis Giglio, Chris Justice, "*Global characterization of fire activity: toward defining fire regimes from Earth observation data",* Global change biology, Volume 14, Issue 7, Pp 1488-1502, 2008

[10] Paul Harmon, Rex Maus, William Morrissey, *"Expert systems: tools and applications",* John Wiley & Sons, Inc, ISBN: 9780 0713 53922, 1988.

[11] Duccio Rocchini, Giles M Foody, Harini Nagendra, Carlo Ricotta, Madhur Anand, Kate S He, Valerio Amici, Birgit Kleinschmit, Michael Förster, Sebastian Schmidtlein, Hannes Feilhauer, Anne Ghisla, Markus Metz, Markus Neteler, *"Uncertainty in ecosystem mapping by remote sensing",* Computers & Geosciences, Volume 50, Pp.128-135, 2013

[12] Christine Hladik, Merryl Alber, "*Classification of salt marsh vegetation using edaphic and remote sensing-derived variables*", Estuarine, Coastal and Shelf Science, Volume 141, Pp. 47-57, 2014

[13] Herbert K Winning, Mike J Hann, "*Modelling soil erosion risk for pipelines using remote sensed data*", Biosystems Engineering, Volume 127, Pp.135-143, 2014

[14] Jinyang Du, John S Kimball, Marzieh Azarderakhsh, R Scott Dunbar, Mahta Moghaddam, Kyle C McDonald, "*Classification of Alaska spring thaw characteristics using satellite L-band radar remote sensing*", IEEE Transactions on Geoscience and Remote Sensing, Volume 53, Issue1, Pp.542-556, 2015

[15] Offer Rozenstein, Tarin Paz-Kagan, Christoph Salbach, Arnon Karnieli, "*Comparing the effect of preprocessing transformations on methods of land-use classification derived from spectral soil measurements*", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume 8, Issue 6, Pp.2393-2404, 2015

[16] Yi Tang, Xinrong Li, "*Set-based similarity learning in subspace for agricultural remote sensing classification*", Neurocomputing, Volume 173, Pp.332-338, 2014

[17] Mohamed A E Abdel Rahman, A Natarajan, C A Srinivasamurthy, Rajendra Hegde, "*Estimating soil fertility status in physically degraded land using GIS and remote sensing techniques in Chamarajanagar district*, *Karnataka, India",* The Egyptian Journal of Remote Sensing and Space Science, Volume 19, Issue 1, Pp.95-108, 2016

[18] Lei Zhu, Li Ma, "*Class centroid alignment-based domain adaptation for classification of remote sensing images*", Pattern Recognition Letters, Volume 83, Pp.124-132, 2016

[19] Alfredo Bonini Neto, Carolina dos Santos Batista Bonini, Beatriz Santos Bisi, Andre Rodrigues dos Reis, "*Artificial Neural Network for Classification and Analysis of Degraded Soils*", IEEE Latin America Transactions, Volume 15, Issue 3, Pp. 503-509, 2017

[20] Wei Wu, Qipo Yang, Jiake Lv, Aidi Li, Hongbin Liu, *"Investigation of Remote Sensing Imageries for Identifying Soil Texture Classes Using Classification Methods",* IEEE Transactions on Geoscience and Remote Sensing, Volume 57, Issue 3, Pp. 1-11, 2018

[21] Bingbo Gao, Anxiang Lu, Yuchun Pan, Lili Huo, Yunbing Gao, Xiaolan Li, Shuhua Li, Ziyue Chen. *"Additional Sampling Layout Optimization Method for Environmental Quality Grade Classifications of Farmland Soil",* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume 10, Issue 12, Pp. 5350-5358, 2017

[22] Soumadip Ghos, Debasish Biswas, Sushanta Biswas, Debasree Chanda Sarkar, Partha Pratim Sarkar, *"Soil Classification from Large Imagery Databases Using a Neuro-Fuzzy Classifier",* Canadian Journal of Electrical and Computer Engineering, Volume 39, Issue 4, Pp. 333-343, 2016

[23] Ming-Der Yang, Kai-Shiang Huang, Yeh Fen Yang, Liang-You Lu, Zheng-Yi Feng, Hui Ping Tsai, "*Hyperspectral Image Classification Using Fast and Adaptive Bidimensional Empirical Mode Decomposition with Minimum Noise Fraction*", IEEE Geoscience and Remote Sensing Letters, Volume 13, Issue 12, Pp.1-5, 2016

[24] Bai Xue, Chunyan Yu, Yulei Wang, Meiping Song, Sen Li, Lin Wang, Hsian-Min Chen, Chein-I Chang. *"A Subpixel Target Detection Approach to Hyperspectral Image Classification",* IEEE Transactions on Geoscience and Remote Sensing, Volume 55, Issue 9, Pp. 5093-5114, 2017

[25] Robert Pike, Guolan Lu, Dongsheng Wang, Zhuo Georgia Chen, Baowei Fei, *"A Minimum Spanning Forest-Based Method for Noninvasive Cancer Detection with Hyperspectral Imaging",* IEEE Transactions on Biomedical Engineering, Volume 63, Issue 3, Pp. 1-11, 2015

[26] Jielian Guo, Xiong Zhou, Jun Li, Antonio Plaza, Saurabh Prasad, *"Superpixel-Based Active Learning and Online Feature Importance Learning for Hyperspectral Image Analysis",* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume 1, Issue 1, Pp. 347-359, 2017

[27] Chunyan Yu, Yulei Wang, Meiping Song, Chein-I Chang, "*Class Signature-Constrained Background- Suppressed Ap*

proach to Band Selection for Classification of Hyperspectral Images", IEEE Transactions on Geoscience and Remote Sensing, Volume 57, Issue 1, Pp. 14-31, 2019

[28] Yansheng Li, Yongjun Zhang, Xin Huang, Jiayi Ma, *"Learning Source-Invariant Deep Hashing Convolutional Neural Networks for Cross-Source Remote Sensing Image Retrieval",* IEEE Transactions on Geoscience and Remote Sensing, Volume 56, Issue 11, Pp. 6521-6536, 2018

[29] Xianghai Cao, Yiming Ge, Renjie Li, Jing Zhao, Licheng Jiao, "*Hyperspectral imagery classification with deep metric learning*", Neurocomputing, Volume 356, Pp. 217-227, 2019

[30] Qi Liu, Zhengtao Li, Shuai Shuai, Qizhen Sun, "*Spectral Group Attention Networks for Hyperspectral Image Classification with Spectral Separability Analysis*", Infrared Physics & Technology, Volume 108, 2020.

[31] Yonghe Chu, Hongfei Lin, Liang Yang, Dongyu Zhang, Yufeng Diao, Xiaochao Fan, Chen Shen, "*Hyperspectral image classification based on discriminative locality preserving broad learning system",* Knowledge-Based Systems, Volume 206, 2020

[32] Xiangtian Meng, Yilin Bao, Jiangui Liu, Huanjun Liu, Xinle Zhang, Yu Zhang, Peng Wang, Haitao Tang, Fanchang Kong, "*Regional soil organic carbon prediction model based on a discrete wavelet analysis of hyperspectral satellite data*, "International Journal of Applied Earth Observation and Geoinformation, Volume 89, 2020

[33] Saeideh Ghanbari Azar, Saeed Meshgini, Tohid Yousefi Rezaii, Soosan Beheshti, *"Hyperspectral Image Classification Based on Sparse Modeling of Spectral Blocks*", Neuro computing, Volume 407, Pp. 12-23, 2020

[34] Onuwa Okwuashi, Christopher E Ndehedehe, "*Deep support vector machine for hyperspectral image classification*", Pattern Recognition, Volume 103, 2020

[35] Yuanshu Zhang, Yong Ma, Xiaobing Dai, Hao Li, Xiaoguang Mei, Jiayi Ma, "*Locality-constrained Sparse Representation for Hyperspectral Image Classification*", Information Sciences, Volume 546, Pp. 858-870, 2020

Authors Profile:

Mr. S. Prasanna Bharathi (Sankara Raghavan Prasanna Bharathi) obtained his Bachelor's degree in Electronics and Communication Engineering from SSN College of Engineering affiliated to University of Madras. Then he obtained his Master's degree in Microelectronics from Victoria University, Melbourne, Australia. He is pursuing his Ph.D in the field of Remote Sensing in Saveetha School of Engineering, SIMATS. Currently, he is working as an Assistant Professor in SRM IST, Vadapalani Campus. He has around 12 years of teaching experience and 4 years of industry experience. He has published around 20 papers in various reputed International Journals, National and International Conferences. His field of interests includes Remote Sensing, VLSI Design, IoT. He is a Member of The Institution of Engineering and Technology and he holds the position of Treasurer in IET Chennai Local Network.

Dr. S. Srinivasan (**Subramanian Srinivasan**) received his Ph.D Degree in the field of IoT Ubiquitous Computing from St.Peters Institute of Higher Education & Research, Chennai. He completed his M.E in Microwave & Optical Engineering from A.C.C.E.Tech, Karaikudi and received his B.E Degree in ECE from Government College of Engineering, Salem. He is currently working as Professor and Program Director to Biomedical Engineering and Biotechnology Department in Saveetha School of Engineering, SIMATS. He has around 21 + years of Teaching experience. He has published around 50 papers in various reputed International Journals, International & National Conferences. He is the Reviewer and Editorial Board Member of various reputed International Journals. He serves as TPC member in various conferences. He is a Member of various Professional Body Organizations like ISTE, IET, IEEE and IACSIT.

Ms. G. Chamundeeswari (Ganesan Chamundeeswari) is currently pursuing her Ph.D in the field of Remote Sensing in Saveetha School of Engineering, SIMATS. She completed her M.E in the field of Communication Systems from SRM University and received her B.E degree in Electronics & Communication Engineering from Madras University. She has around 12 years of teaching experience. She has published around 15 papers in various reputed International Journals, National and International Conferences. Her fields of Interest are Remote Sensing, Digital Image Processing, Digital Signal Processing, Machine learning and Internet of Things

Dr. Ramesh, received his Bachelor of Engineering from Karnatak University and Master of Engineering (Engineering Mechanics) from IIT Madras. Professor Ramesh received his doctoral degree from IIT, Roorkee. He has a total experience of 19 years in teaching, research and industry in various capacities and is being instrumental in driving the institute to deliver quality-packed education to the students. A few awards he has won are National Doctorate Fellowship for Ph.D., and the best lecturer award from SMIT. He is responsible/member for Academic Council, Board of Research, Educational Technical Unit and IRB (Institute Review Board). To his credit, he has 16 peer-reviewed journals published in National and International Journals. His areas of interest include sedimentation studies, structural dynamics and mechanics of structures. He teaches subjects like Structural Engineering, Design of Reinforced Concrete and Structural Analysis to all undergraduate and postgraduate students.

AGENT BASED EFFECTIVE RESOURCE ALLOCATION USING CUCKOO AND HARMONIC SEARCH (ABACH)

Abstract:*Cloud computing is a necessary component of current technology. It contains a feature called Resource Allocation that is worth mentioning. Server and virtual machine (VM) and task consolidation are all examples of consolidation issues in cloud computing. Cloud computing is a collection of corporeal and virtualized possessions delivered to clients via the internet on a request and pay-per-use basis. Cost and makespan are two features that are evaluated in job scheduling and resource allocation. Resource distribution and task scheduling must be carefully organised and optimised together in order to obtain improved task scheduling performance. Resource allocation in distributed systems has grown increasingly difficult due to these issues. By lowering manufacturing costs and enhancing system productivity, group technology has been successfully employed to address resource allocation difficulties.Multi-impartial optimization influences scheduling problems in the cloud computing environment, but single-objective algorithms are usually used. The algorithms must solve multi-objective glitches that differ greatly from single-objective optimization procedures and methodologies. Meta-heuristic algorithms have consistently demonstrated their ability to solve multi-impartial optimization glitches for this purpose. In the proposed method where a innovative approach using a uniqueamalgamation of cuckoo and harmonic exploration algorithm which are helpful for a proper scheduling of resources. To strengthen the concept of our approach three parameters as makespan time, utilization or resources and cost were considered. The obtained results were compared with the state of art of the existing tactics which clearly showed that our approach is better.*

Keywords*: Cloud Computing (CC),Cloud Technology (CT),Cuckoo Search Optimization (CSO),Egg Laying Radius (ELR),Gravitational Search Algorithm (GSA), Harmony Search Memory (HM),Heterogeneous Earliest Finish Time in simulations (HEFT), Hybrid Genetic Algorithm (HGA), Hybrid Gravitational Search Algorithm (HGSA),National Institute of Standards and Technology (NIST),Virtual Machine (VM).*

1. INTRODUCTION

‘‘CC is a methodology for providing on-demand networking pool of configurable of programmable computational power that can be quickly supplied and released with no administration effort or network operator contact.," according to NIST. [1] defines CC as applications transported as services over the Internet via data center hardware along with software; the cloud service operating systems is thought to as a cloud. CC is an extension of the ICT paradigm in which business applications' complex services may be exposed and accessed through the internetCloud computing thus is a SaaS and utility computer programme which does not use small sized cloud services, but which is based on management virtualization.

Typically large Data Centers are used to support cloud computing with millions of powerful processors. These computer systems can be located in an adapting existing containing all the hardware and software of the cloud servers. Data centres can be shared with the owner, from the other hand, which can lease cloud services from the service provider facility. The Data Center will therefore accommodate as many customers as possible, and will therefore be responsible for a variety of applications. In order to do this, virtualization is often used to follow a variety of operating and maintenance issues in the data centre, where users are given network technologies rather than physical platforms.

Virtualization is a powerful cloud computing backend that facilitates high-level scalability, agility and cloud computing availability. Cloud infrastructure finances in the cloud computing environment are virtualized and supported by several critical support technologies for cloud clients as VMs. The allocation

of resources for this environment is therefore linked to the infrastructure capacity for the incoming VM's (hosts). Server consolidation is usually used in a data centre to maximise the use of physical resources. This technique of Resource Management combines VMs with hardware resources. This approach leads to multivariate mapping to satisfy a variety of demands, including the number of Processor available, the hard disc space required for storing, the physical memory and the network bandwidth assigned. It is a sort of allocation of resources aimed at maximizing the use of physical cloud resources. Another way to say this is that virtualization technology is the powerful tool used on a cloud platform to manage dynamic resources.

In cloud computing, the problematic of heterogeneity and platform insignificance of contributors'requirements can be better managed by encapsulating the service in VM and plotting it to each physical server. While CT is extensively acknowledged as a means of providing a variety of services, it faces a number of challenges in terms of power consumption, resource management, security, big data and service quality. The comparatively high operational costs and electricity consumption of data centres are well-known flaws, with significant environmental consequences. With rising energy consumption and isessential for high computational power, offering energy efficient resolutions has developed one of the top concerns as a result of this deficit. As a result, there is a lot of effort going into finding effective ways to strike a compromise between reducing data centre power usage and maintaining service performance and quality.

In order to increase resource efficiency, current resource management systems rarely analyse the world security of VM placement. Because applications resource demands shift over time, the existing VM-to-physical-node mapping may no longer be appropriate for forthcoming workloads. New blistering nodes will emerge in the neighboring future, resulting in new integrative approaches. Resource reallocation may be resulting in supplementary overheads such as leisure time, migration time and service deterioration. The durability of a VM allocation strategy would be assessed throughout dynamical resource conformation. The problem of allocation of resources is an NP-hard problem, that is a sort of combinatorial problem. An ecological design algorithm can anticipate an ideal solution in polynomial time.

2. RELATED WORK

Several scheduling approaches for cloud computing are presented in the literature. Here, we look at some of the work that has been done on task scheduling. In CC, Zhu et al. [2] developed a scheduling system. They devised bidding value calculation rules for both backward and forward announcement-bidding stages. They also used an agent-based dynamic scheduling technique called ANGEL to create autonomous and periodic activities in clouds. In a virtualized cloud context, this solution addressed the challenges of schedulability, priority, scalability, and real time. Despite the fact that this strategy did not take into account communication and dispatching durations, it was unable to improve resource utilisation rates. Furthermore, among the most difficult difficulties in Clouds is process scheduling, which is the difficulty of achieving the participant's QoS while reducing the amount of workflow execution. To solve the issues, Alkhanak et al. [3] presented an expense work flow management in CC. The taxonomy group organises the cost-conscious key aspects of WFS in cloud technology in aspects of QoS implementation, foundation function, and framework design.

Shi et al. [4] developed an Energy-Efficient Task Scheduling for Time-Constrained Work in Local Mobile Clouds. They begin by developing mobile device computational equations. Then, pinpoint the problem and devise a probabilistic scheduling method.How to use MCC to enable mobile devices to run complicated real-time applications while maintaining great energy efficiency remains a difficult topic

in mobile cloud computing. To solve the problem,it is created a multi-objective work scheduling system. In cloud-based disassembly, Jiang et al. [5] devised a multi-objective method for work development and resource distribution. They came up with two goals: shortening the makespan and lowering the total cost. The non-conquered sorting GA II was used to resolve the multi-objective problematic. This solution proved effective, although it required the most time to execute and increased the system's complexity.

3. EXISTING WORK

Multi-objective-Oriented CSOBased Resource SAfor Clouds (MOCSO): The main contribution of this approach les at consuming less time and cost to provide an efficient resource allocation algorithm. The main advantage of this approach is that when there are more parameters to consider and, in such state, also resource allocation is done well.Choudhary et al. [6] offer a HGSA for bi-objective scheduling processes that reduces both the makespan and the execution cost. The HGSA outperforms the basic GSA, HGA, and HEFT. For tackling the short-term hydrothermal arrangement problem, Nguyen and Vo [7] suggest using the Modified CS (MCS) algorithm. The goal of the hydrothermal development problem is to diminish the over-all cost of thermal generators while accounting for loading effects, power balance constraints, generator operating limits andwater availability. It uses cuckoo search optimization technique. The downslope of this approach is that it works for few optimization parameters only.

An ERM for Prioritized Users In Cloud Environment Using Cuckoo Search Algorithm (ERMPUCSA): The positive side of this approach like at an innovative method to allocation the resource depending on load. With less computation and exaction time best resource allocation is carried out.Zhang.et.al. [8] suggested a cloud-based dynamically heterogeneity aware provisioning system that is suitable for DCP in heterogeneity server farms. Here, classification is done first, followed by prediction, and then DCP is done. It used cuckoo search algorithm. The negative side of this algorithm is that it works for a prescribed datacenter only.

Power-Aware Resource Reconfiguration Using GA in Cloud Computing (EPMOGA): the success point of this approach lies at efficient resource allocation during variable workloads. This approach is based on genetic algorithm. The positive side of this approach is that it defined dynamical nature resource allocation algorithm. By forecasting future resource bandwidth utilization of scheduled VMs, an energy-efficient wealth distribution framework [9] is presented to minimize physical node overload occurrence for made too many mistakes' clouds. The most serious issue in cloud computing is energy usage. Particularly in the multicore age [10], it gets more significant. Certain approaches for improving application SLAs are also discussed, as well as some energy management solutions for specific applications: business process and stream processing [11]. The downfall of this approach lies as it is limited to work with few objectives.

MCS for Resource Allocation on Social Internet-of-Things (MCRA): The main contribution of this approach lies at producing an optimal reserveassortment technique. The altered algorithm of cuckoo search follows. In SIoT, the authors recognized object-object interaction protocols and architecture. A SIoT-based programming tool customs semantic models for the basic/low-level presentation of knowledge and production regulations for high-level thought in order for devices from several manufacturers to capture their variability[12][13].Its advantages are observed in less time an efficient allocation of resources. its pitfall is it works for few parameters only.

Author	Contribution	Methodology	Advantage	Limitations
Syed Hamid Hussain Madni et al. [14]	Less expected time for completion and less cost of resources	CSO	Best suitable for multi objective scheduling of resources	Work only for few optimization features
Aneena Ann Alexander et al. [15]	Allocation of resources based on load	Cuckoo search algorithm	Computation and execution time are minimized	Limited to work with only one datacenter
Li Deng et al. [16]	Resource allocation for variable workloads	Genetic Algorithm	Dynamic nature resource allocation	Does not perform well with more objectives
Himanshu Jindal et al. [17]	An optimal resource selection technique	Modified Cuckoo search algorithm	In less time best resource allocation	Limited to few parameters

Table 1: making comparison of existing works

4. PROPOSED METHOD

Proposed Algorithm

1. $\boldsymbol{Solution\ Generation}$: $solutions\ for\ assigning\ various\ tasks\ to\ Virtual\ Machines$
   $Let\ PM = \{PM_1, PM_2, PM_3, \ldots\ldots\ldots\ldots\ldots PM_p\}$
   $VM = \{VM_1, VM_2, VM_3, \ldots\ldots\ldots\ldots\ldots VM_q\}$
   $T = \{T_1, T_2, T_3, \ldots\ldots\ldots\ldots\ldots T_m\}$
   $t = \{t_1, t_2, t_3, \ldots\ldots\ldots\ldots\ldots t_n\}$
   $PM = Physical\ Machine\ , VM = \ Vritual\ Machine, T = Task, t = Sub\ task$,
2. $Calculation\ of\ Fitness$: $each\ individual\ fitness\ value\ is\ evaluated\ and\ stored\ for\ future\ reference$
   $$FF = Min\ [\beta_1(EC) + \beta_2(MC) + \beta_3(MU) + \beta_4(1 - CR) + (P)]$$
3. **Cuckoo search algorithm-based update**
   New solution
   $$\boldsymbol{S_i^{New}} = \boldsymbol{S_i^{(t+1)}} = \boldsymbol{S_i^{N(t)}} + \boldsymbol{\alpha \oplus Levy(n)}$$
4. **Harmony search algorithm-based update**
   New solution
   $$\boldsymbol{S_i^{New}} = \begin{Bmatrix} S_i^{New} \in \{S_i^1, S_i^2 \ \ldots\ldots\ldots\ldots S_i^{HMS} \ \ along\ with\ probability\ HMCR \\ S_i^{New} \in S\ with\ probability\ (1 - HMCR) \end{Bmatrix}$$

   Pitch adjustments are given as
   $$\boldsymbol{S_i^{New}} = \begin{Bmatrix} S_i^{New} \in \{S_i^1, S_i^2 \ \ldots\ldots\ldots\ldots S_i^{HMS} \ \ along\ with\ probability\ HMCR \\ S_i^{New} \in S\ with\ probability\ (1 - HMCR) \end{Bmatrix}$$

   Pitch adjustments are calculated as
   $$\boldsymbol{S_i^{New}} = \begin{Bmatrix} S_i^{New} \pm rand * b \ \ with\ probability\ PAR \\ S_i^{New} \in S\ with\ probability\ (1 - PAR) \end{Bmatrix}$$

   Where

HMCR (harmony Memory Based Considered Rate), PAR (Pitch Adjusting Rate)

6. **Hybridization**
   Solution of both Cuckoo and harmony approach are generated and compared and the best choice is selected while which may be either cuckoo or harmony approaches.
7. **Criteria for Termination**
   The algorithm only continues once the maximum number of iterations has been completed, and the arrangement with the fitness value is picked and determined to be the optimal option for job scheduling.

Cuckoo search algorithm

Optimization comprises a wide variety of areas and applications. Because time, money, and resources are all finite, optimization is becoming increasingly vital. Energy-efficient designs and sustainable solutions to many work-related issues, for example, necessitate a fundamental shift in perspective and design practice. People, from the other hand, seek clever and intelligent things, and computational intelligence has emerged as a promising field with far-reaching implications.

X.S. Yang and S. Deb expanded the Cuckoo optimisation algorithm in 2009 [18]. Instead of simply random isotropical progression, the Cuckoos laying procedure combined with Levy Flight were the first signs. Later, the 2011 Cuckoo optimization algorithm was studied in detail by R. Rajabioun [19]. This one starts with an initial population like other evolutionary algorithms. The following steps are taken with this method. We suppose that there are certain eggs in this population. At first, the host bird will keep these eggs alongside its own eggs in other types of bird's nests. Indeed, this lazy bird makes other birds play an unintended role in the survival of their generation. A few of the eggs less similar to the eggs of the host bird will be identified and destroyed. Cuckoos improve and learn, in fact, to lay eggs more like the eggs of the host bird and to know the fake eggs for host birds. More eggs in each area are survived and more eggs are survived and this area is given more attention, and this is the variable which the Cuckoo optimization algorithm needs to optimize.

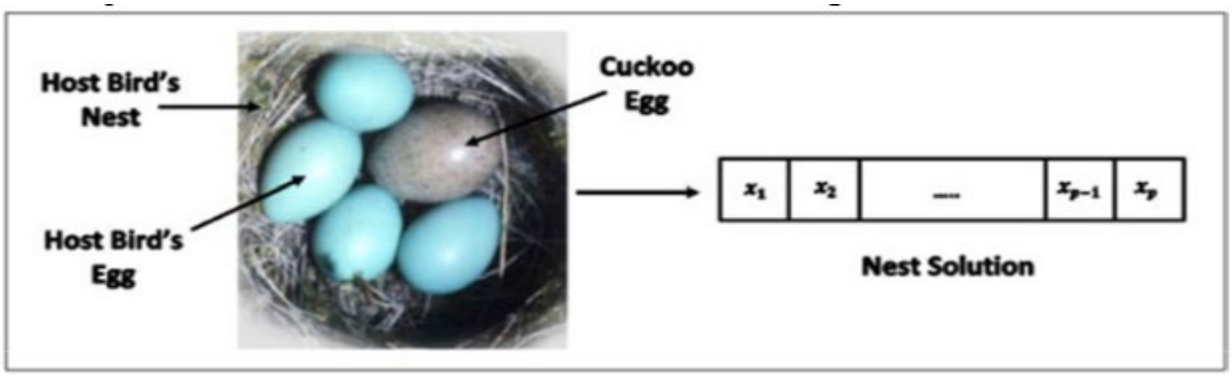


Fig1: In the Cuckoo search algorithm, depiction of a nest solution.

Levy Flights:

In the long run, Levy flights do have authority to perform a comprehensive investigation near the solution, which can then be accompanied by a huge step. In most optimization circumstances, Levy flights are said to be utilised to quickly search for a new optimal method [20]. In order to enhance the solutions, the left leg is linked with the values obtained by Levy flights. To move after one option to additional, the stage unit, that is considered an optional move, is employed. You can move around in search space by making that little step, kth steps, or a large step. These are regulated by Levy flights and are coupled to the 0-1 interval, assisting in the selection of the proper step length.

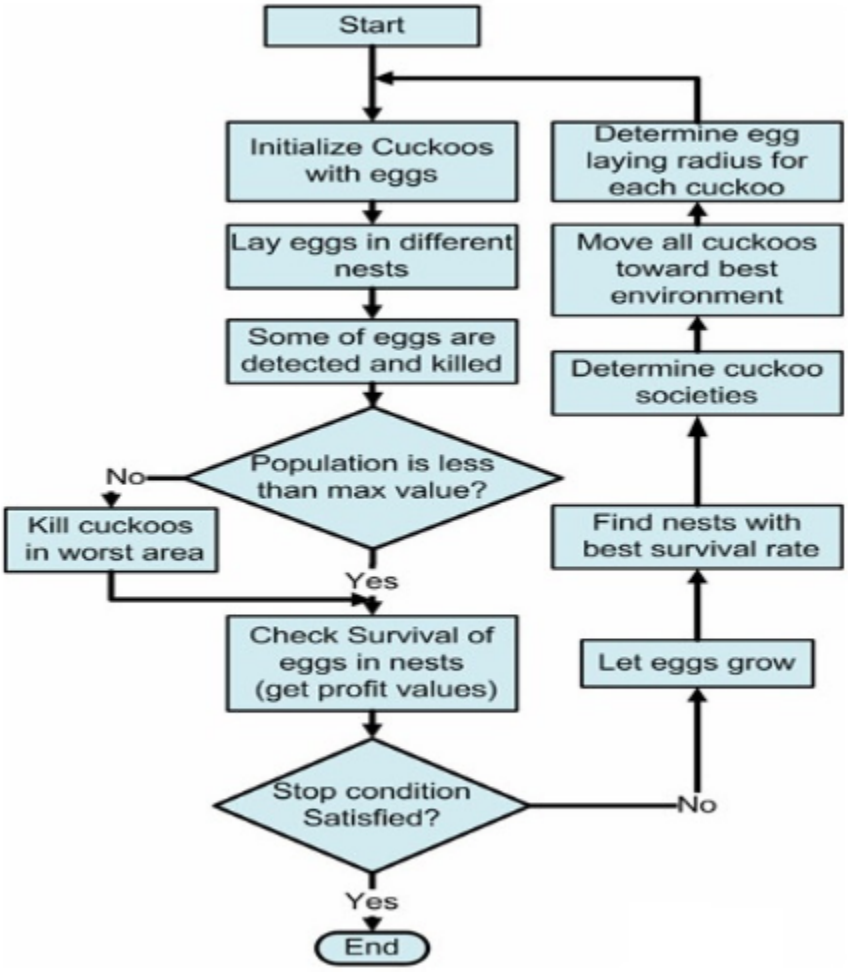


Fig 2: Cuckoo algorithm flowchart

Pseudo code for CSA [21]

1: objective function $f(X), X = (f(x_1, x_2, \ldots\ldots\ldots x_d))^T$
2: Createpreliminary population of n host nest x_i *(i=1,2,3……. n)*
3: ***while*** *t<Max_iteration do*
4: Consider a cuckoo randomly by Levy flights
5: Evaluationof quality F_i
6: Select a nest among n (say, j) randomly
7: ***if*** $F_i > F_j$ ***then***
8: substitute j with newly obtained solution;
9: ***end if***
10: A fraction (P_a) *of worse nests are uncontrolled and chance for the new one are built;*
11: remain with the best solutions
12: replace the best solution rank as the first one
13: ***End While***
14: verify the results and visualization

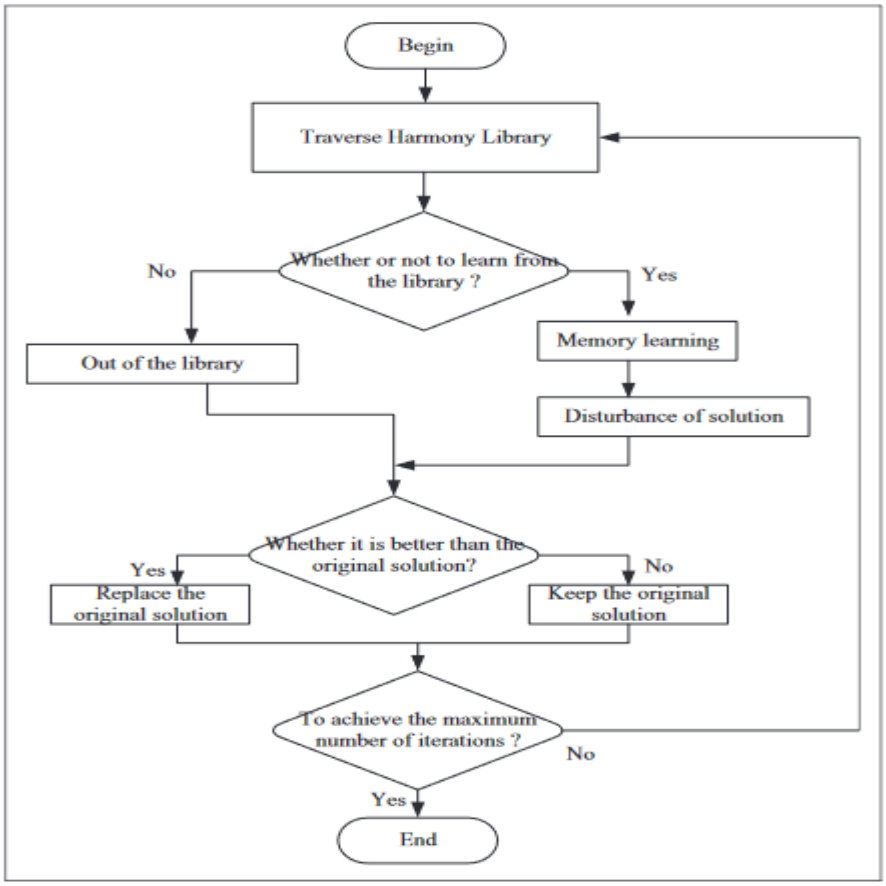


Fig 3: Harmony search algorithm

Harmony Search Algorithm

In general musicians try various variations of the music batters stored in their memories when musicians are composing the harmony. In fact, this search for perfect harmony is similar to the process by which optimum solutions are found to technical problems. HS is actually based on harmony improvisation's working principles.

Pseudo code for HSA [22]

Algorithm:

1: prepare the harmony memory along with HMS erraticallyproducedresolutions
2: ***repeat***
3: follow the steps to get a new solution
4: ∀variables decisions do
5: HMCRProbability uses a value of one of the solutions and change the value slightly with PARprobability
6: then use aarbitrary value for with a choice variable
7: ***end for***
8: in harmony value replace it with the worst solution with the best one
9: a new solution is placed in case of worst solution
10: ***end if.***

5. EXPERIMENTAL SET UP

This work is indented to measure the real-time performance of the cloud resource scheduling algorithms in terms of utilization of cloud resources, cost and makespan time. The standard cloud scheduling procedure codes are fetched from [23] and the proposed method is coded using VC++ programming language.

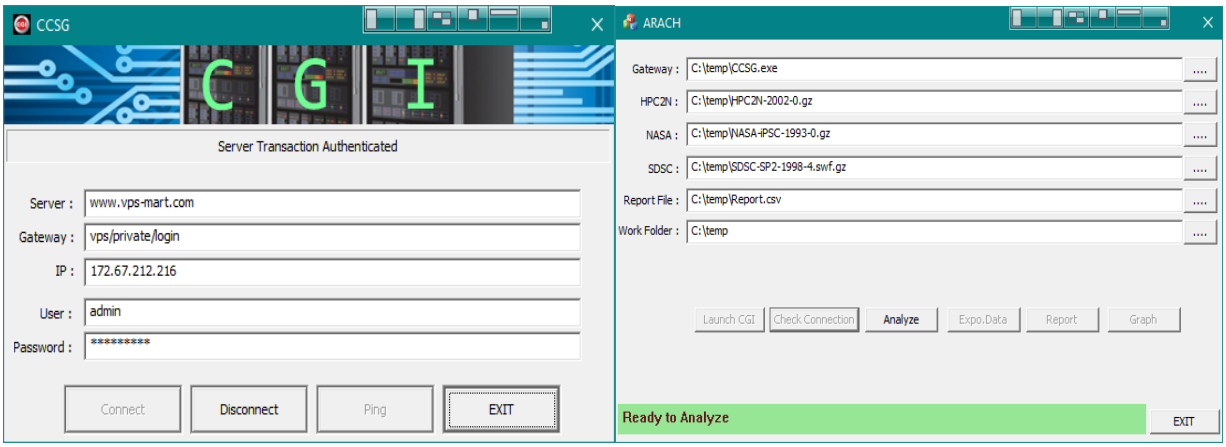


Fig 4: The User interface and CGI data

To communicate with the Common Gateway Interface (CGI), which connects to a leased Windows Virtual Private Server from [24], a dedicated User Interface (UI) is built using Visual Studio IDE [25]. Standard and proposed cloud resource distribution techniques are configured on the Express Windows VPS with dual-core processor, 2GB RAM, 60GB SSD, 50Mbps bandwidth, and Free DNS. The performance monitor measures the required performance parameters and sends them to a specific UI through CGI. The UI can then save the measured data as a report file and plot the corresponding graphs.

6. RESULTS AND DISCUSSION

The proposed approach carried out three parameters makespan time, utilization and cost with respect to three databases especially on NASA, HPC2N and SDSC. Here all the three considered parameters were found with respect to the three databases and were compared with existing and proposed approach.

Makespan:Makespan is a method for determining the maximum execution time by computing the total completion time of all tasks if they're all assigned. The expectations of cloud users will not be met on time if the makespan of a single cloudlet or job is longer than the expectations of cloud users [26]. This study shortens the time it takes to map all tasks on virtual machines.

$$f(x) = \max \prod_{i=1}^{m} T_i \quad \forall \in N, i = 1,2,3, \ldots \ldots . . m$$

The term "makespan time" refers to the amount of time it takes to complete a calculation task. When compared to existing state-of-the-art methodologies, the obtained values clearly show that the proposed method is the best. The higher the value, the shorter the makespan time.

Makespan Time [HPC2N]					
No. of cloudlets	MOCSO	ERMPUCSA	EPMOGA	MCRA	ARACH
200	2044	2793	2729	3008	1676
400	2391	3318	3189	3568	1970
600	2805	3798	3588	4136	2331
800	3246	4289	4017	4662	2633
1000	3615	4835	4478	5208	2984
1200	4020	5384	4969	5769	3407
1400	4429	5886	5445	6422	3734
1600	4904	6427	5977	6963	4100
1800	5359	7044	6410	7601	4561
2000	5818	7577	6914	8247	4958

Table 2: Makespan Time for HPC2N dataset

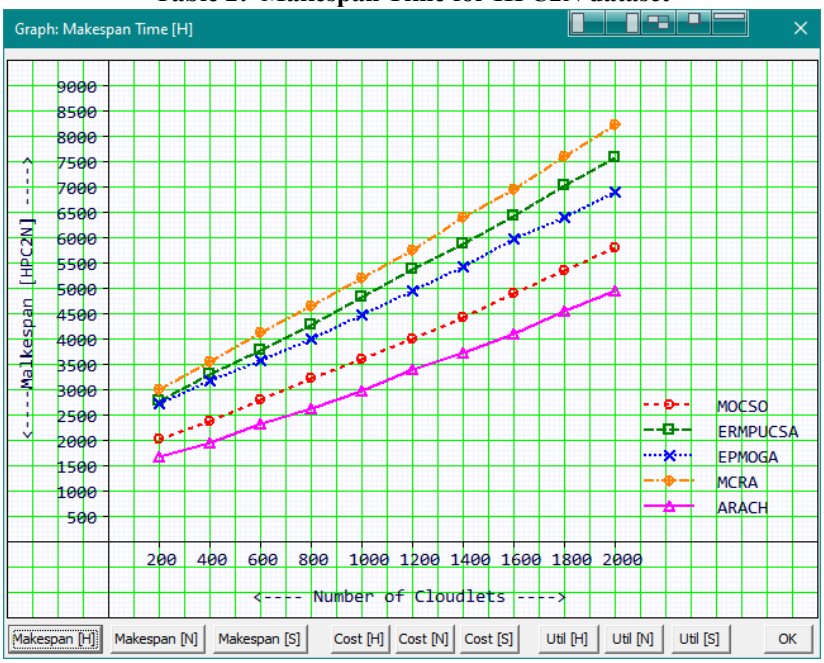


Fig 5: Makespan Time graph for SDCS dataset

Makespan Time [NASA]					
No. of cloudlets	MOCSO	ERMPUCSA	EPMOGA	MCRA	ARACH
200	73	79	86	110	72
400	149	176	141	176	120
600	230	268	227	286	191
800	314	344	341	399	277
1000	401	474	443	521	357
1200	535	611	544	646	454
1400	638	747	674	798	564
1600	782	887	832	960	680
1800	917	1040	965	1133	803

2000	1074	1206	1130	1344	929

Table 3:Makespan Time for NASA dataset

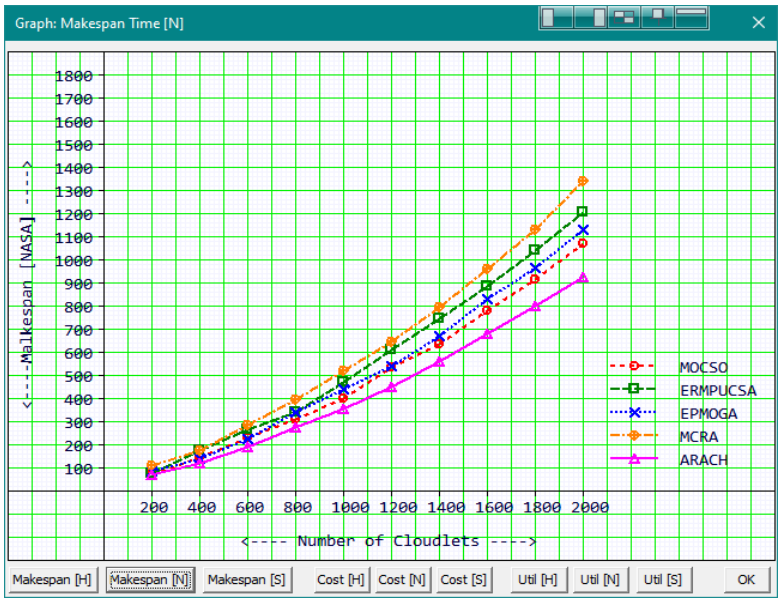


Fig 6: Makespan Time graph for SDCS dataset

Makespan Time [SDCS]					
No. of cloudlets	**MOCSO**	**ERMPUCSA**	**EPMOGA**	**MCRA**	**ARACH**
200	226	350	311	441	160
400	370	627	583	807	394
600	617	924	845	1208	552
800	765	1333	1102	1557	715
1000	968	1714	1429	2082	914
1200	1285	2138	1817	2737	1113
1400	1485	2779	2225	3503	1366
1600	1819	3471	2834	4481	1630
1800	2203	4283	3479	5726	1854
2000	2664	5441	4262	7325	2245

Table 4: Makespan Time for SDCS dataset

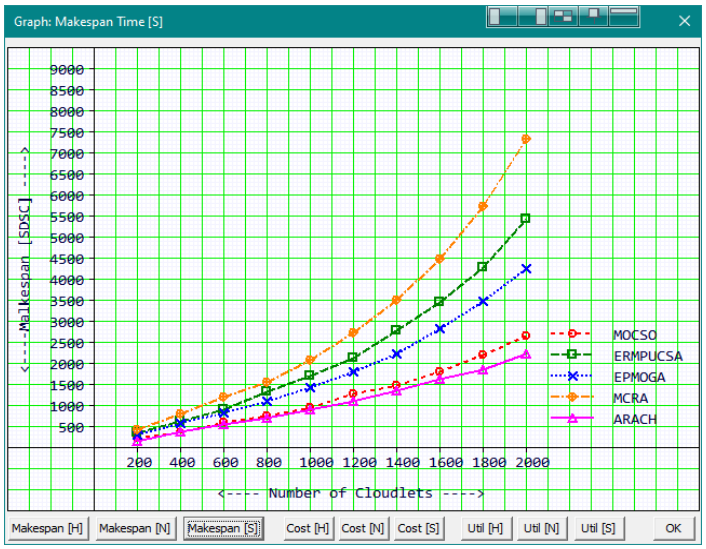


Fig 7: Makespan Time graph for SDCS dataset

Cost:Cost refers to the total quantity produced in relation to the number of resources used or utilized, as well as the amount which cloud providers get from the cloud users. The true goal obtained by minimizing costs for cloud users though increasing gain of income for cloud resource providers through effective utilization [27]. Assuming that only the cost of a VM varies from alternative to alternative depending on significant VMs' and time description as defined by cloud service providers.

$f(y) = \sum_{i=1}^{m} resource^{i}\,(C_i * T_i) \forall\ i \in N, i = 1,2,3, \ldots\ldots m$. Where C_i is the $i^{th} resource\ cost * T_i$ is the i^{th} resource utilization.

When our recommended methodology is applied to three datasets, it is noticed that with less cost, more resources are captured, which is extremely obvious when the acquired results of the suggested method are compared to the existing state of the art. The less the cost values the better is the approach.

Cost [HPC2N]					
No. of cloudlets	**MOCSO**	**ERMPUCSA**	**EPMOGA**	**MCRA**	**ARACH**
200	26	32	34	40	18
400	25	35	34	42	18
600	27	37	38	45	19
800	30	41	41	47	21
1000	33	47	45	53	23
1200	38	50	50	62	27
1400	43	58	55	70	29
1600	51	66	64	78	35
1800	57	75	73	87	39
2000	63	86	81	99	43

Table 5:Cost spent on resources with HPC2N dataset

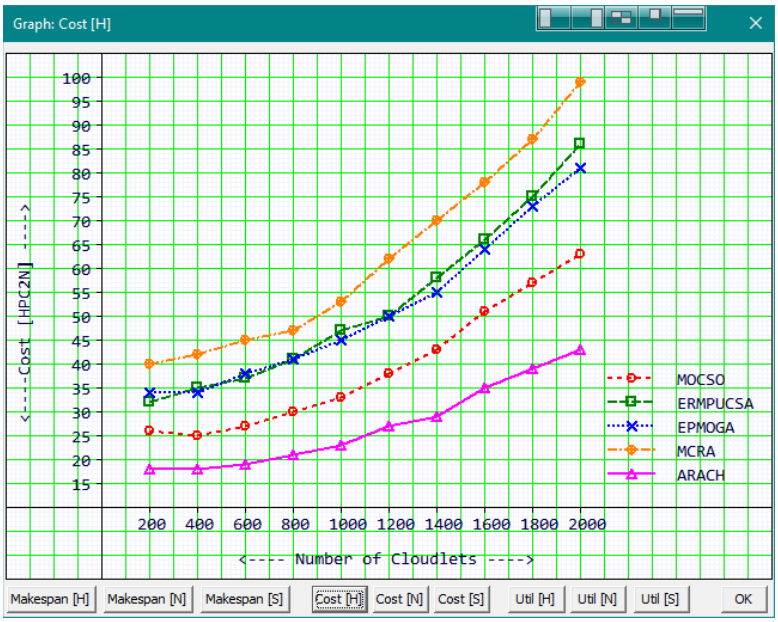


Fig 8: Cost graph for HPC2N dataset

Cost [NASA]					
No. of cloudlets	**MOCSO**	**ERMPUCSA**	**EPMOGA**	**MCRA**	**ARACH**
200	0	12	9	27	0
400	1	19	16	35	0
600	2	28	23	46	1
800	3	36	29	53	1
1000	4	42	36	62	2
1200	5	49	43	70	2
1400	6	54	51	77	3
1600	7	63	57	85	3
1800	8	70	65	91	4
2000	9	76	71	100	4

Table 6: Cost spent on resources with NASA dataset

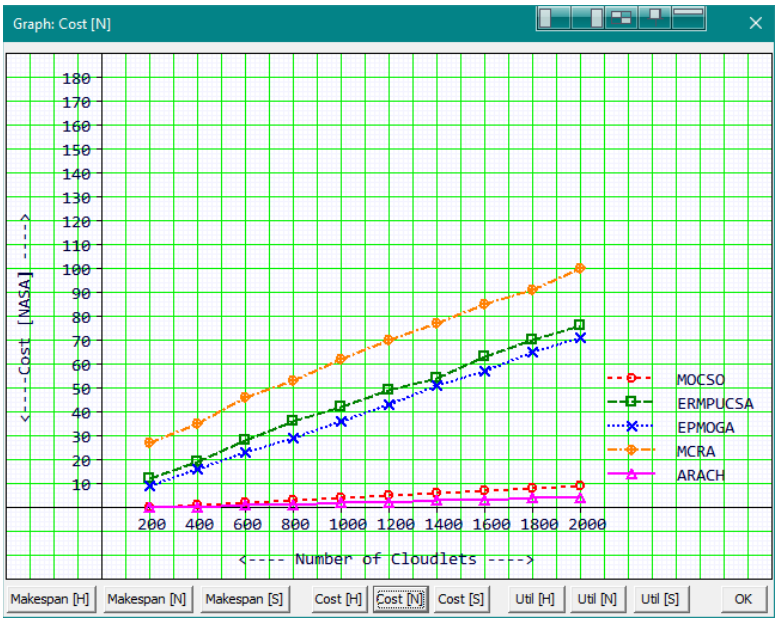


Fig 9: Cost graph for NASA dataset

Cost [SDCS]					
No. of cloudlets	**MOCSO**	**ERMPUCSA**	**EPMOGA**	**MCRA**	**ARACH**
200	1	14	9	19	1
400	3	16	11	22	2
600	4	18	13	25	3
800	6	21	16	28	5
1000	8	24	18	31	6
1200	10	27	21	35	8
1400	13	31	24	40	10
1600	16	35	27	45	12
1800	19	40	32	51	14
2000	22	46	36	58	16

Table 7: Cost spent on resources with SDSCS dataset

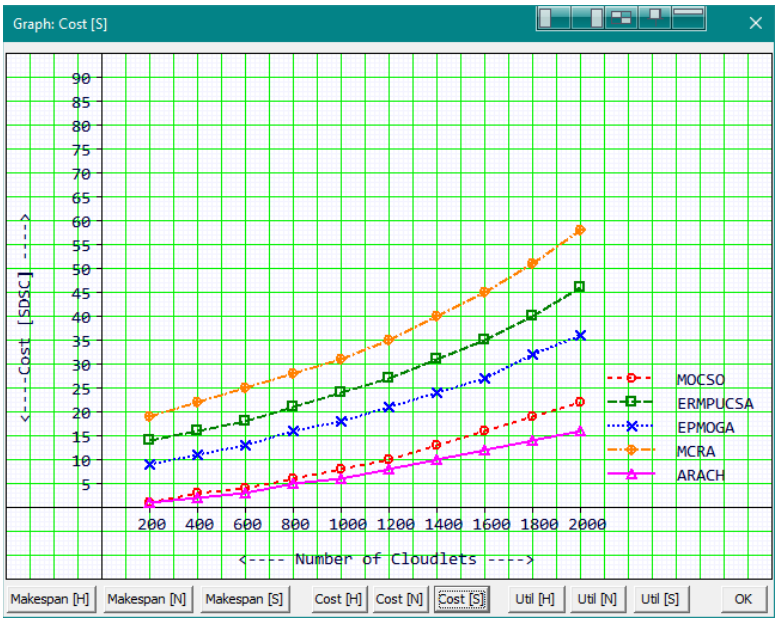


Fig 10: Cost graph for SDCS dataset

Utilization: The total number of resources utilized in data centers is referred to as utilization. The major goal is to make efficient use of resources by competent mapping [28]. It aids cloud providers in increasing profit and revenue while successfully meeting the needs of cloud consumers.

$$f(z) = \frac{\sum_{i=1}^{m} resource^{i}\,(C_i * T_i)}{\max \coprod_{i=1}^{m} T_i}$$

$$\forall\ i \in N, i = 1,2,3, \ldots \ldots . . m$$

When it comes to resource sharing, resource use is crucial. The available resources must be effectively divided in this situation. The greater the number, the more resources are being used. Our proposed strategy appears to have worked well for the three datasets, with more resources being used as evidenced by the obtained findings.

Utilization [HPC2N]					
No. of cloudlets	**MOCSO**	**ERMPUCSA**	**EPMOGA**	**MCRA**	**ARACH**
200	10.81	9.1	8.32	8.25	15.28
400	11.12	9.7	8.61	8.91	14.91
600	10.97	10.26	8.08	9.49	14.3
800	11.16	9.07	8.5	9.17	15.29
1000	11.14	8.69	8.1	9.79	15.5
1200	10.18	9.28	8.35	8.97	15.58
1400	10.91	8.63	8.07	8.95	14.2
1600	10.38	9.37	8.59	8.06	15.4
1800	10.41	9.53	8.77	8.69	14.5
2000	11.28	8.53	8.66	9.68	15.37

Table 8: Utilization of Resources with HP2CN dataset

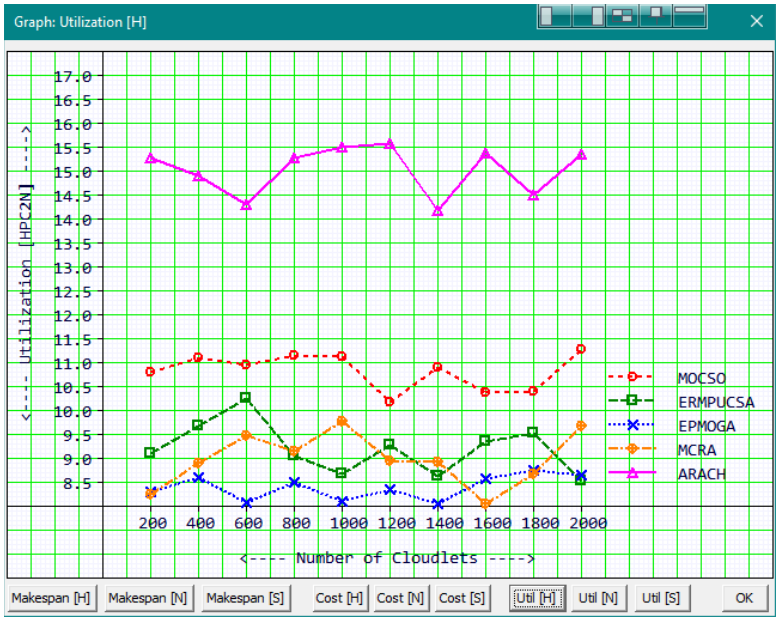


Fig 11: Resources Utilization graph with HPC2N dataset

Higher values of resource consumption are derived from the table, and the same are mimicked in the graph, based on the obtained values. The higher the values, the greater the resource use, and the best resource utilisation is shown. When the proposed method's results are compared to the existing state of the art, it is evident that we achieved better outcomes. For the three considered data sets HPC2N, NASA and SDCS the highest resource utilization values obtained are 15.37, 20.299 and 27.746, and also it is clear that for SDCS dataset the resource utilization is high when compared with the remaining two considered datasets.

Utilization [NASA]					
No. of cloudlets	**MOCSO**	**ERMPUCSA**	**EPMOGA**	**MCRA**	**ARACH**
200	41.9709	37.555302	44.841202	33.7002	54.2083
400	35.18554	33.639729	36.945263	28.3279	49.1093
600	30.39667	25.682405	30.635975	25.4749	45.4748
800	22.64102	21.163153	24.810719	21.203	38.4235
1000	18.7536	16.534086	19.099457	16.022	34.5666
1200	13.72333	12.207283	15.746853	11.3238	31.3136
1400	9.478241	8.913244	11.879416	9.20593	26.6748
1600	6.963081	6.314257	10.443508	6.15615	23.5791
1800	6.261339	4.541913	7.644475	4.84323	21.0689
2000	4.772063	3.228214	8.86231	3.50739	20.2999

Table 9: Utilization of resources with NASA dataset

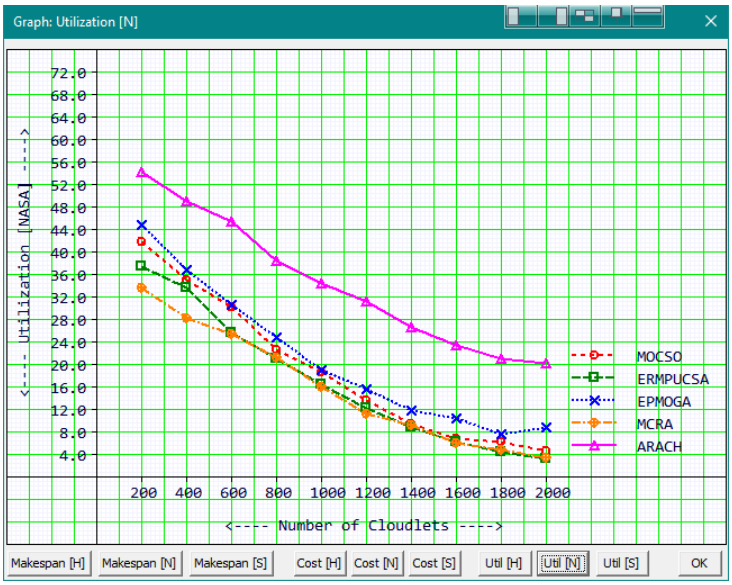


Fig 12: Resources Utilization graph with NASA dataset

Utilization [SDCS]					
No. of cloudlets	**MOCSO**	**ERMPUCSA**	**EPMOGA**	**MCRA**	**ARACH**
200	66.1554	43.171299	58.980099	38.5866	79.9442
400	56.20223	35.108646	51.086987	33.3457	68.9329
600	49.74176	31.756483	41.291817	27.77	61.3782
800	40.86816	27.982384	35.815498	21.3839	51.2146
1000	31.14937	22.473104	27.43243	16.4403	42.8719
1200	25.81346	20.469999	22.486555	10.9415	38.2385
1400	20.41947	15.644387	19.521086	6.04136	33.5171
1600	16.76032	15.442533	14.66276	6.28281	28.2837
1800	15.33714	12.214144	13.700529	3.87673	26.6223
2000	13.4399	12.188146	12.859795	3.87511	27.746

Table 10: Utilization on resources with SDCS dataset

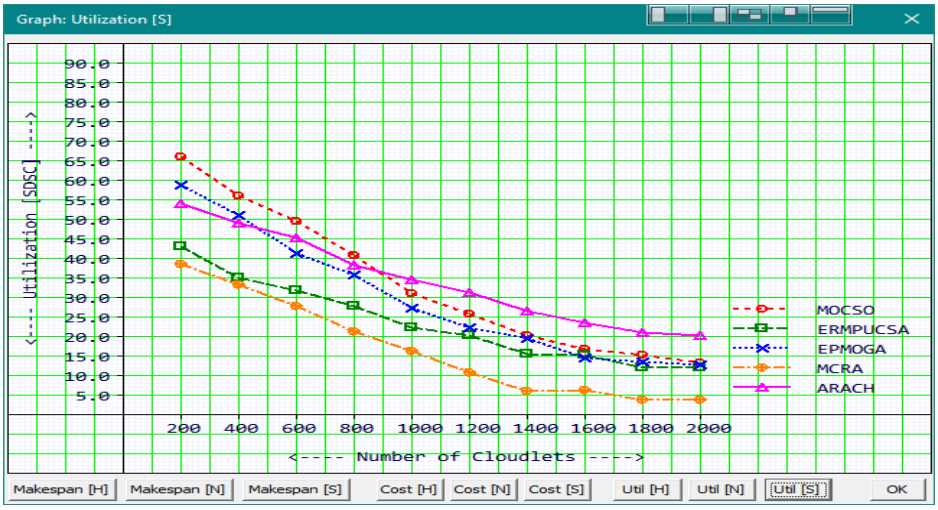


Fig 13: Resources Utilization graph with SDCS dataset

7. CONCLUSION

The proposed approach, which is based on three metrics, namely makespan time, resource usage, and cost spent for resource allocation in the defined simulation environment, was tested on three datasets: NASA, SDCS, and HPC2N. When the suggested methods' findings were compared to the state of the art of existing methods, it was discovered that the proposed methods produced better values for the parameters under consideration. In the future, we'd like to expand our technique which could work well with almost all dataset by including additional parameters and work with dynamical changeable loads, and we'd like to propose a permanent solution in this regard using trending technologies.

REFERENCES

[1] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.I., 2010. A view of cloud computing. Communication. ACM 53 4, 50–58. https://doi.org/10.1145/ 1721654.1721672.

[2]. Zhu, X., Chen, C., Yang, L. T., & Xiang, Y. (2015). ANGEL: Agent-based scheduling for real-time tasks in virtualized clouds. IEEE Transactions on Computers, 64(12), 3389–3403.

[3]. Alkhanak, E. N., Lee, S. P., & Khan, S. U. R. (2015). Cost-aware challenges for workflow scheduling approaches in cloud computing environments: Taxonomy and opportunities. Future Generation Computer Systems, 50, 3–21.

[4]. Shi, T., Yang, M., Li, X., Lei, Q., & Jiang, Y. (2016). An energy-efficient scheduling scheme for time constrained tasks in local mobile clouds. Pervasive and Mobile Computing, 27, 90–105.

[5]. Jiang, H., Yi, J., Chen, S., & Zhu, X. (2016). A multi-objective algorithm for task scheduling and resource allocation in cloud-based disassembly. Journal of Manufacturing Systems, 41, 239–255.

[6]. Choudhary, A.; Gupta, I.; Singh, V.; Jana, P.K.: A GSA based hybrid algorithm for bi-objective workflow scheduling in cloud computing. Future Gener. Comput. Syst. 83, 14–26 (2018)

[7]. Nguyen, T.T.; Vo, D.N.: Modified cuckoo search algorithm for multi-objective short-term hydrothermal scheduling. Swarm Evolution Computing. 37, 73–89 (2017).

[8] Qi Zang, Mohamed FatnZhani, Raouf Boutaba and Jospeh L Hellerstein, ”Dynamic Heterogeneity Aware Resource Provisioning in the Cloud”,IEEE 2013,DOI 10.1109/ICDS.2013.28

[9] M. Dabbagh, B. Hamdaoui, M. Guizani, and A. Rayes, “An energy-efficient VM prediction and migration framework for overcommitted clouds,” IEEE Transactions on Cloud Computing, 2016.

[10] J. Liu, K. L. Li, D. K. Zhu, J. J. Han, and K. Q. Li, “Minimizing cost of scheduling tasks on heterogeneous multicore embedded systems,” ACM Transactions on Embedded Computing Systems, vol. 16, no. 2, 2016.

[11] J. Xu, C. Liu, X. Zhao, S. Yongchareon, and Z. Ding, “Resource management for business process scheduling in the presence of availability constraints,” ACM Transactions on Management Information Systems, vol. 7, no. 3, article 9, 2016

[12] J. E. Kim, X. Fan, and D. Mosse, “Empowering end users for social internet of things,” in Proceedings of the Second International Conference on Internet-of-Things Design and Implementation. ACM, 2017, pp. 71–82.

[13] D. Hussein, S. N. Han, G. M. Lee, N. Crespi, and E. Bertin, “Towards a dynamic discovery of smart services in the social internet of things

[14]Madni, S. H. H., Latiff, M. S. A., Ali, J., &Abdulhamid, S. M. (2018). Multi-objective-Oriented Cuckoo Search Optimization-Based Resource Scheduling Algorithm for Clouds. Arabian Journal for Science and Engineering. doi:10.1007/s13369-018-3602-7.

[15]Alexander, A. A., & Joseph, D. L. (2016). An Efficient Resource Management for Prioritized Users in Cloud Environment Using Cuckoo Search Algorithm. Procedia Technology, 25, 341–348. doi:10.1016/j.protcy.2016.08.116 .

[16] Alexander, A. A., & Joseph, D. L. (2016). An Efficient Resource Management for Prioritized Users in Cloud Environment Using Cuckoo Search Algorithm. Procedia Technology, 25, 341–348. doi:10.1016/j.protcy.2016.08.116

[17] Jindal, H., Singh, H., & Bharti, M. (2018). Modified Cuckoo Search for Resource Allocation on Social Internet-of-Things. 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC). doi:10.1109/pdgc.2018.8745772
[18] X. S. Yang and S. Deb, Cuckoo search via Lévy Flights, In: World Congress on Nature & Biologically Inspired Computing (NaBIC2009). IEEE Publications, pp. 210–214, 2009.
[19] R. Rajabioun, Cuckoo Optimization Algorithm, In: Applied Soft Computing journal, vol. 11,pp.5508 - 5518, 2011
[20] amaruzaman, A.F.; Zain, A.M.; Yusuf, S.M.; Udin, A.: Levy flight algorithm for optimization problems–a literature review. Appl. Mech. Mater. 421, 496–501 (2013).
[21] Shehab, Mohammad, A. Khader and M. Laouchedi. *"*a hybrid method based on cuckoo search algorithm for global optimization problems *1." (2018).*
[22] D. Weyland, A critical analysis of the Harmony Search Algorithm – How not to solve Suduko. Oper. Res. Perspect., 2 (2015) 97-105.
[23] https://github.com/
[24] https://visualstudio.microsoft.com/
[25] https://www.vps-mart.com/windows-vps
[26] Abdulhamid, S.M.; Latiff, M.S.A.; Idris, I.: Tasks scheduling technique using league championship algorithm for makespan minimization in IaaS cloud. ARPN J. Eng. Appl. Sci. 9(12), 2528–2533 (2015).
[27] Madni, S.H.H.; Latiff, M.S.A.; Coulibaly, Y.; Abdulhamid, S.I.M.: Recent advancements in resource allocation techniques for cloud computing environment: a systematic review. Cluster Computing 20(3), 2489–2533 (2017)
[28] Zhang, Q.; Cheng, L.; Boutaba, R.: Cloud computing: state-ofthe-art and research challenges. J. Internet Serv. Appl. 1(1), 7–18 (2010).

ENHANCING EFFORT ESTIMATION IN GLOBAL SOFTWARE DEVELOPMENT USING NEURO-FUZZY DEEP LEARNING NEURAL NETWORKS (NFDLN)

Abstract:*For the successful project planning and management in global software development (GSD), the major issues are cost estimation and effort allocation. The focus of research investigations are to provide timely estimation of the probable software development for enabling the industries to deliver software and in a cost effective manner meeting time deadlines. The researchers worked with various models and algorithms for the GSD projects to punctiliously predict the cost. The global industry is also increasingly producing its products through a diversity of scientific, engineering, technical and managerial teams worldwide. The approximate size of the software is traditionally used to measure software costs and schedules. Advanced estimation methods take into account a variety of variables such as the project's purpose, available personnel expertise, time constraints, efficiency constraints, technology requirements, and so on. Typically, estimation is based on a model developed with the assistance of experienced workers. Estimating software costs is critical since it entails a significant financial and strategic commitment. However, accurate estimation remains a challenge due to the fact that the computational models used for software project management and estimation do not account for the true dynamic nature of software development. To address the complexity and versatility associated with the cost drivers. To achieve at optimized cost factors, we used the cuckoo algorithm in our model. Our proposed model NFDLN (Neuro Fuzzy Deep Learning Neural Network) gives highly accurate effort estimation results when compared with actual cost. The accuracy of the effort prediction in software cost approximation was validated by the parameters Magnitude of Relative Error (MRE), Balanced Relative Error (BRE) and prediction interval. Our proposed model results are better than the state of the art methods.*

Keywords:*Global Software Development (GSD)*, Effort estimation, *Lines of Code (LOC), Software Project Management (SPM). Neuro-Fuzzy Algorithmic (NFA), Artificial Neural Network (ANN)*

1. INTRODUCTION:

One of the most challenging aspects of the project is software estimating. For decades, project managers have struggled to assess the work, cost, and duration of initiatives needed to set schedules and budgets. The challenge is estimating those metrics early in the project lifecycle when the limits of each industry must be determined, and there is significant ambiguity about the ultimate product's functions [1]. Limited awareness of potential contributing factors and dangers, management or client pressure, older software approximation methodologies relying onprofessional opinion can all lead to inaccurate then often over-optimistic projections. Consequently, this may significantly affect completing projects on time, budget, and adequate

standard [2] [3]. Despite continual improvements in project and software development processes, recent surveys show that only one-third of ventures succeed[4]

For the rest, either a cost or time overrun happens, or the product fails to satisfy the customer's expectations, leading to project termination. Although widespread adoption of an agile method to software development has slowed the previously indicated trend, the project failure rate remains high. Despite frequently chastised researchers[4] by the Standish Group, which is primarily aimed at blocking access to assess data and limiting research to American companies, the project success problem is real. It is seen by construction companies daily[5]. The description of McKinsey fits many IT jobs and therefore makes further use of GSD [39] many software features can be offshore for the software.

Sahay states that global development remains unexamined mainly in the empirical context. New organizational patterns develop, and contemporary political problems encounter rapid economic growth in formerly stagnant areas that still live alongside the newly prosperous populations. GSD has become a powerful economically and technologically powerful force in this context. The US, UK, Australia, Western Europe, and Korea are major outsourcing customers. The United States has consumed 5,5 trillion dollars in 2000 and about 17,6 trillion dollars by 2005. Indiaalone is estimated to reach 90 billion dollars by 2010[6]. In Ireland, India, and Israel, major 'Tech nopoles' emerged to emerge sources in Russia, the Philippines, and China[7].

Software is available all around us. Software plays a crucial role in operating everything from kitchen devices to aircraft in the contemporary world of technology. The development of all this source code is now a genuinely global sector. Computers were born about 60 years ago as an international sector and are now a global enterprise that is growing exponentially, scope and geography. This global industry is also increasingly producing its products through a diversity of scientific, engineering, technician, and manager teams worldwide. GSD provides a detailed approach to GSD management and discusses current trends and their future implications. There will be a variety of options and the repercussions of each one. The available supply models run from completely internally to predominantly externalized or offshore. This leads to a debate on the global development of software and its variants. A particular focus will be put on understanding the evolution of this phenomenon and the definition of its core features. Since the software is a global business, companies, countries, and cultures have a great deal of interplay. These aspects will be described, and the patterns of technology transfer will be examined between the countries involved in global software development. These exchanges led to the establishment of widespread practices used in GSD and offshore commitments. A detailed description of these practices provides a practical guide on managing offshore projects and is open to anyone who is confronted with this challenge.The major reasons for this are a shortage of hard and soft expertise in teams, a lack of coordination between stakeholders[8], and inappropriate assessment and planning of the projects[9]. Because once a budget and timetable are defined, it may not be possible to change them due to costs surpassing benefits or a lack of a competitor time to market, precise predicting of effort and timeframe necessary to implement software package at the initial stage markedly increases the chances of project completion. As a result, there may be a reduction in capabilities or even lost opportunities due to this.

To assure quality and monitor the same way that on-site inspectors function in the physical structures of companies, testing firms will emerge. These companies will undertake on-site spot checks, coding standards, code audits, safety tests, and integration tests.As indicated by the increasing expansion in demand for high-quality software and expanding investments in software projects, project management is one of the most important markets in the world. The median company spends 4% to 5% of its turnover on computer technology, with more new businesses, such as banking and telecommunications companies, paying more than 10% [10].In project management, predicting the expense of a software system is a critical task. Cost overestimation and underestimation can both be problematic for a business. Overestimation, for example, can lead to a corporation losing contracts or spending billions, whereas underestimation can lead to low quality, delayed, or incomplete software systems. The cost of software is determined mainly by effort [11]. As a result, models for forecasting software expenses can be utilized as decision support tools, allowing researchers to investigate the impact of specific requirements and development team characteristics on the payment of a project under development. A rapidly evolving market needs software products with ever-increasing capability, reliability, and performance.

Furthermore, to remain competitive, senior managers are increasingly demanding that IT teams produce more quality and functionality with fewer resources than ever before [12]. In conjunction with the growing complexity of software, the global trend towards the transfer of development from individual contract workers to dispersed projects has resulted in a confident basis [40]for making or purchase decisions for software companies.

Software engineers ultimately need to strive to achieve all these goals by using remarkable advances in rapidly changing (and often immature) technology.Cuckoo algorithm used in software effort estimation also play an important role [41][42][43]

2. RELATED WORK

The use of effort estimating Data is becoming more popular, particularly in the corporate world. Recently, there has been a significant change in the business sector that necessitates precise effort assessment to get better results early in the project lifecycle and avoid failure. Estimating the cost and effort of software development is an essential issue for SPM. Predicting software development effort is a key and crucial challenge for any complicated and huge software industry's effective administration. As a result, estimating effort has become increasingly critical [13]. Understanding the current hypothesis around GSD, it is informative to step back a bit and take a glance at how technological developments have occurred over the long term. We must remember that for hundreds of years, the international economy has been with us. GSD differentiates from the old form of trade by semi-connected communications, which were impossible over the centuries. The natural state of trade between GSD countries has been there for ages, however. Although communications were not immediate, technical matters were the subject of a "dialogue" between countries. Chinese paper was published in China a thousand years ago, and in China, Iranian windmills became known. In India, the first cotton textiles were developed, but the whirling wheel was transported to India from Iran.Furthermore, estimates for any program are to save money by maintaining a competitive balance between software quality

and cost while also considering the relevance of profit. Avoiding project abortion and restarts is another crucial part of project management [14].

Effort Estimation of GSD objectives:

- **Benchmarking in projects:** The effort estimating approach should allow for software project benchmarking in progression effort and productivity. Comparing software project economic output and exertion between different companies is now very important, in the context of fast-growing global development, to enable us to make a decision and select/manage software providers.
- **Change in management:** The cost estimate strategy should be straightforward to reapply during the software development cycle to facilitate the management of changes in the project scope, including changes in the (non-)functional software requirements.
- **Improvement in the process:**The effort estimating approach should aid in the comprehension and enhancement of development processes involving effort and productivity.
- **Improving productivity:**The effort estimating approach should help identify the aspects that have the most impact on development productivity. At least, in the long run, achieving objectives usually entails increasing productivity; nevertheless, productivity can be increased without improving related procedures, for example, by conveying more capable individuals to the project rather than enlightening training developments.
- **Managing risk in projects:** The method for estimating work should help with project risk management. This involves the bidirectional incorporation of effort estimate and risk management, in which effort approximation uses risk management's outputs and gives hazard management with input. Furthermore, the technique should address estimated uncertainty explicitly.
- **Managing project overhead:** The management overhead, or the expense of adopting and keeping an estimation approach, should be as low as possible.
- **Project cost negotiation:** In computer procurement, the effort estimating approach should help the communication and conflict resolution process between the players involved.
- **Tracking and planning of projects:** By delivering accurate and trustworthy projections at multiple levels of project granularity, the effort estimating approach should enable good project planning. Depending on the project type, there are two main approaches to project planning. In the form of total projects, effort estimation of GSD is the foundation for software planning capability. In the setting of repaired projects, effort estimate serves as a foundation for scheduling and budgeting software projects.

3. EXISTING METHODS:

EEDANN: Easy way of comparisons was discovered**.** Estimation of effort quickly is the advantage. The main methodology used in this approach is MATLAB 10 NN Tool**.**[15] used a 2-2-1 architecture back-propagation ANN on a NASA dataset with 18 projects. The input was KDLOC, and the Output was development approach and effort. He received an MMRE score of

11.78. [16] utilized the MATLAB NN toolbox to predict effort. He had employed Lopez-recommended Martin's dataset, which comprises 41 project data. He used the DC, MC, and LOC as inputs to create a 3-3-1 neural network. The only outcome was development time.Being domain-specific is the drawback of this approach.

EEEAM: The main contributory approach of this method is reduced measure and effort estimation of GSD.It uses IEAM-RP Methodology. The advantages of this approach lie in managing the diversity preservation and capabilities of convergence speed. Because of their population-based search strategy, nature-inspired algorithms were shown to be more suitable for tackling software cost estimates challenges. In recent years, many nature-inspired methods have been proposed for the cost assessment of software projects. [17] proposed a memetic approach for estimating software costs. In [18], IEAM-RP is also used to calculate the modularity value. In this study, IEAM-RP is used to tune the Sheta model's parameters (a, b, c, d) for which MRE and MMRE are optimized over current models. For evaluating software development efforts, the authors of [19] proposed a genetic algorithm-based parameter tuning strategy for the COCOMO II model. One of the evolutionary algorithms used to address single-objective optimization problems is the IEAM-RP [20].Limited to only two parameters is the only pitfall of this Model.

Author	Contribution	Methodology	Advantage	Limitations
Jagannath S et al. [21] (EEDANN)	An easy, effective way of comparisons	MATLAB10 NN Tool	A moderate value effort estimation of GSD	Domain-specific
Tribhuvan S et al. [22] (EEEAM)	Reduced Measure and effort estimation of GSD	IEAM-RP	Diversity preservation and convergence speed capabilities	Limited to only two parameters
Y Amelia E et al. [23] BAT	The effort estimation of GSD produces values having more closer to actual effort for a given dataset	BAT Algorithm	The smallest MMRE values are generated	A bit time-consuming process
Przemyslaw P et al. [24] EML	Provides a quick decision support tool	SVM,MLP,GLM	Results are accurate	The accuracy of the method needs to be validated
P Rajwani et al. [25] MLFNNBP	Better effort estimation of GSD results	MLFFNN	Better Results and accurate forecast for effort estimation of GSD	It does not work for determining the uncertainty

Table 1: compares all the existing methods.

BAT: Obtaining effort estimation of GSD values of software values which are very nearly the same as the values obtained from a project. A new BAT algorithm is used—calculation of very minimal MMRE values. COCOMO model optimization strategies include neural networks and fuzzy logic [26]. To solve the problem, a multi-layer Neural Network with feed-forward and back-propagation is commonly used. Only the hidden layer utilizes a sigmoidal activation function, while the output layer utilizes a linear process. The NASA dataset was also used to test this Model. In 1981, Boehm published COCOMO [27], a research paper describing the effort and schedule estimation model. COCOMO became one of the most prominent estimating models in the 1980s as a result of Boehm's publication. In the year 2000, Boehm created COCOMO II, a fresh approach to COCOMO. This new technique is superior to COCOMO since cost drivers have been improved, and COCOMO II is also more accurate in resolving software project estimates issues. The negative side of this approach is that it shows the project is taking a bit of a time-consuming process.

EML:Provides a successful support tool that helps make quick decisions. Some of the popular Machine Learning Algorithms such as SVM, MLP, GLM is used. [28]The MMER,MBRE, and MAR are still extensively employed by academics as evaluation standards to compare outcomes. Furthermore, to prevent the overfitting of ML models, a k-fold cross-validation approach is used for validation.Aside from the many strategies available, such as deletion,imputation, and their application, rely on the dataset, according to [29]. The obtained results are accurate, which is the main plus point to be considered. Obtained results accuracy needs to be validated.

MLFNNBP: This approach is famous in estimation efforts better compared to the existing methods. It used the MLFFNN methodology. Because of its ease of use and ability to measure effort in person-months for a project at multiple junctures, [30] is one of the most widely used empirical cost modeling methods in the software industry. [31] experimented with using ANN to estimate software effort. ANN training is essentially a nonlinear least-squares problem that can be tackled using typical nonlinear least-squares approaches. [32] is now in use. The hidden layer uses a tangent sigmoid transfer function, while the output layer uses a linear one.Accurate results and forecasts are obtained. Its drawback lies in uncertainty determination.

4. PROPOSED METHOD

Our proposed method uses two algorithms, the cuckoo algorithm, and the Neuro-fuzzy algorithm, for effort estimation of GSD.

Cuckoo algorithm:

Start

Step 1. Initialize

initialize the generation counter $G=1$ and the detection rate p_a; initialize the population P of n host nests arbitrarily.

Step 2. While $G<MaxGeneration$ do

all cuckoos are sorted.

Arbitrarily choice a cuckoo i and contrivance Levy flights to substitute its resolution.

Appraise its fitness F_i.

Arbitrarily pick another nest j.

if(F_i<F_j)
Substitute j by the new resolution.
end if
Arbitrarily produce a fraction (p_a)of new nests and substitute the inferior nests.
Retain the most delicate nests.
Categorize the population and discover the best cuckoo for this cohort.
Pass the existing best to the following generation.
G=G+1.

Step 3. end while
End.

Neuro-fuzzy algorithm for effort estimation of GSD

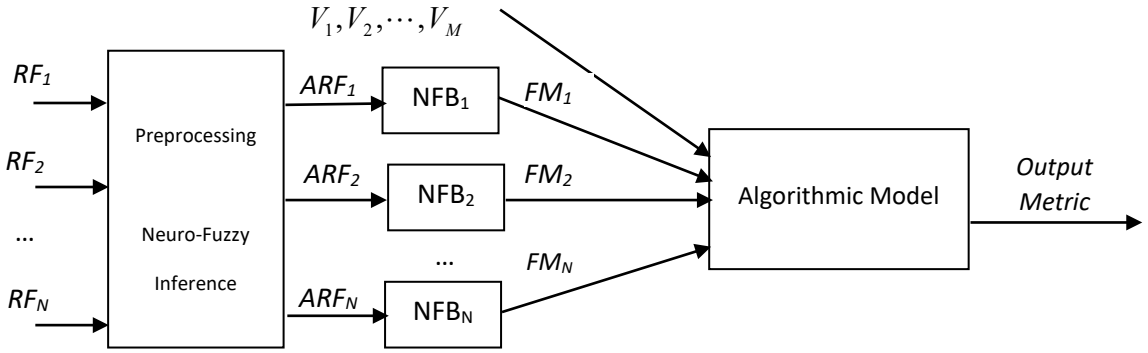


Figure 1. NFA Model for estimation

Here

ARF is AdjustableFactor Rating
FM is Numerical Factor
Mo is Output based Metric.
M is No. of another Algorithmic variable in the Model
N is the No. offactors helping for contribution
NFB is the Neuro based Fuzzy Bank,
RF helps rate the factor
V is Algorithmic Model input,

Neuro-Fuzzy Algorithm(NFA)

Step 1: receive and adjust contribution factors rating as input
Step 2: calculate membership values for every fuzzy rule (I,k)
Step 3" calculate the firing strength for every fuzzy rule (I,k)
Step 4: calculate the summation for all firing strengths
Step 5: normalize the firing strength for every fuzzy rule.
Step 6: calculate the reasoning result for each fuzzy rule (I,k)
Step 7: compute the final reasoning result for all the proposed rules

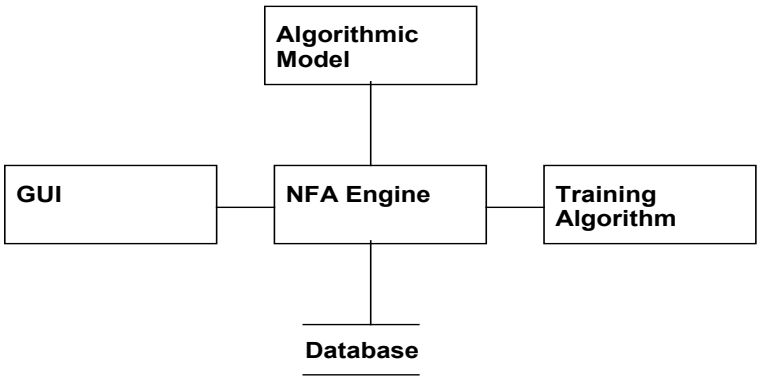


Figure 2. Neuro-Fuzzy Algorithmic (NFA) Tool for Estimation

Activation function used

$$O_N^i = \mu_{RF_N}(ARF_i)$$

There will be a common front-end GUI, back-end database, and training algorithm. Various types of estimation will be simple plug-in of algorithmic models to the NFA engine, and the effort for additional extension will be small.

- ***GUI Front-End:*** Two implementations are available for the GUI front end. Figure 1 initializes the PNFIS on the first screen. The user can change the definition of interdependence from different contributing factors based on predefined defaults. The second display uses the pattern of the algorithm. The user is able to enter factor values and give other inputs to the Model. The Output of the GUI is a set of the NFA-motor model outputs.

- ***NFA Engine***: The NFA motor is NFA Tool's brain. The NFB, as is shown in the above figure, is implemented.

  Database:The database includes appreciated Data for comparison of actual and estimated data. Comparison is aimed at training the NFA motor and improving it. The project data is included in the algorithmic Model and grows with the user's collected local project data. For various project data points, different weights can be assigned in the NFA engine training. Also included in the database are parameter values of different ratings. A set of failures, PNFIS (Pre-Processing Neuro-Fuzzy Inference System) overwritten values or values can be included in the values of the parameter as the NFA engine matures.

- ***Training Algorithm:***If the engine is to be perpetually strengthened, the learning algorithm is essential. A simple MMRE mechanism is used to evolve the parameters with project data. The factor ratings in the parameters are monotonous.

5. EXPERIMENTAL SETUP:

NASA dataset from Link [33] is used to evaluate the Proposed NFDLNN method. Dataset is divided into ten equal segments, and the benchmark parameters such as BRE, MRE, PI, MBRE, and MMRE are measured. An exclusive User Interface (UI) is designed using Visual Studio IDE [34] to evaluate the proposed method and generate the report file based on the observed results. The UI uses Operations Research SDK Tools (OR-Tools) [35] to perform standard and customized processes for calculating the benchmark parameters. The dataset input is accessed in ARFF [36] format, and the report file is generated in CSV [37] format.

6. RESULTS AND DISCUSSION

The parameters considered to show the strength of the proposed method are MRE, BRE,PI.

The magnitude ofRelative Error (MRE):The average of MRE across a set of n observations is MMRE. MMRE can be expressed mathematically as follows: it is denoted as

$$MRE_i = \frac{|Actual\ Effort - Predicted\ Effort|}{Actual\ Effort}$$

MRE	
% Of Data	MRE
10	3.22
20	3.32
30	3.25
40	3.05
50	3.4
60	3.03
70	3.44
80	3.62
90	3.35
100	3.61

Table 2: Showing obtained MRE values from the proposed method concerning % of data

Where MMRE

$$MMRE = \frac{1}{N}\sum_{1}^{N} MRE_i$$

MMRE					
EEDANN	EEEAM	BAT	EML	MLFNNBP	NFDNLL
4.98	18.84	34.25	17	14.48	3.33

Table 3: MMRE values obtained from the existing and proposed methods

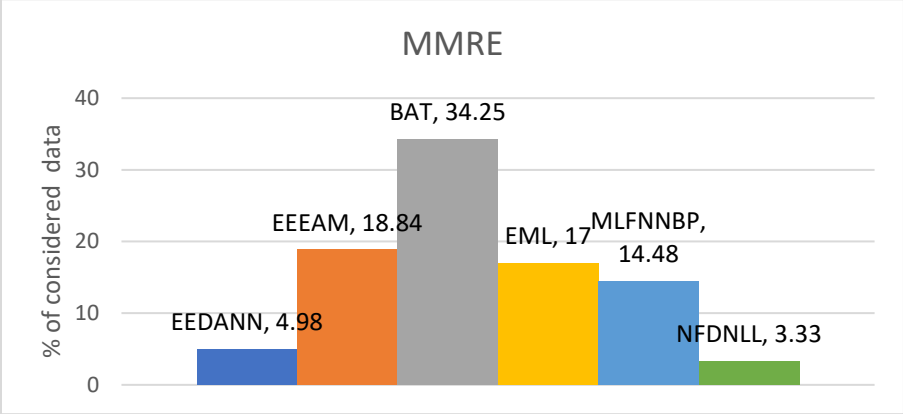


Figure 3: MMRE values for various existing methods along with the proposed method concerning the percentage of Data considered

Balanced Relative Error (BRE):Lesser the BRE values indicate the stability of a system in prediction; from the above values, it is clear that the proposed method BRE values are much less compared with the state of the art of the existing approaches. It is defined as

$BRE = \frac{Actual\ effort - Estimated\ effort}{\min(Actual\ effort - Estimated\ effort)}$ **Mean of BRE is termed as MBRE**

BRE	
% Of Data	BRE
10	0.1253
20	0.1305
30	0.1316
40	0.1347
50	0.1269
60	0.1342
70	0.1332
80	0.1251
90	0.1263
100	0.1348

Table 4: Showing obtained BRE values from the proposed method with respect to % of data

Figure 4: showing the obtained BRE values vs. % of data from the proposed method

MBRE		
EEDANN	EML	NFDNLL
0.16	0.2	0.13

Table 5: MBRE values obtained from the existing and proposed methods

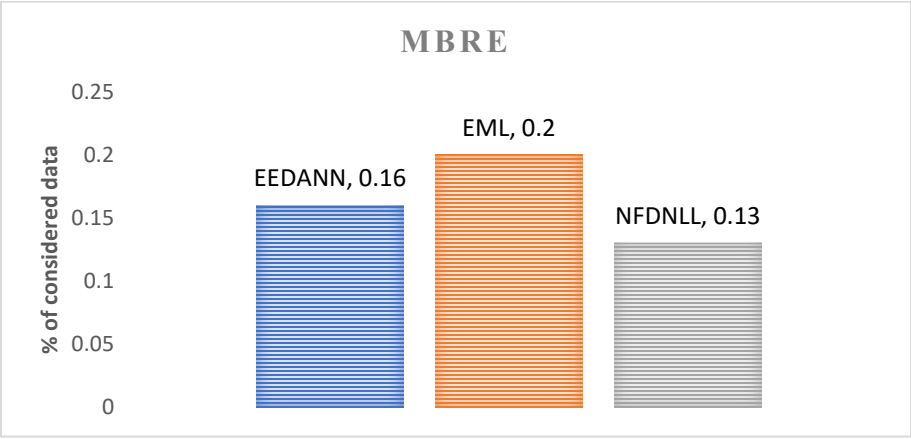


Figure 5:Obtained MBRE values for various methods with respect to the percentage of considered Data

Prediction Interval (PI):The prediction interval denotes the time period between one parameter and the next; the smaller the gap, the more useful the method. It is defined as

$$PI = \begin{Bmatrix} \frac{Estimated\ effort_{\ new\ project}}{1 - BRE} & BRE \leq 0 \\ Estimated\ Effort\ (1 + BRE), & BRE > 0 \end{Bmatrix}$$

PI	
% Of Data	PI
10	75.14
20	75.73
30	73.56
40	73.52
50	74.1
60	75.22
70	74.05
80	74.6
90	75.03
100	73.81

Table 6: Showing obtained PI values from the proposed method concerning % of data

It is evident from the acquired values that almost all existing and suggested methods are operating in parallel with the identical values, causing all associated methods graphs to be superimposed over each other. Its mean values are obtained by averaging all the PI values obtained for a different amount of data collection. This implies that PI still needs to be improved.

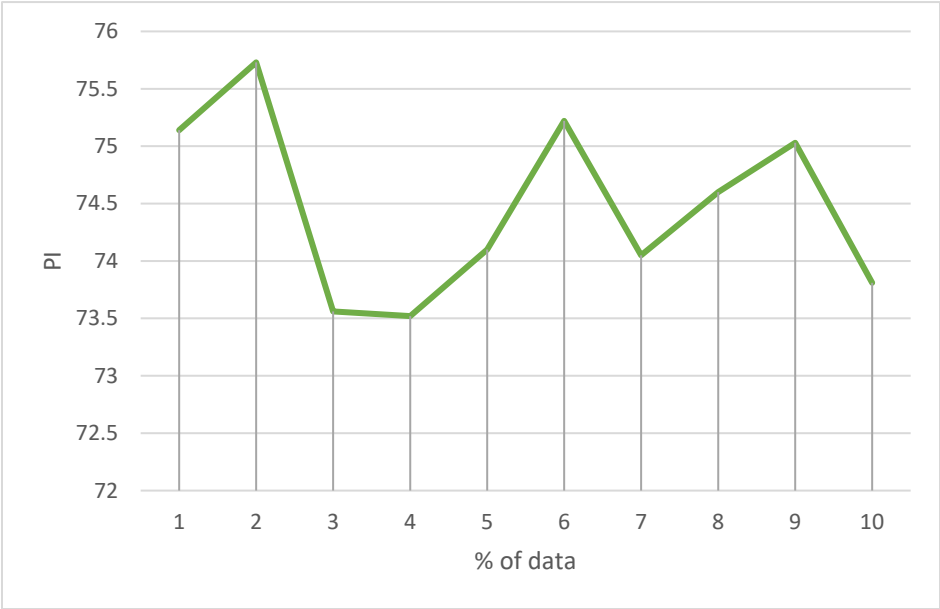


Figure 4: Obtained PI values from the proposed method concerning % of data

MPI

EEDANN	EEEAM	EML	NFDNLL
70	72.42	69	74.48

Table 7: MPI values obtained from the existing and proposed methods

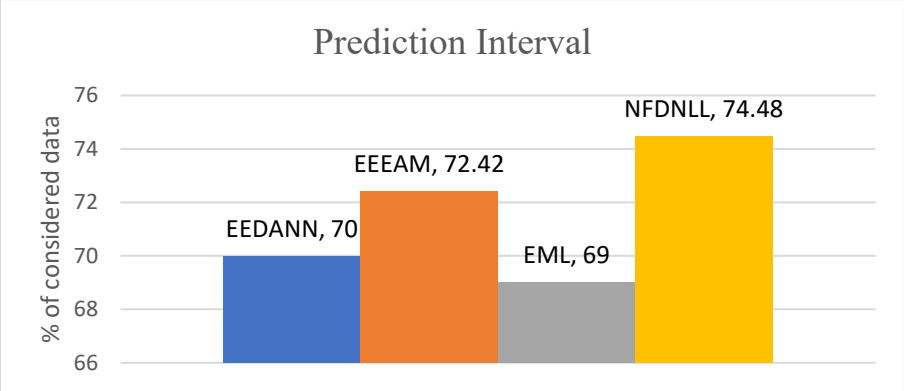


Figure 4:Predition Interval values obtained for various methods concerning the percentage of considered Data

7. CONCLUSION

We proposed a new method for estimating GSD effort based on Neuro Fuzzy deep learning Logic and a combination of Cuckoo search in our technique, which helps to improve the efficiency of the global software development effort.The obtained results were further refined with the deep learning approach which is used in the proposed method. As of total their parameters were considered with means as MMRE, BRE, PI. The obtained values from the proposed method seem to be novel as 3.33, 0.13, 74.48, (average values obtained from the table 2,4,6) on NASA dataset the parameters were calculated by considered parameters over percentage of data. The lower MMRE and BRE values indicates that this method is more accurate and stable than the existing methods. The obtained results are satisfactory when compared with the state of the art methods. We plan to expand our specific method by adding more parameters that will help us calculate GSD's more accurate effort in the future

REFERENCES

[1] Boehm, Barry W. "A spiral model of software development and enhancement." Computer 21.5 (1988): 61-72.

[2]. Alami, A. (2016). Why Do Information Technology Projects Fail? Procedia Computer Science, 100, 62–71. DOI: 10.1016/j.procs. 2016.09.124.

[3]. Minku, L. L., & Yao, X. (2013). Ensembles and locality: Insight on improving software effort estimation. Information and Software Technology, 55(8), 1512–1528. doi:10.1016/j. info. 2012. 09. 012

[4] Eveleens, J. Laurenz, and Chris Verhoef. "The rise and fall of the chaos report figures." IEEE Software 27.1 (2010): 30.

[5] Carvalho, M. M. de, Patah, L. A., & de Souza Bido, D. (2015). Project management and its effects on project success: Cross-country and cross-industry comparisons. International Journal of Project Management, 33(7), 1509–1522. doi:10.1016/j.ijproman.2015.04.004.
[6] Lewin A., and Mani M., Next-generation offshoring: The globalization of innovation. Offshore Research Network, April 11, 2007.
[7] Sahay S., Nicholson B., and Krishna S., 2003. Global IT Outsourcing: Software Development across Borders. Cambridge University Press.
[6] Lehtinen, T. O. A., Mäntylä, M. V., Vanhanen, J., Itkonen, J., & Lassenius, C. (2014). Perceived causes of software project failures – An analysis of their relationships. Information and Software Technology, 56(6), 623–643. doi:10.1016/j.infsof.2014.01.015 .
[7] Savolainen, P., Ahonen, J. J., & Richardson, I. (2012). Software development project success and failure from the supplier's perspective: A systematic literature review. International Journal of Project Management, 30(4), 458–469. doi:10.1016/j.ijproman.2011.07.002 .
[8] Schwalbe, Kathy. Information technology project management. Cengage Learning, 2015.
[9] Trendowicz A., Münch J., Jeffery R. (2011) State of the Practice in Software Effort Estimation: A Survey and Literature Review. In: Huzar Z., Koci R., Meyer B., Walter B., Zendulka J. (eds) Software Engineering Techniques. CEE-SET 2008. Lecture Notes in Computer Science, vol 4980. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22386-0_18
[10] Charette, R.N.: Why Software Fails (Software Failure). IEEE Spectrum 32(9), 42–49 (2005).
[11] R. Agarwal, M. Kumar, Y. Mallick, R. Bharadwaj, D. Anantwar, Estimating software projects, Software Engineering Notes 16 (4) (2001) 60–67
[12] Jørgensen, M., Løvstad, N., Moen, L.: Combining Quantitative Software Development Cost Estimation Precision Data with Qualitative Data from Project Experience Reports at Ericsson Design Center in Norway. In: International Conference on Empirical Assessments of Software Engineering (2002).
[13] Balich, I. and Martin, C. (2010) 'Applying a feed-forward neural network for predicting software development effort of short-scale projects,' in Eighth ACIS International Conference on Software Engineering Research, Management, and Applications, IEEE Computer Society, IEEE, pp.269–275.
[14] Molokken, K. and Jorgensen, M. (2005) 'A comparison of software project overruns – flexible versus sequential development models,' IEEE Transactions on Software Engineering, October, Vol. 31, No. 9, pp.754–766.
[15] Jaswinder Kaur, Satwinder Singh, Dr. Karanjeet Singh Kahlon, PourushBassi, "Neural Network-A Novel Technique for Software Effort Estimation," International Journal of Computer Theory and Engineering, Vol. 2, No. February 1 2010, page:17-19.
[16] Rohit Bhatnagar, Vandana Bhattacharjee and Mrinal Kanti Ghose, "Software Development Effort Estimation –Neural Network Vs. Regression Modeling Approach", International Journal of Engineering Science and Technology, Vol. 2(7), 2010, page: 2950- 2956.

[17] K. Mishra, A. Tripathi, S. Tiwari, and N. Saxena. Evolution-based memetic algorithm and its application in software cost estimation. Journal of Intelligent & Fuzzy Systems, 32(3):2485–2498, 2017.
[18] K. K. Mishra, H. Bisht, T. Singh, and V. Chang. A direction-aware particle swarm optimization with a sensitive swarm leader. Big Data Research, 2018
[19] C. S. Yadav and R. Singh. Tuning of COCOMO II model parameters for estimating software development effort using ga for promise project data set. International Journal of Computer Applications, 90(1), 2014.

[20] T. Singh, A. Shukla, and K. Mishra. Improved environmental adaption method with real parameter encoding for solving optimization problems. In Advances in Computer and Computational Sciences, pages 13–20. Springer, 2018.
[21] Singh, J. and Bibhudatta Sahoo. “Software Effort Estimation with Different Artificial Neural Network.” (2011).
[22] Singh, T., Singh, R., & Mishra, K. K. (2018). Software Cost Estimation Using Environmental Adaptation Method. Procedia Computer Science, 143, 325–332. doi:10.1016/j.procs.2018.10.403
[23] Amelia Effendi, Y., Sarno, R., & Prasetyo, J. (2018). Implementation of Bat Algorithm for COCOMO II Optimization. 2018 International Seminar on Application for Technology of Information and Communication. doi:10.1109/isemantic.2018.8549699
[24] Pospieszny, P., Czarnacka-Chrobot, B., & Kobylinski, A. (2018). An effective approach for software project effort and duration estimation with machine learning algorithms. Journal of Systems and Software, 137, 184–196. doi:10.1016/j.jss.2017.11.066
[25] Rijwani, P., & Jain, S. (2016). Enhanced Software Effort Estimation Using Multi-Layered Feed Forward Artificial Neural Network Technique. Procedia Computer Science, 89, 307–312 . DOI: 10.1016/j.procs.2016.06.073
[26] K. Langsari and R.Sarno. "Optimizing Effort and Time Parameters of COCOMO II Estimation using Fuzzy Multi-Objective PSO." International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2017. https://doi.org/10.1109/EECSI.2017.8239157.
[27] M. Alajlan, "Optimization of COCOMO II Model for Effort and Development Time Estimation using Genetic Algorithms," no. March 2016.
[28] Idri, A., Hosni, M., & Abran, A. (2016). Improved estimation of software development effort using Classical and Fuzzy Analogy ensembles. Applied Soft Computing, 49, 990–1019, doi:10.1016/ j.asoc. 2016.08.012.
[29] Huang, J., Li, Y.-F., & Xie, M. (2015). An empirical analysis of data preprocessing for machine learning-based software cost estimation. Information and Software Technology, 67, 108–127. DOI: 10.1016 /j.infsof.2015.07.004.
[30] M. Madheswaran and D. Sivakumar, Enhancement of Prediction Accuracy in Cocomo Model for Software Project using Neural Network, In Proceedings of International Conference on IEEE Computing, Communication and Networking Technologies (ICCCNT), 2014, pp. 1–5, (2014).
[31] Anupama Kaushik, Ashish Chauhan, Deepak Mittal, and Sachin Gupta, Cocomo Estimates using Neural Networks, International Journal of Intelligent Systems and Applications, vol. 4(9), pp. 22, (2012).
[32] S. Ajitha, T. V. Suresh Kumar, Evangelin D. Geetha, and K. Rajani Kanth, Neural Network Model for Software size Estimation using use Case Point Approach, In International Conference on IEEE Industrial and Information Systems (ICIIS), pp. 372–376, (2010).
[33]. http://promise.site.uottawa.ca/SERepository/datasets/cocomonasa_v1.arff
[34]. https://visualstudio.microsoft.com/
[35]. https://developers.google.com/optimization/install
[36]. https://www.cs.waikato.ac.nz/ml/weka/arff.html#:~:text=An%20ARFF%20(Attribute%2D Relation %20File,the%20 Weka%20machine%20learning%20software.
[37]. https://docs.fileformat.com/spreadsheet/csv/
[38]Hastie, Shane, and Stéphane Wojewoda. "Standish group 2015 chaos report-q&a with jennifer lynch." (2015
[39]https://www.mckinsey.com/careers/search-jobs/jobs/analyst-software engineer data engineer data scientist campusstudents-41876.

[40] Jørgensen, M., & Sjøberg, D. I. K. (2003). An effort prediction interval approach based on the empirical distribution of previous estimation accuracy. Information and Software Technology, 45(3), 123–136. doi:10.1016/s0950-5849(02)00188-x

[41]Kumari, S., & Pushkar, S. (2018). Cuckoo search based hybrid models for improving the accuracy of software effort estimation. Microsystem Technologies. doi:10.1007/s00542-018-3871-9.

[42]Puspaningrum, A., & Sarno, R. (2017). A Hybrid Cuckoo Optimization and Harmony Search Algorithm for Software Cost Estimation. Procedia Computer Science, 124, 461–469. doi:10.1016/j.procs.2017.12.178.

[43]Parwita, I. M. M., Sarno, R., & Puspaningrum, A. (2017). Optimization of COCOMO II coefficients using Cuckoo optimization algorithm to improve the accuracy of effort estimation. 2017 11th International Conference on Information & Communication Technology and System (ICTS). doi:10.1109/icts.2017.8265653.

Hybrid Fuzzy C-Means Clustering Segmentation of Tumor Region from MRI Images

Abstract - In medical diagnosis, brain tumor identification and evaluation is important. The proposed study focuses on segmenting anomalies in MR DICOM axial brain slices, since that format has the advantage of preserving broad metadata. The axial parts suppose a symmetric line of symmetry (LOS) is the left and the right portion of the brain. A semi-automatic device has been built in the sense of a DICOM analysis to mine both usual and abnormal structures from each MR slice. In this paper we suggest an algorithm for identification of human brain tumors in magnetic imaging (MRI) that contains a Kernel Fuzzy C Means (KFCM). This procedure mainly removes the color, texture and place characteristics of each pixel by choosing the required color space depending on the grey image strength. Secondly, updated membership will be calculated using a fuzzy C-means (FCM) Algorithm in the cluster centroid to cluster data points and, final, the kernel FCM clustering algorithm is used to detect the tumors location by updating its membership function obtained on the basis of the various features of the tumor image, including Contrast, Energy, Homogeneity and Correlation. The simulation results demonstrate that the algorithm proposed is able to better detect abnormal and normal tissue with less grey strength detachment in the human brain.

Keywords: DICOM processing, MR brain segmentation, fuzzy clustering, feature extraction, gray level intensity, Region Description.

1. INTRODUCTION

Digital image processing (DIP) is an emerging field of biological science such as tumor identification, recognition, disease detection, and essential parts of the human body. The automated detection of tumors is an important aspect of the medical science [1]. The human body consists of multiple cell types of which the brain (also the body's transformer) performs quite a important role. The brain is the most essential part of our nervous system [2]. In comparison, it is the central nervous human system kernel.

Increased diagnostic significance is provided by standard Digital Imaging and Medical Communications (DICOM) image formats. MR imaging systems that cooperate with DICOM conform to a particular standard on digital medical image archiving and correspondence. DICOM (.dcm) files contain metadata, such as patient analysis, settings of equipment and the mode of image attributes, scale, bit depth and measurements. A regular number of tags is grouped into the DICOM header object. These are grouped into categories including image pixels, the image plane, MR/CT image and patient information [3,4]. These are listed as patient information. Based on data elements across each class, the size of this header varies. For instance: the picture plane system takes several key parameters including slice location, slice position and pixel distance. The spatial relation between the slices is determined from these parameters. DICOM enables the development of private tags that identify open data elements in the created application. Different imagery modes store DICOM digital files, and have higher amounts of metadata than other formats. DICOM offers harmonization, which thoroughly analyses the research patient and is compatible with several trade tools [5].

The abnormal growth of the tissue or central spinal cord can interrupt the brain's proper function. Brain cancer death rates for the USA are 12,674 per year, 1106 per month, 254 per week, and 34 per day, according to the National Cancer Institute Statistics (NCIS) survey. It indicates that the detection of brain tumors at advanced stages is very important to saving lives [6]. Though, very high speed and precision can be used for tumor detection processes. This can only be achieved by the use of MR images and suspect regions are extracted from complicated medical images through MR image segmentation. The typical flow diagram of MR image segmentation is shown in figure1. Experts manually perform the detection of a brain tumor. However, some problems remain, as they take a lot of time and various experts can vary greatly in their segmentation of the MR image [7]. In addition the outcome of the detecting of tumors can differ by the same physician under various conditions, with segmentation outcomes varying with the luminosity and contrast of the screen. For these

purposes it is increasingly important to automatically detect brain tumors. Automatic brain tumor detection may enhance tumor survival risk. There is no specific procedure for brain tumor detection in the medical community [8].

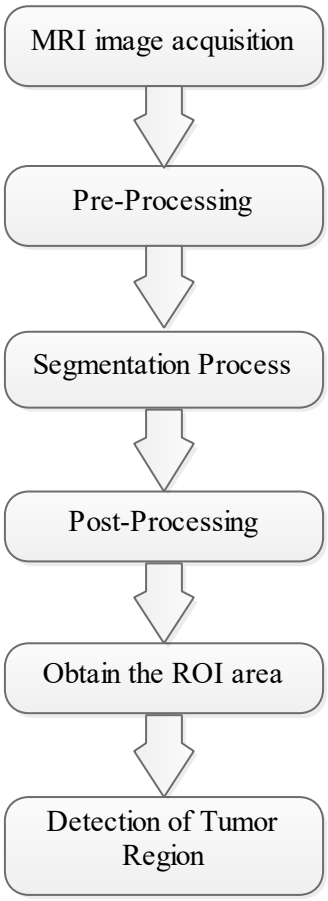


Figure 1. Typical flow diagram of MRI segmentation

Detection of brain tumors ensures that not only the infected area of the brain is known but also the shape, scale, border and location of the tumor. For imaging the brain, various imaging methods like the MRI, computed tomography (CT), positron emission tomography (PET) etc. are used. The MRI scan or CT scan will more commonly assess the brain tumor pathology. Though, the CT scan might produce radiation which is dangerous to the human body, while the anatomic tissue formation of the brain is specifically visualized by MRI [9]. MRI is a tool for the development of accurate photographs of organ and tissues by means of a magnetic field and a radio wave. Researchers are actively scrutinizing and allowing MR images incredibly complex, so that pathologists can obtain a greater background in diagnosing patients [10].

This paper outlines the main contributions:

- The benefit of this research analysis is the Kernelized fuzzy (image-enhancing) clustering strategy that efficiently decreases the field of focus of the brain segments.
- We used DICOM's attributes, such as the patient in image position, pixel distance and image orientation, as key to the generation of brain structures model and volumetric study, for each brain slice in this work.
- The fuzzy clustering is used for improving the image in the identifying of the objects of concern by correctly choosing the number of clusters 'k.' Building on the silhouette metric, the necessary k is selected between 2 and 9 clusters (k).
- The proposed study is initiated with an image consistency, similarity, and statistical tests, and is initially checked on the brain MR sequence of BRATS dataset through anomaly extraction. The

overall dice scores for ten tumor segmentation experiments have shown positive findings. The method is also checked and verified against expert data on the clinical MR brain sequence.

2 LITERATURE SURVEY

Magnetic resonance imaging for detection and visualization of the internal structure of the human body is also used in the medical field. It is primarily used to find variations in the tissues in the body and are slightly greater than the CT scan. This method is also well suitable for the diagnosis of brain tumors and cancer imaging. In general, CT utilizes ionizing radiation in contrast while MRI uses the powerful magnetic field to align nuclear magnetization which results in changes to the radio frequency alignment observable by the scanner. An image based tumor identification analysis is given in this article [11,12]. One of the major causes of rising death in children and adults has been brain tumor. The technique of CT-scan is used for photographs of the injured brain area. The images of the CT scans have been seen in gray-sized pictures because the CT scan system embraces this color type and promotes the tumor's image identification. Any brain clotting that indicates damage of any type can be observed in color as dark grey. The method of extracting parameters is simply like gathering and plotting information per pixel [13]. The images from CT-Scan show that a white and brain-damaged cell tumor is seen in black color, so binary pixel sizes showing the brain-damaged cells to 0 and shows a tumor to 1, so further analyses such as MATLAB tests and charts can be conducted by the extraction process. This procedure will separate the patient with a damaged brain from the normal patient. The tumor can also be clearly detected based on the image [14]. We also used k-means, c-fuzzy clustering and segmentation techniques in this article.

Fuzzy C-means became a soft technique of clustering where each pixel will belong to two or more cluster memberships. The Description of distances from cluster centers to models identify the targets [15]. Although only image strength values are taken into consideration and no filters are not used, it offers high noise and enhanced standard of segmentation [16]. A K-means and FCM algorithm is given, in which K-means initially segmentation, FCM further segment on the image and an FCM technique detects an approximate segmented tumor by exact selection of the cluster [17]. The traditional FCM is restricted to noise sensitivity, while in K it is clustered. However, before segmentation the threshold should be fine, as the complex structure of the brain is very difficult [18].

3. PROPOSED APPROACH

This section describes an idea employed by the proposed technology. There are mainly six modules available for the proposed system: image acquisition, pre-processing, segmenting, post-processing, extraction of features, stage detection. Filtering is done on the MR image during preprocessing. In separate segmentation work, K-means, Fuzzy C, and Kernelized Fuzzy C mean algorithms are used [19]. For function extraction, thresholding is used. In this work, two different datasets of brain slices were used. First of all, the KAGGLE dataset is a real-world clinical dataset, which contains 22 brain MR DICOM slices. Secondly, to evaluate the performance of the proposed model, the ZENODO data set has been used. In addition, ten patients with approximately 200 brain parts have been acquired in the Zenodo dataset in this work. The proposed solution to the segmenting and study of MR DICOM slice is set out in this section. Initially, pre-processing of DICOM slices is feasible. Segmentation of DICOM slices preprocessed is subject to fuzzy clustering for the improvement of the image. The silhouette metric is used to select the right clusters. The enhancement by morphological procedures of derived structures. The ROI is then collected by image post-processing techniques like MCW, RG and DRLS and checked using resemblance tests the tumor is extracted. Figure 2 illustrates the complete architecture.

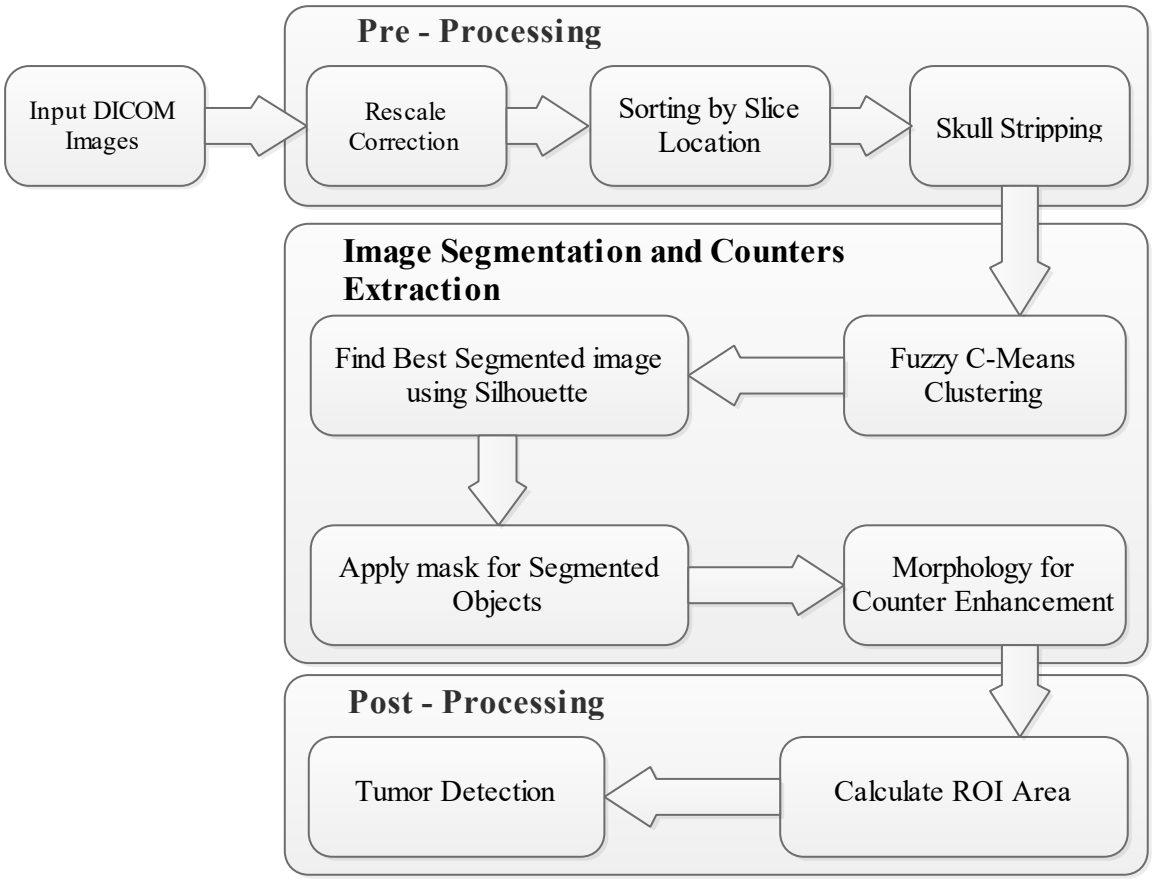


Figure 2. Proposed MRI DICOM images Tumor detection

A. MRI image acquisition

1) Data collection: MRI input images from the various open access web research libraries in DICOM, MHA and JPEG formats have been obtained.

2) Conversion of the file format: most of the images that were obtained were in DCM format; thus they were exported to a functional format. Interaction toolkit for diagnostic imaging. We have used the programmed tool for the .DICOM image.

3) Normalization of Size: images from multiple sources have been acquired; thus, variable sized images. Both pictures are scaled to 255 * 255 pixels normalized.

B. Image pre-processing

1. Remove color components for Fuzzy C-means and K-means by changing to grey mode

2. Improve image consistency with a medium filter.

3. Chart the histogram for amplitude distribution studying and analyzing of the pixels.

4. Adjustment of intensity if necessary.

5. Equalization of histogram.

6. This section formats the image as needed in additional steps

C. Segmentation

This article contrasts three major segmentation approaches focused on clusters called KFCM, Fuzzy C-Means and K-Means. When a group is clustered, unlabeled pattern sectors are separated or grouped into a few clusters such as the related patterns. These classes are referred to as clusters.

3.1. K-Means Clustering Algorithm

K-means clustering is an unsupervised algorithm for learning. This gives a very simple way to classify a certain number of clusters, i.e. a group of data like $i_1, i_2, i_3, ...i_n$ into K clusters. The key idea for this algorithm is to set K centers for each group [20]. One for each group. Randomly select the K cluster centers. Distance measurement plays a key role in the performance of this algorithm. Multiple distance measuring techniques such as Euclidean, Manhattan and Chebychev etc. are available for this algorithm. Yet it depends entirely on the type of data we are going to cluster that you choose a correct distance calculation method. We can, however, use Euclidean as a metric distance, since it is fast, robust and easy to comprehensive. The classic K-means algorithm is described step-by-step [21] as follows:

Algorithm 1:- K-means clustering algorithm

Assume that, $A = a_1, a_2, a_3, ...a_n$ be the set of data points and $C = c_1, c_2, c_3, ...c_n$ be the set of centers.

1: Define 'K' cluster number.

2: Identifying cluster 'c' spontaneously.

3: Measure the distance from the cluster centre and data point.

4: Data point for the cluster centre with a minimum distance from the cluster centre Both cluster centers. All cluster centers.

5: The following is re-calculation of the cluster centre.

$$X_n = \frac{1}{c_n} \sum_{m=1}^{c_n} A_n \tag{1}$$

Where 'c_n' is the n^{-th} cluster number of data points.

6: Recalculate the distance from each newly acquired cluster core to each data point.

7: If no reassigned data point then stop, otherwise repeat steps 3 to 6.

The distance called Euclidean is determined from each pixel to each centre of the cluster. In distance all pixels are separately compared to all cluster centers. The pixel contributes to a cluster that is shorter across both. The core is then replenished. Each pixel is then compared again for all centers [22]. This method continues until the centre converges, with a cumulative number of iterations testing the convergence. The clustering efficiency of that algorithm is optimized several times with a different initialization by repeating K-means to classify best centroids. It improves machine performance and facilitates multi-dimensional vectors. This algorithm is then programmed to decrease an objective parameter known as the squared error function:

$$X_n = \sum_{n=1}^{c} \sum_{m=1}^{c_m} \left(\| a_n - c_m \| \right)^2 \tag{2}$$

where, $\| a_n - c_m \|$ is the Euclidean distance between a_n and C_m, the c_n is the number of n^{-th} cluster data points and the c is the number of cluster centers

3.2. Fuzzy C-Means Clustering Algorithm

Bezdek introduces the FCM clustering algorithm, which is a clustering strategy in which each data pixel belongs to two or more clusters. A much closer the data to the cluster centre is to the particular cluster hub [23, 24]. The more its membership is. The objective function of the fluid-specific P cluster centroids Q is to be reduced.

$$S_i(P,Q)=\sum_{n=1}^{N}\sum_{m=1}^{C}(p_{mn})^i\left(\|a_n-c_m\|\right)^2;\ 1\le i\le\infty \tag{3}$$

Where $A=a_1,a_2,a_3,.....a_n$ is an Equation (3) $X\times N$ data matrix and i is any real number that is greater than 1. X, N and C represent the dimension of each ai vector, the number of vectors (image pixel numbers), and the cluster number. $P_{mn}\in P(X\times N\times C)$ is called vector an n-th cluster membership, which is consistent with $P_{mn}\in[0\ \ 1]$ and $\sum_{P_{mn}}=1,\ n=1,2,3,.....,N.$ The membership function is as follows:

$$P_{mn}=\sum_{x=1}^{c}\left(\frac{\left(\|a_n-c_m\|\right)^{\frac{2}{i-1}}}{\left(\|a_n-c_x\|\right)^{\frac{2}{i-1}}}\right) \tag{4}$$

Where $C=c_1,c_2,c_3,.....c_n$ is the matrix $A\times C$. The n-th cluster function centre is now determined as follows:

$$C_i=\frac{\sum_{n=1}^{N}(P_{mn})^i\times m}{\sum_{n=1}^{N}P_{mn}^{\ i}} \tag{5}$$

Where i is any individual number that exceeds 1, the level of fuzziness of the $d^2(a_n,c_m)$ is controlled. It is a similitude calculation between the two and is defined:

$$d^2(a_n,c_m)=\|a_n-c_m\|^2 \tag{6}$$

In this case, $\|.\|$ can be defined as a straight Euclid or as a widely agreed distance like Mahalanobis. The MR image pixel intensity $X=k$ is indicated in the vector A. With the continuous P and C update, the FCM algorithm can optimize $S_i(P, Q)$, repeatedly until $\left\|(P_{mn})^{(k)}-(P_{mn})^{(k+1)}\right\|\le e$ $e=0$ to 1, where k is the number of iterations. The classical foggy C-means progressively clustering algorithm is seen as follows:

Algorithm 2:- Fuzzy C-means clustering algorithm

Suppose that the set of data points is $A=a_1,a_2,a_3,.....a_n$ and that the set of centers is $C=c_1,c_2,c_3,.....c_n$.

1: Fix cluster number c, $2\le c\le m$. where m = data item number. Fix, i where $1<i<\infty$. Choose the any metric μ internal product caused.

2: Initialize the fuzzy c partition $P(0)$.

3: In stage l, l = 0, 1, 2, ...,

4: Measure the Equation (4) Fuzzy Membership Function P_{mn}.

5: Then use equation (5) to calculate the fluidized centers Ci.

6: Repeat steps 2 and 3 to meet the minimum X or $\left\|(P_{mn})^{(k)}-(P_{mn})^{(k+1)}\right\|\le e$ value.

3.3. Proposed Kernelized fuzzy C-means method (KFCM)

The inner product algorithm can be implemented indirectly in space F. This trick can be used for clustering kernelized fuzzy C-means algorithms. The clustering centre, i.e. the clustering centers in functional space, is a common ground of these algorithms to represent the combined δ(a_n,) sum of both [25, 26]. The objective function of the kernelized FCM algorithm is as follows:

$$X_n = \sum_{n=1}^{c} \sum_{m=1}^{N} P_{mn}^{\ i} \left(\left\| d(a_n) - d(c_m) \right\| \right)^2 \tag{7}$$

where δ is a nonlinear, implicit map as previously defined. Here δ (cm) is no longer represented as a linear cumulative value among all δ (a_n) as such dual expression, but is still considered in the original space as a mapped point (image) of c_m, and with the kernel substitution trick we have:

$$\begin{aligned} \left\| d(a_n) - d(c_m) \right\|^2 &= \left(d(a_n) - d(c_m) \right)^T \left(d(a_n) - d(c_m) \right) \\ &= d^T(a_n) d(a_n) - d^T(a_n) d(c_m) \\ &- d(a_n) d^T(c_m) + d^T(c_m) d(c_m) \\ &= K(a_n, a_n) - 2K(a_n, c_m) + K(c_m, c_m) \end{aligned}$$

In GRBF kernel $K(a,c) = \exp\left(- \left\| x - c \right\|^2 / s^2 \right)$, $K(a_n, a_n) = 1$, $K(c_m, c_m) = 1$, $d^T(a_n) d(c_m) - d(a_n) d^T(c_m)$ from equation (3) and (5) we get

$$X_i = 2 \sum_{n=1}^{C} \sum_{m=1}^{N} P_{mn}^{\ i} (1 - K(a_n, c_m)) \tag{8}$$

The purpose of this paper is to evaluate the Kernel Fuzzy C-means (KFCM) validity criterion for MRI data sets. The GRBF kernel provides better segmentation findings in noise-corrupted images of simulated MRs than the dependent polynomial algorithms [27] was shown. We only look for the best index for the stable kernelized fuzzy C-mean clustering on the GRBF kernel. The objective function Xi in Equation (8) is identical to the FCM algorithm under P limitation can be reduced.

$$P_{mn} = \frac{\sum_{k=1}^{c} \left(1 - K(a_n, c_k) \right)^{1/(i-1)}}{\left(1 - K(a_n, c_k) \right)^{1/(i-1)}} \tag{9}$$

The centre of the cluster c_m of:

$$cm = \frac{\sum_{m=1}^{N} P_{mn}^{\ i} \left(K(a_n, c_m) a_n \right)}{\sum_{m=1}^{N} P_{mn}^{\ i} K(a_n, c_m)} \tag{10}$$

Algorithm 3 Kernelized Fuzzy C-means clustering algorithm.

Stage 1. For any positive constant, patch c, t_{max}, m >1 and ε > 0.

Stage 2: $P_{mn}^{\ 0}, C, x$ membership initialization.

Stage 3: In the case of $t = 1, 2, ..., t_{\max}$.

(a) Upgrade the whole prototype C_n^t with equation (10).

(b) Upgrading all memberships P_{mn}^t with equation (9).

(c) Evaluate $E^t = \max_{m,n} \left| P_{mn}^t - P_{mn}^{t-1} \right|$ to quit if $E^t \pounds e$;

Finish

3.4 Improvement Morphological Operations

Masking of images can be used to denote the front, background or possible background. Contour-masking is important for further study and distinguishes the objects from the original images. The outliers such as air may be excluded from the existing brain parts. The fuzzy method of clustering reveals the best segmented clusters. They shape a binary mask that overlays the individual parts to obtain the corresponding contour intensities. In order to ensure similarity, mutual information (M) is determined between a contour mask and the respective slice [28]. The weighted (Hi) addition to the initial slice (IS_i) of the contour mask (Ci) is:

$$H_i = \frac{1}{C_i} e^{-\frac{M_i - M_{\min}}{M_{\max} - M_{\min}}} \tag{11}$$

Where Mi represents Ci and OSi mutual information. For the overall Ci, Mmax, Mmin is the most and least mutual information. The obtained binary mask of the selected parts which be blurred due to noise and texture after the extraction of the structures from the clustering process. A kind of contrast improvement strategy, mathematical morphology supports selectively enhancing the small diagnostic contour properties, overlaid on such a composite background. Dilation applies pixels to the slice contour edges. The number of additional pixels integrated in the mask image depends on its form and scale. The method of dilation is carried out by:

$$C \oplus SE = \left\{ z \mid (\hat{SE})_z \cap C^1 \neq j \right\} \tag{12}$$

where C is a binary mask pixel, SE is the structuring aspect that is first reflected as SE c, and then z is a reflected object. The whole process enhances the binary mask to prevent missing pixels across all directions, specifically at contour borders. Erosion is also carried out by equation 13,

$$C \ominus SE = \left\{ \left\{ z \mid (SE)_z \subseteq C \right\} \right\} \tag{13}$$

The SE moves by specifying that z is limited to C. Erosion eliminate pixels and sharpens the boundaries of the object. Depending on the height of SE, the amount of pixels removed. Erosion strips the natural and irregular connected contours, which help to remove ROI efficiently at the post-processing levels.

3.5 Quantizing and Validating Tumors

High solidity ventricles and tumor area are used in the removed artifacts. The image-post-processing techniques are used to collect ROI from the brain structures to reduce the ROI from the removed objects. The tumor size is quantified by area and perimeter after determination of irregular areas. In order to determine the structural intersection of the Ground Truth (GT) including its clinical slices with the extracted ROI, validity metrics are used [29]. The segmentation procedure's output is validated by similarities like Dice, Jaccard, False Positive Rate (FPR), and False Negatives Rate (FNRs). The statistical propagation of these steps is:

$$Jaccard\left(I_{gt}, I_{ROI}\right) = \left(I_{gt} \cap I_{ROI}\right) / \left(I_{gt} \cup I_{ROI}\right) \tag{14}$$

$$Dice\left(I_{gt}, I_{ROI}\right) = 2\left(I_{gt} \cap I_{ROI}\right) / \left(I_{gt} \cup I_{ROI}\right) \tag{15}$$

$$FPR\left(I_{gt}, I_{ROI}\right) = \left(I_{gt} / I_{ROI}\right) / \left(I_{gt} \cup I_{ROI}\right) \tag{16}$$

Where, *Igt* gives a segmented picture with the suggested technique to the ground truth and *IROI*.

3.6 Feature Extraction

It is still the procedure of getting more details about the shape, texture, color and contrast of an image. Texture interpretation is essentially an essential parameter of human visual perception and the framework of machine learning. It is being used mainly by choosing main features to maximize the accuracy of the diagnostic system. Haralick et al.[30] implemented Gray Level Cooccurrence Matrix

(GLCM) and texture function as one of the most frequently used image analysis applications. This procedure takes two steps to remove medical images from the functionalities. The first step is the GLCM calculation and the second is the estimation of the texture characteristics based on the GLCM. Highly dynamic nature, the retrieval of the related functionality is important throughout the brain MR images of diversified tissues like WM, GM and CSF. The identification and various stages of tumours (tumour stage) and therapeutic response evaluations could enhance textual outcomes and analyses. For some of the helpful features, the statistics form is given below

(1) Contrast (C_{on}): In contrast, a pixel and its neighbor are measured in intensity over the image. Contrast is 0 for an image that is constant and defined as

$$Con = \sum_{i=0}^{x-1} \sum_{j=0}^{y-1} (i - j)^2 f(i,j) \tag{17}$$

(2) Correlation (Cor): The correlation attribute defines the spatial dependency between the pixels. For a perfectly positive or negative picture, the correlation is 1 or -1. Correlation is NaN for a constant image.

$$cor = \frac{\sum_{i=0}^{x-1} \sum_{j=0}^{y-1} (ij) f(i,j) - M_i M_j}{s_i s_j} \tag{18}$$

where M_i and s_i are the mean and standard deviation in the horizontal spatial domain and M_j and s_j are the mean and standard deviation in the vertical spatial domain.

(3) Energy (En): The quantifiable sum of pixels pair repeatments could be described as Energy. Energy is an image similitude calculation parameter. If the Haralicks GLCM function defines energy, then it is also known as the second angular moment. Energy is 1 and is defined as a continuous image

$$En = \sqrt{\sum_{i=0}^{x-1} \sum_{j=0}^{y-1} f^2(i,j)} \tag{19}$$

(4) Homogeneity (Hg): Returns a value that calculates the similarity to the GLCM diagonal of the distribution of components. Homogeneity 1 is defined for a diagonal GLCM

$$Hg = \sqrt{\sum_{i=0}^{x-1} \sum_{j=0}^{y-1} \frac{f(i,j)}{1 + |i - j|}} \tag{20}$$

3.7 Image Quality Metrics

IQM plays an important role in designing image processing algorithms. IQM may be used to measure the output of the processed image. Quality of image is defined as a function of an image which measures the image degradation processed by comparison with an optimised image. People are typically observers and participants of most imaging technologies, so subjective process measurement of image quality is known to be an accurate method [31]. The use of the subjective approach, however, is limited in real time implementations because of its complications and difficulties in its execution. In recent years, however, quantitative approaches have been more commonly used for the measurement of image quality. In this study, we discuss multiple measurements of image quality and assess their statistical actions for seven focused indicators of the SFF system.

(i) Mean Squared Error (MSE)

MSE is a very easy and regular measure of distortion. The MSE between the referential image and the processed image $(x' \; y)$ is expressed as follows:

$$MSE = \frac{1}{xy} \sum_{i=1}^{x} \sum_{j=1}^{y} \left(A_{ij} - B_{ij} \right)^2 \quad (21)$$

where A_{ij} and B_{ij} are the reference image and processed image pixel values respectively. The MSE value calculates the difference between an image processed and the reference image. This increases the smaller value of the MSE.

2.Peak Signal to Noise Ratio (PSNR)

PSNR (Peak Signal to Noise Ratio) is one of the most widely used metric metrics for calculating the reconstruction efficiency. It represents the relationship between the maximal power of a signal and the power of corrupting sound and is typically seen in decibel size. The following can be said about PSNR:

$$PSNR\,(db) = 10 \log \frac{255^2}{MSE} \quad (22)$$

A higher PSNR value implies higher quality restoration.

3.Structural Content (SSIM)

The following is reflected in this quality metric:

$$SSIM = \frac{\sum_{i=1}^{x} \sum_{j=1}^{y} \left(A_{ij} \right)^2}{\sum_{i=1}^{x} \sum_{j=1}^{y} \left(B_{ij} \right)^2} \quad (23)$$

A higher SC value indicates that the picture is of low quality.

4.Normalized Cross Correlation (NCC)

The measure NCC demonstrates the contrast of the picture processed with the reference picture. The following was represented by NCC:

$$NCC = \sum_{i=1}^{x} \sum_{j=1}^{y} \frac{\left(A_{ij} * B_{ij} \right)}{A_{ij}^2} \quad (24)$$

5.Maximum Difference (MD)

MD produces the highest error difference between the processed and reference image. The following are described below,

$$MD = Max\left(\left| A_{ij} - B_{ij} \right| \right) \quad (25)$$

$$i = 1, 2, 3, \ldots x \;\&\; j = 1, 2, 3, \ldots y$$

The maximum variation value, the lower the image quality.

6.Normalized Absolute Error (NAE)

The NAE quality measure can be expressed as follows

$$NAE = \frac{\sum_{i=1}^{x}\sum_{j=1}^{y}\left(\left|A_{ij} - B_{ij}\right|\right)}{\sum_{i=1}^{x}\sum_{j=1}^{y}\left(A_{ij}\right)} \quad (26)$$

A higher NAE value reveals a low quality image,

7.Average Difference (AD)

The average differences for the processed and source image are made by AD. The following AD can be expressed:

$$AD = \frac{1}{xy}\sum_{i=1}^{x}\sum_{j=1}^{y}\left[A_{(i,j)} - B_{(i,j)}\right] \quad (27)$$

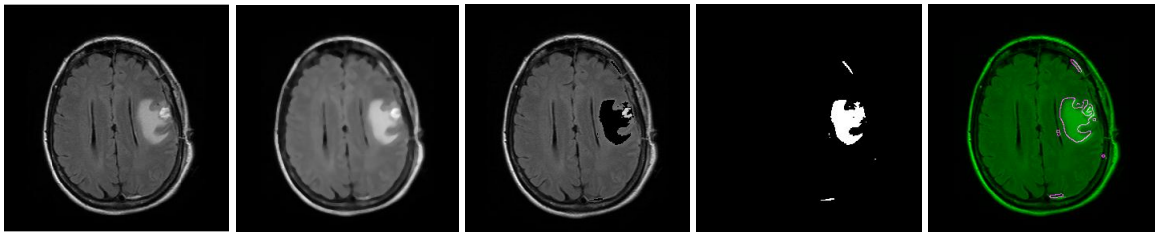
It should ideally be zero.

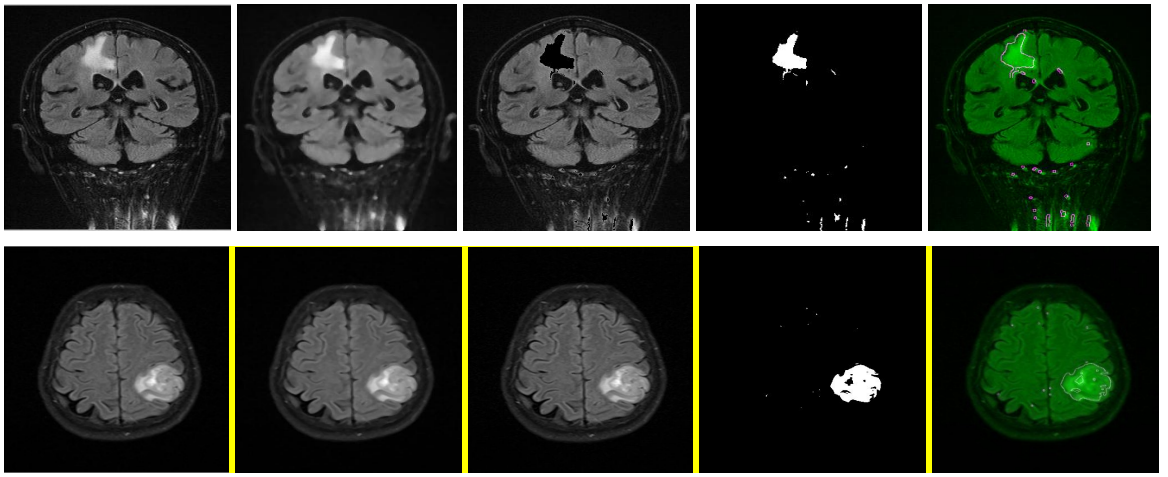
4. RESULTS AND DISCUSSION

A database of 40 MRI brain tumor images, where the first 3 images and the next database of 2 liver images are shown in Figure 3(a). The collection was compiled with numerous complex images of brain tumors. These images have been obtained from the kaggle and pre-processing in our algorithmic program for effective use. Then, by MATLAB 2014(a), we processed these original images and produced the final implementation database in figure 3. The tumors in such images have become so vital that they are too difficult to detect too easily by the ordinary people.

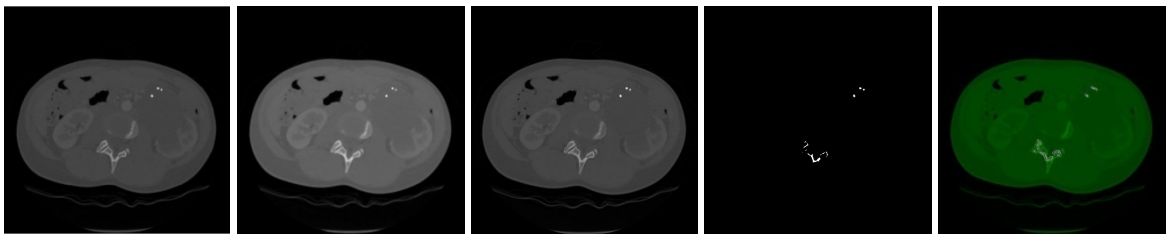
The preprocessing of MR images is very important for further processing to enhance the visual effect of an image. The captured images in the dataset are usually of such low quality that noise is removed and the image sharpened. The obtained image in the dataset is preprocessed and converted into a two-dimensional matrix, and the RGB image converted into grey image. A median filter is used to extract the noise from the background. Then the image is changed with an optimized operation, a histogram dependent operation and a histogram-based adaptive operation. Improving an image usually requires enhancing the image contrast.

Initially, multiple attributes are indirectly extracted. Each part of the brain tumor should also not be avoided, including a small part of the brain tumor. The first step is to process the input image via Gaussian and Medium filters described in Figure 3. (b). Segmentation images from the region are illustrated in Figure 3 (c). The first segmentation of the image using Karinalized FCM (KFCM) template is then shown in Figure 3(d) based on its grey intensity and color temperature, where C=4 is segmented. The tumor is then filtered again through a median filter. So the tumor has been detected and represented red line using an enhanced FCM algorithm, which depends mainly on the distance Euclidean from the cluster centre to each data centre. This might be important in order to understand the significance of this modified and integrated technique. Finally, Figure 3(e) shows the area of the tumor.





Is the above image,is CT image or MRI image (All the input images are MRI)



(a) (b) (c) (d) (e)

Figure 3. representing (a) Input MR images, (b) Pre-processed images, (c) ROI-based segmentation images, (d) FCM extracted tumor-affected area, (e) Area of the tumor-affected region.

Why all (e) images are in green color the final output images are extracted tumor area and the remaining areas are masked

Evaluation of output and segmented images the tumor extracted field estimation of trained MR images seen in table 1. The contrast between trained MR images was more evident from the observation than the tested MR images, while the homogeneity of trained MR images was lower than the tested MR images. In contrast to studied MR images, energy and homogeneity are even more found in qualified MR images. This proposed technique has been used to produce photographs of the brain tumor using mathematical structural information such as contrast, correlation, energy and homogeneity. The differences in statistical textural feature values of trained and tested brain tumors were found to be very useful in manipulating the performance of the KFCM in training and testing.

Table 1. The performance analysis of segmented images with the calculation of area

Algorithms	K-means		Fuzzy C-means		FCM Threshold		Kernel FCM	
Images	I1_Brain1	I2_liver	I1_Brain1	I2_liver	I1_Brain1	I2_liver	I1_Brain1	I2_liver
Tumour Area	0.0204	0.0001	0.0125	0.0001	0.019	0.0013	0.0187	0.0421
Number of Pixel	1619	35	992	37	1530	333	1486	10787
r_compression ratio	30.49	2004.18	50.87	1935.99	32.32	359.22	33.72	5292.71
r_bits per pixel	0.2624	0.0040	0.1573	0.0041	0.2477	0.0223	0.2372	0.0015
n_compression ratio	8.8359	14.3108	8.7455	16.9773	8.9654	17.3807	8.7465	21.4632

n_bits per pixel	0.9054	0.5590	0.9148	0.4712	0.8923	0.4603	0.9147	0.3727
Time	21.8758	70.4322	21.8758	70.4322	21.8758	70.4322	21.8758	70.4322

The MRI brain and liver image features were extracted successfully. The contrast and correlation of brain MR images were higher than those of liver MR images. However, the energy, homogeneity and correlation were higher in liver MR images than in brain MR images. Therefore, there were statistically significant differences in the variables of texture features is shown in Table 2. Thus, the heterogeneity or complexity of brain MR images was higher than those of liver MR images, while more homogenous and uniform appearances were observed for liver MR images.

Table 2. Statistical Features Analysis of segmented images

Algorithms	K-means		Fuzzy C-means		FCM Threshold		Kernel FCM	
Images	I1_Brain1	I2_liver	I1_Brain1	I2_liver	I1_Brain1	I2_liver	I1_Brain1	I2_liver
Contrast	0.1311	0.0035	0.1113	0.0035	0.1261	0.0322	0.1459	0
Correlation	0.9332	0.7428	0.9081	0.7567	0.9320	0.7474	0.9191	NaN
Energy	0.9573	0.9997	0.9730	0.9996	0.9596	0.9967	0.9602	1
Homogeneity	0.9977	0.9999	0.9980	0.9999	0.9977	0.9994	0.9974	1

The results of the applied target measures were analyzed with the image statistical metrics MSE, PSNR, NCC, AD, SSIM, MD and NAE. The results of the objective measurements measured statistically are described in table 3. In the absence of noise it contrasts and gives the ideal values of these metrics the different statistical metrics of the depth maps and all-in-center picture. Table 3 shows that Laplacian operators' statistical metrics are similar to the optimal values. With LAPD focus measurement, slightly improved outcomes are obtained. For instance, MSE is a basic and normal measure of distortion. The MSE value is the difference between the processed image and the reference image. Significantly higher MSE the outcome is much better. PSNR is also one of the most commonly used quality metrics for calculating the consistency of the reconstruction. A higher PSNR value implies a higher quality reconstruction.

Table 3. MR Image Quality Metrics

Algorithms	K-means		Fuzzy C-means		FCM Threshold		Kernel FCM	
Images	I1_Brain1	I2_liver	I1_Brain1	I2_liver	I1_Brain1	I2_liver	I1_Brain1	I2_liver
Mean Square Error	1352.1	2652.02	4593.02	2385.27	3997.7	1920.79	1322.05	134.67
Peak Signal to Noise Ratio	16.8207	13.8950	11.5098	14.3554	12.1127	15.2960	16.9183	26.8381
Normalized Cross-Correlation	0.7094	1.9301	1.2851	1.7092	1.1527	1.5914	0.8558	0.9756
Average Difference	2.5837	-32.97	-19.478	-22.026	-14.826	-18.48	2.4658	0.1020
Structural Content	0.5480	0.5833	0.5368	0.5945	0.6964	0.7367	0.9894	0.9932
Maximum Difference	227	229	247	229	247	229	227	104
Normalized Absolute Error	0.4032	1.0619	1.2032	0.9852	1.1155	0.8852	0.4310	0.0923

5. Conclusions

This paper gives a higher result than traditional schemes of the proposed KFCM algorithm. This algorithm is often found to be optimal relative to standard systems. When the proposed KFCM algorithm detects human brain tumors, SSIMs are 35.1% and 34.89% above the FCM threshold, FCM and K-mean algorithms respectively. Furthermore, the PSNR of the KFCM algorithm is 11.04% more and the MSE among all the other algorithms is slightly smaller. Moreover, the tumor region of the proposed KFCM algorithm has been decreased by 3 percent and statistics for extracting energy and homogeneity nearly achieved by the value of one (1). However, with the proposed KFCM algorithm the time taken to diagnose the brain tumor is very limited than traditional algorithms.

References

[1] Lei, T., Jia, X., Zhang, Y., He, L., Meng, H., & Nandi, A. K. (2018). Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering. IEEE Transactions on Fuzzy Systems, 26(5), 3027-3041.

[2]Stefanatos, R., & Sanz, A. (2018). The role of mitochondrial ROS in the aging brain. FEBS letters, 592(5), 743-758.

[3] Kanniappan, S., Samiayya, D., Vincent PM, D. R., Srinivasan, K., Jayakody, D. N. K., Reina, D. G., & Inoue, A. (2020). An Efficient Hybrid Fuzzy-Clustering Driven 3D-Modeling of Magnetic Resonance Imagery for Enhanced Brain Tumor Diagnosis. Electronics, 9(3), 475.

[4] Czajkowska, J., Kawa, J., Czajkowski, Z., & Grzegorzek, M. (2013, June). 4-D fuzzy connectedness-based medical image segmentation technique. In Proceedings of the 20th International Conference Mixed Design of Integrated Circuits and Systems-MIXDES 2013 (pp. 519-524). IEEE.

[5] Godinho, T. M., Lebre, R., Silva, L. B., & Costa, C. (2017). An efficient architecture to support digital pathology in standard medical imaging repositories. Journal of biomedical informatics, 71, 190-197.

[6] Alam, M. S., Rahman, M. M., Hossain, M. A., Islam, M. K., Ahmed, K. M., Ahmed, K. T., ... & Miah, M. S. (2019). Automatic human brain tumor detection in MRI image using template-based K means and improved fuzzy C means clustering algorithm. Big Data and Cognitive Computing, 3(2), 27.

[7] Joseph, R. P., Singh, C. S., & Manikandan, M. (2014). Brain tumor MRI image segmentation and detection in image processing. International Journal of Research in Engineering and Technology, 3(1), 1-5.

[8] Vishnuvarthanan, G., Rajasekaran, M. P., Subbaraj, P., & Vishnuvarthanan, A. (2016). An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images. Applied Soft Computing, 38, 190-212.

[9] Angulakshmi, M., & Lakshmi Priya, G. G. (2017). Automated brain tumour segmentation techniques—a review. International Journal of Imaging Systems and Technology, 27(1), 66-77.

[10] Al-Moteri, M. O., Symmons, M., Plummer, V., & Cooper, S. (2017). Eye tracking to investigate cue processing in medical decision-making: A scoping review. Computers in Human Behavior, 66, 52-66.

[11] Hunnur, S. S., Raut, A., & Kulkarni, S. (2017, July). Implementation of image processing for detection of brain tumors. In 2017 International Conference on Computing Methodologies and Communication (ICCMC) (pp. 717-722). IEEE.

[12] Chanchlani, A., Chaudhari, M., Shewale, B., & Jha, A. (2017). Tumor detection in brain MRI using clustering and segmentation algorithm. IJARIIE-ISSN (O)-2395-4396, 3.

[13] Koundal, D., Kadyan, V., Dutta, P., Anand, V., Aggarwal, S., & Gupta, S. (2020). Computational techniques in biomedical image analysis: overview. Advances in Computational Techniques for Biomedical Image Analysis, 3-31.

[14] Vilela, P., & Rowley, H. A. (2017). Brain ischemia: CT and MRI techniques in acute ischemic stroke. European journal of radiology, 96, 162-172.

[15] Teichgraeber, H., & Brandt, A. R. (2019). Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. Applied energy, 239, 1283-1293.

[16] Lei, T., Jia, X., Zhang, Y., He, L., Meng, H., & Nandi, A. K. (2018). Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering. IEEE Transactions on Fuzzy Systems, 26(5), 3027-3041.
[17] Nongmeikapam, K., Kumar, W. K., & Singh, A. D. (2017). Fast and automatically adjustable GRBF kernel based fuzzy C-means for cluster-wise coloured feature extraction and segmentation of MR images. IET image processing, 12(4), 513-524.
[18] Alagarsamy, S., Kamatchi, K., Govindaraj, V., Zhang, Y. D., & Thiyagarajan, A. (2019). Multi-channeled MR brain image segmentation: A new automated approach combining BAT and clustering technique for better identification of heterogeneous tumors. Biocybernetics and Biomedical Engineering, 39(4), 1005-1035.
[19] Sheela, C. J. J., & Suganthi, G. (2020). Morphological edge detection and brain tumor segmentation in Magnetic Resonance (MR) images based on region growing and performance evaluation of modified Fuzzy C-Means (FCM) algorithm. Multimedia Tools and Applications, 1-14.
[20] Khanmohammadi, S., Adibeig, N., & Shanehbandy, S. (2017). An improved overlapping k-means clustering method for medical applications. Expert Systems with Applications, 67, 12-18.
[21] Salem, S. B., Naouali, S., & Chtourou, Z. (2018). A fast and effective partitional clustering algorithm for large categorical datasets using a k-means based approach. Computers & Electrical Engineering, 68, 463-483.
[22] De Brabandere, B., Neven, D., & Van Gool, L. (2017). Semantic instance segmentation with a discriminative loss function. arXiv preprint arXiv:1708.02551.
[23] Mishro, P. K., Agrawal, S., Panda, R., & Abraham, A. (2020). A novel type-2 fuzzy C-means clustering for brain MR image segmentation. IEEE Transactions on Cybernetics.
[24] Guo, L., Chen, L., Chen, C. P., & Zhou, J. (2018). Integrating guided filter into fuzzy clustering for noisy image segmentation. Digital Signal Processing, 83, 235-248.
[25] Farahani, F. V., Ahmadi, A., & Zarandi, M. H. F. (2018). Hybrid intelligent approach for diagnosis of the lung nodule from CT images using spatial kernelized fuzzy c-means and ensemble learning. Mathematics and Computers in Simulation, 149, 48-68.
[26] Hu, G., & Du, Z. (2019). Adaptive kernel-based fuzzy c-means clustering with spatial constraints for image segmentation. International Journal of Pattern Recognition and Artificial Intelligence, 33(01), 1954003.
[27] Pham, T. X., Siarry, P., & Oulhadj, H. (2019). A multi-objective optimization approach for brain MRI segmentation using fuzzy entropy clustering and region-based active contour methods. Magnetic resonance imaging, 61, 41-65.
[28] Eresen, A., Birch, S. M., Alic, L., Griffin, J. F., Kornegay, J. N., & Ji, J. X. (2018). New similarity metric for registration of MRI to histology: Golden retriever muscular dystrophy imaging. IEEE Transactions on Biomedical Engineering, 66(5), 1222-1230.
[29] Khalifa, F., Soliman, A., Elmaghraby, A., Gimel'farb, G., & El-Baz, A. (2017). 3D kidney segmentation from abdominal images using spatial-appearance models. Computational and mathematical methods in medicine, 2017.
[30] Alharan, A. F., Fatlawi, H. K., & Ali, N. S. (2019). A cluster-based feature selection method for image texture classification. Indonesian Journal of Electrical Engineering and Computer Science, 14(3), 1433-1442.
[31] Kalaiselvi, T., & Karthigai Selvi, S. (2018). Investigation of Image Processing Techniques in MRI Based Medical Image Analysis Methods and Validation Metrics for Brain Tumor. Current Medical Imaging, 14(4), 489-505.

Arivazhagan Thamaraiselvi, Srikrishna Subramanian*, Kaliyan Naveenkumar and Sivarajan Ganesan

Distribution System State Estimation Considering Renewable Energy Sources with Optimized PMU Placements

https://doi.org/10.1515/sample-YYYY-XXXX

Received Month DD, YYYY; revised Month DD, YYYY; accepted Month DD, YYYY

Abstract: Among renewable, wind and solar energy are the most prominent and favorable substitutes to meet mankind's future electricity requirements. Typically, these resources combined in distribution network to provide local distribution users. The amalgamation of these upgrades into the distribution network may modify the malfunction and network topologies, which may fail due to their pre-set state. Therefore, an advanced and accurate tracking device must continuously monitor topology changes, which is the Phasor Measuring Unit (PMU). In this study, a new technique based on the Sea Lion Optimization Algorithm (SLnO) is proposed to determine the finest number of PMUs and employment positions, i.e., the power structure is completely perceptible. In addition, the cost of system losses can be calculated with or without wind energy to achieve energy reserves. The energy losses before and after the connection of the wind turbine are compared in order to realize energy reserves. This comparison of energy reserves made by means of Realistic and Stochastic platforms. Comparison and proposed formulas offered in the IEEE 15, 33, 69 and 85 bus distribution networks explored to prove their effectiveness. Moreover, the proposed methodology produces trusted results than those of other methods in the literature.

Keywords: Coverage, Distribution systems, Energy reserves, PMU, Renewable energy sources, Sea Lion Algorithm, State estimation.

1 Introduction

1.1 Challenges of RES Assimilation in DSSE

The electricity grid is considered to be the critical, divergent and abundant man-made active grid. To enhance the proficiency of the electricity supply and to lower the operating costs and give users the choice of which electricity provider is needed to deregulate, restructure and decentralization of the network. The modern refurbishment of this grid presents a number of challenges;

- A key stage is the development of sophisticated substructures that can capture the dynamics of the network.
- Support multiple applications with diverse needs

As a result of these meetings, the emerging electricity industries are facing unique challenges, particularly in planning, generation and operation. Previously, SCADA systems only provided static conditions for the electrical grid. Synchrophasor technology is a success because it makes it possible to synchronize phase data in real time on the network. The PMU is part of synchrophasor tools that measures the time-stamped values of frequency (f), voltage (V) and current (I) phasors in the electrical grid. It measures up to 1 rate per primary cycle. The main features have made PMUs a powerful and widely deployed observation tool in electric networks. The second sight of the distribution network is to deliver quality of power to the users with lowest possible cost and eco-friendly. R ES assimilated in the distribution network to carry out the variable demand with decarbonizing effect. When encompassing Renewable Energy Sources in the distribution system, operation of distribution network has become more complex. Due to a variation in active power flow in a renewable system, meaning that, the wind does not always blow whenever we wanted.

First Author :Arivazhagan Thamaraiselvi, Research scholar, Annamalai university, Department of electrical engineering, annamalainagar, India, e-mail: thamaraiarivazhagan96@gmail.com
***Corresponding author: Srikrishna Subramanian,** Professor and Head, Department of Electrical Engineering, Annamalainagar, India.
e-mail: ss02551@annamalaiuniversity.ac.in
Third Author: Kaliyan Naveenkumar, Assistant professor, Department of Instrumentation and Control Engineering, Sri Manakula Vinayagar Engineering College, Pondicherry, India. e-mail: naveenk.aueee@gmail.com
Fourth Author: Sivarajan Ganesan, Associate Professor, Department of Electrical & Electronics Engineering, Govt. College of Engineering, Salem, India.
e-mail: ganeshshriraj@gmail.com

Sometimes the sun does not shine due to cloud in the sky, power quality problems like harmonics, frequency and voltage variations, maintaining stability is one of the major challenges in the integrations, storage of RES, protection of equipment, optimal placement of RES, islanding of renewable sources from the distribution system. While modelling the challenge is to understand their variability and interaction with other resources. To support the reliable integration of DER at higher levels appropriate modelling and method is necessary system protection management is a herculean task.

1.2 State of the art methods

The placement of PMUs taking into account the minimal count of PMUs necessary to keep the system visible has posed a specific problem during the past three decades. PMU places a significant role in the assessment of its state, because it has excellent merit of achieving overall system observability, security and extensive control of electrical systems. The concept of Observable Propagation Depth (OPD) excludes the depth of the bus's PMU measurements [1]. To represent the OPD of each bus, new rules for the spread of observability were proposed which are valid at the incomplete observable state.

The different metaheuristic algorithm such as Hybrid algorithm of the Genetic Algorithm (GA) and Minimum Spanning Tress (MST) [2] method, Mixed Integer Linear programming (MILP) and Non Linear Programming (NLP) [3], Binary Integer linear programming [4], Binary Cuckoo Search (BCS) method [5], Taguchi Binary Bat Algorithm [6], Exponential Binary Particle Swarm Optimization Algorithm (EBPSO) [7], Modified Greedy Algorithm [8], Modified Binary Cuckoo optimization [9], Ant Colony Algorithm (ACO)[10], Best First Search Algorithm [11] are executed for the PMU placement to ensure comprehensive monitoring, the WAMS data traffic index and the cost installation index.

Furthermore, load loss and repositioning after malfunction featured as conjunction with the relaying functions of the measuring devices in the OPP model [12]. The article [13] provides a comprehensive overview of security plans for renewable integrated power network, including distribution, transmission and microgrids. Conversely, the paper [14] presents the infiltration effect of the wind turbine at the distribution level, the proximity of the wind turbine error and the position of the wind turbine relative to the grid. In addition, A performance analysis of a wind turbine as a distributed generation unit is provided in (15). Moreover, the implications of integrating large-scale wind mill into the distribution network are detailed [16]. Considering the flexibility, both HV and LV explored the robustness of the model, particularly ZT and DT usage, voltage regulation, power flow and harmonics in the network. Moreover, the paper solves the probability restricted load flow [17] by changing the optimization, complex to consider random models of wind power generation and demand / supply of electric vehicles. Article [18-19] outlines the integration of the wind farms system with regard to the impact on voltage quality and stability of the power system.

A perfect PMU placement model for electrically controlled islands, so that the electrical network is observable under controlled island conditions and normal operating conditions [20-21]. Now a day, PMUs are essential devices in the modern distribution system, in particular, on the way to the smart grid. The issue of placement of PMUs in the distribution presented in the context of system restructuring [22-24]. Voltage Stability Index [25] to identify the most important bus for voltage drop in radial distribution systems due to active and reactive power change in the line. The Tab. 1 shows the consolidate literature survey for the proposed method.

1.3 Research hiatus and motivation

The precise economic benefits of incorporating wind power into the distribution system, such as reducing electricity losses and improving voltage profiles are investigated by many researchers. Although most of these works are related to the stochastic nature of the system, therefore it is important to note the exact estimate of yearly energy reserves stochastically through the integration of wind energy. In the proposed works, energy reserves were resolute by computing the charge of losses from system with and without wind energy.

Energy losses prior to and later than the installation of the windmill is compared with the gain in energy reserves. This makes it possible to compare energy saving achieved by means of Realistic and stochastic platforms. The stochastic load flow (SLF) was resolved by means of the Monte Carlo Simulation method. This process was performed using 24-hour wind energies data for winter, summer seasons and 24-hour load demand curves for both the seasons. Fig. 1 illustrates proposed location of the PMU with the integration of the RS into the distribution system. The red color circle depicts the PMU sites and the green circle represents the placement of wind farms in IEEE 33 distribution system.

Tab.1 Consolidated Literature Survey for Proposed Method

Authors	Year	Approaches	Objective		OPP & RES Integration
			Min OPP	RES Integration	
Guo et al.	2020	Observability Propagation Depth	✓		
Devi and Geethanjali	2020	Genetic Algorithm and Minimum Spanning Tree method	✓		
Fan	2019	Mixed Integer Linear Programming (MILP) and nonlinear programming (NLP)	✓		
Su et al.	2019	Binary Integer linear programming	✓		
Babu et al.	2018	Binary Cuckoo Search (BCS) method	✓		
Ling et al.	2018	Load loss and Reconfiguration		✓	
Basetti et al.	2017	Taguchi Binary Bat Algorithm	✓		
Maji et al.	2017	Exponential Binary Particle Swarm Optimisation Algorithm	✓		Research Gap
Zhang	2017	Modified Greedy Algorithm	✓		(No reports)
KazemiKaregar & Dalali	2016	Modified Binary Cuckoo Optimisation	✓		
Mouwafi et al.	2016	Ant Colony Algorithm	✓		
Jain et al.	2016	Impact of wind turbine penetration		✓	
Venkatesh	2014	Best First Search Algorithm	✓	✓	
Syahputra et al.	2014	Performance analysis of wind power		✓	
Shafiullah et al.	2013	Influences of integrating large-scale wind energy		✓	
Vlachogiannis	2012	Probabilistic constrained load flow		✓	
Tande	2012	OPP considering system-controlled islanding	✓		
Proposed method		Impact of RES integration in DSSE considering PMU	✓	✓	✓

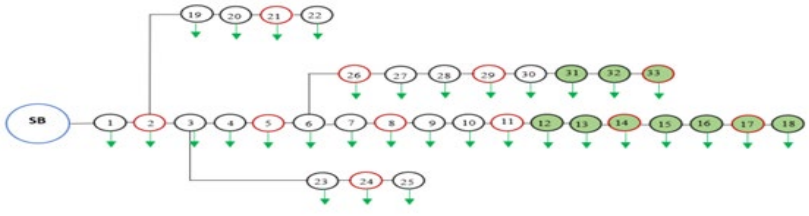


***Fig.1**. PMU placement with integration of RES in IEEE 33 bus distribution system*

1.4 Objectives

The objectives of this work are:

i. To find out the effective placement of PMU with lesser number and maximum redundancy index value. Due to the effective placement, the network has been complete observable and subject to smart grid.
ii. The integration of RES into the distribution grid enhances the voltage profile and provides a consistent power supply to utilities. This will avoid the outage of the weak buses.
iii. Annual energy reserves are compared using RLF and SLF computation. Analyze the 24-hour power reserves from radial distribution system by incorporating wind turbines into practical 24-hour wind data.

iv. Testing the relation of grid energy reserves for winter and summer. The cost of energy reserves is calculated statistically from the annual total energy losses of the systems. It compares energy reserves achieved based on Realistic and Stochastic platforms. The study was carried out using 24-hour data of wind for the winter and summer seasons in the 24-hour load curves for both the seasons.

1.5 Work flow

This paper arranged as follows: Section 2 illustrates formulation for optimal PMU and assimilation of RES. In section 3, the proposed SLnO algorithm which has been discussed for solving problem formulation. Section 4, provides simulation of PMU placement and integration of RES conferred for numerous test systems. Finally, section 5 wraps up the paper.

2 Problem formulation

2.1 Metaheuristic method based optimal PMU placement

While determining the finest number of PMUs, it is presumed that each and every PMU has adequate number of stations for measuring the bus voltage and that all the current of the incident branch of a particular bus are current phasor. For a 'n' bus power system, the placement of ideal PMU problem is stated as

$$\min \sum_{i=1}^{n} w_i * x_i \tag{1}$$

Subject to

$$f(X) \geq 1 \tag{2}$$

$$x_i = \begin{cases} 1, & if\ PMU\ is\ installed\ at\ busi \\ 0, & if\ no\ PMU\ is\ installed \end{cases} \tag{3}$$

Where,

n = number of buses in the system,
w_i = installation cost of PMU at bus I,
f(x) = vector function representing constraints and its rows correspond to one or more observable islands resemble to each critical measurement.
X= binary decision variable (0/1) that has well defined inputs.

The average overall cost per PMU (cost for procurement, installation and commissioning) ranged from $40,000 to $180,000 (U.S. Department of Energy Office 2014).

For a system which does not have standard measurement and injections, the PMU placement problem considered as finding the minimum of PMUs such that a bus must reach at least by the set of PMUs. To begin with, A binary connectivity matrix formed. The elements of matrix A defined as,

$$A_{i,j} = \begin{cases} 1, & if i = j \\ 1, & if\ i\ and\ j\ are\ connected \\ 0, & for\ all\ other\ cases \end{cases} \quad (4)$$

$$f(X) \geq 1 \quad (5)$$

$$X = [x_1, x_2, \ldots x_N]^T \quad (6)$$

$$\boldsymbol{x_i} \in \{\mathbf{0}, \mathbf{1}\} \quad (7)$$

2.2 Modelling of wind turbine

The inaccuracy of wind turbine generation at all times is primarily due to changes in wind velocity and air denseness. Since, wind speeds often vary, it is an involuntary variable in energy flow computations. This article deals with the Weibull distribution that was used to model wind velocities. It is a two-parameter function that provides a mathematical description of wind velocity [23]

$$f(v) = \frac{k}{k}\left(\frac{v}{c}\right)^{k-1} \exp\left(\frac{-v}{c}\right)^{k} \qquad 0 \leq v \geq \infty \quad (8)$$

Where,

v=wind velocity in m/s

k=shape parameter

c= scale parameter.

The output power of a windmill assumed by the following equation [19]:

$$P_w = \begin{cases} 0, & v \leq v_{ci} \\ k_1 v + k_2 v, & v_{ci} < v < v_r \\ P_r, & v_{ci} < v < v_r \\ 0, & v > v_{co} \end{cases} \quad (9)$$

Where,

P_w= wind turbine output in MW
V_{ci}= wind turbine cut in speed in m/s
V_{co}= wind turbine cut out speed in m/s
V_r= Nominal wind turbine velocity in m/s
P_r= Nominal wind turbine capacity in MW
$k_1 = P_r/(V_r - V_{ci})$ and $k_2 = -k_1 * V_{ci}$

2.3 Modeling of stochastic load

Load demands are ambiguous and that ambiguity were managed by the stochastic distribution function. The stochastic loads on each bus are incorporated into the load flow studies by showing the load in the form of randomized variables distributed with a variation of an average value. The load data of each bus is presumed to correspond to randomized variables with a Gaussian distribution [19].

$$f(P_{L,i}) = \left(\frac{1}{\sigma_{PL,i}\sqrt{2\pi}}\right) exp\left(-\frac{\left(P_{L,i} - \mu_{P_{L,i}}\right)^2}{2\sigma_{PL,i}{}^2}\right) \quad (10)$$

$$f(Q_{L,i}) = \left(\frac{1}{\sigma_{QL,i}\sqrt{2\pi}}\right) exp\left(-\frac{\left(Q_{L,i} - \mu_{Q_{L,i}}\right)^2}{2\sigma_{PL,i}{}^2}\right) \quad (11)$$

Where,

$P_{L,i}$ and $Q_{L,i}$= active and reactive load demand at bus number i
$\mu P_{L,i}$,$\mu Q_{L,i}$,$\sigma P_{L,i}$, and$\sigma Q_{L,i}$= mean and standard deviation values of load active and reactive powers respectively.

The SLF integrated all uncertainty variables into the system using the MCS [25].

2.4 Modeling of the stochastic substation voltage

The voltage of the substation can vary immediately as a function of the load changes. This prompt change in sub-station voltage is modelled with a normal distribution function [24].

$$f(v_{s/s}) = \left(\frac{1}{\sigma_{v_{s/s}}\sqrt{2\pi}}\right) exp\left(-\frac{\left(v_{s/s}-\mu_{v_{s/s}}\right)^2}{2\sigma_{v_{s/s}}{}^2}\right) \quad (12)$$

Where,

$V_{\underline{s}}$ = nominal substation bus voltages (1.00 p.u.)

$\mu V_{\underline{s}}$, $\sigma V_{\underline{s}}$= mean and standard deviation values of substation voltage respectively.

The optimum location of the wind turbine made it possible to obtain a voltage Stability index (VSI).

2.5 Energy losses cost

According to the equation given in [22], the annual cost of energy losses (COSTEL) can be calculated as follows:

$$\text{COSTEL} = (Total_{Realpowerloss})*K_P*K_E*L_{sf}*8760) \quad (13)$$

Where,

K_P = Annual demand cost of power loss ($/kW),

K_E = Annual cost of energy loss ($/kWh),

L_{sf} = Loss factor.

It can define in terms of load factor (L_f) as:

$$L_{sf} = K_c*L_f+1-K_c*L_{f2} \quad (14)$$

The standard values used in the computations are [21]: K_c = 0.2, L_f= 0.47, K_P = 57.6923 $/kW, K_E = 0.00961538 $/kWh.

2.6 Energy reserves

Energy reserves / cost savings from energy losses are obtained by:

$$E_S = COSTEL_{woWT} - COSTEL_{WT} \quad (15)$$

Where,

$COSTEL_{woWT}$= cost of energy losses without wind turbine integration

$COSTEL_{WT}$= cost of energy losses with wind turbine integration.

2.7 Wind cost

The operating cost of the wind turbine per hour t ($Cost_{wind,t}$) can be computed as follows [17]:

$$Cost_{wind,t} = a_W U_W + b_W P_{W,t} \quad (16)$$

Where

a_W = Fixed wind energy cost

b_W= Variable wind energy cost coefficient

U_W = Coefficient indicating thewind turbine availability at time t (1- available,

0-not available)

P_W = Output capacity of the wind turbine per hour t in MW

The values of coefficients used to determine hourly wind energy costs are [17] a_W = 12.8 $, b_W = 75 $/MWh, U_W = one for all the 24 hours of study.

3 Optimization algorithm for solving the OPP and wind model

3.1 Necessity of metaheuristic optimization algorithm

Metaheuristic optimization mechanisms well known for solving optimization problems by the reproduction of physical or biological phenomena. They are more common because they are built around modest notions and are easier for implementation.

Teaching Learning Optimization (TLBO), Interior Search Algorithm (ISA), League Championship Algorithm (LCA), Harmony Search (HS) and Colliding Bodies Optimization (CBO) are the other metaheuristic algorithms that are inspired by human behavior.

Population-based metaheuristic mechanisms share popular characteristics without regard to their nature. There are two key phases of the Exploration and Exploitation research process. Operators should remain as a part of the optimizer in exploring the global research space. At this time, travels should be selected at random. When the operation phase is to take place after the exploratory phase., and its phase should explore in detail the identified research space. In other words, the operation used in the area identified by the survey phase. Any metamorphic mechanism is challenged to balance investigation and functioning due to the nature of the optimization problem.

3.2 Sea lion optimization (sLno) algorithm in ephemeral

This article proposes a novel naturally enthused metaheuristic optimization algorithm called the Sea Lion Optimization (SLnO) [30-31] algorithm to solve the PMU positioning problem model in the distribution network. The SLnO method monitors sea lions' hunting behavior in the wild. Furthermore, the sea lions' whiskers used to find prey attract them.

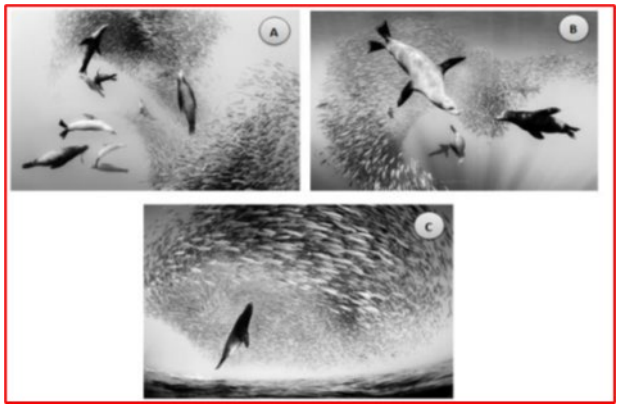


Fig. 2 *Hunting behavior of Sea Lions: (A) Chasing, Approaching, and Tracking Prey, (B) Encircling, (C) Stationary Situation and Attack*

The main stages in the hunting behaviour of sea lions were illustrated in Fig. 2 and as follows:

- Stalking and hunting with their whiskers.
- The Subgroup invites other members to join their subgroup, hunts prey and moves.
- Attack against the target (prey).

In this work proposal, these sealion hunting techniques were mathematically designed. Fig. 3 shows the employment flowchart in the OPP, considering the SLnO algorithm.

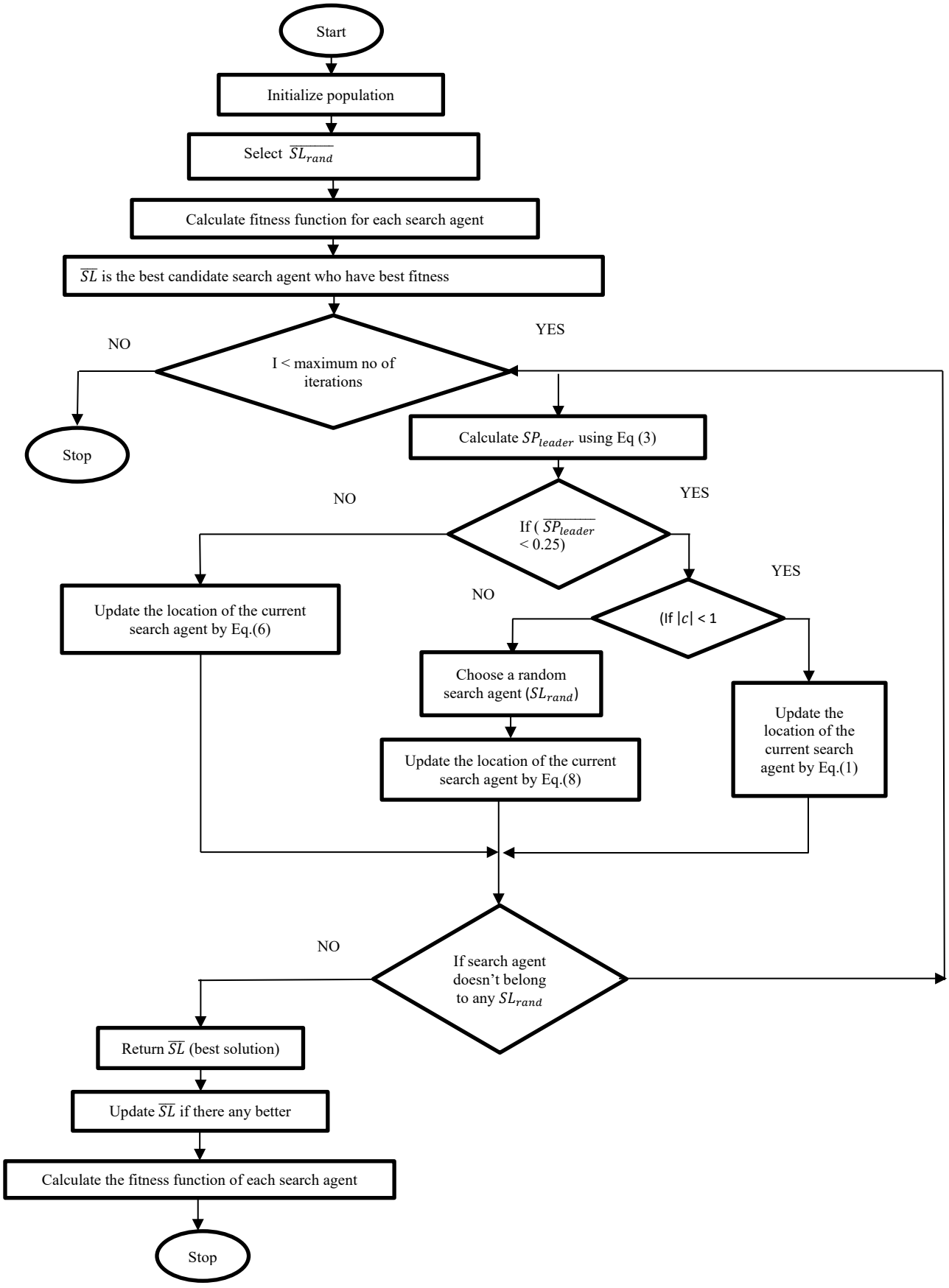


Fig.3 *Flowchart of SLno Algorithm with RES Integration with OPP Model*

4 Results and discussion

A simulation study was built on MATLAB 2015a and is performed on personal computer. It is organized in Intel® Core™ i3-5005U / 2GHz / 4GB RAM and 64-bit operating systems. The PMU deployment was formulated with the SLnO algorithm and tested on different IEEE standard test systems. To validate the effectiveness of the algorithm, OPP carried out to address the following instances.

Case 1. PMU placement in DSSE by SLnO algorithm.

Case 2. Energy losses cost with and without wind integration

Case 3. Energy loses savings for winter and summer conditions.

The various test system such as IEEE 15, 33, 69 and 85 bus distribution networks have been performed to achieve the above case 1. The IEEE 33 bus system has executed to solve the cases (2 and 3).

4.1 PMU placement in DSSE by using sLno algorithm

As part of this work, the observation assessment was performed under normal operating conditions. The optimal number of PMUs and related placements achieved using the proposed methodology for different test systems. These findings are presented in Tab. 2. The simulation results enable multiple placements with the same number of PMUs. Optimal positioning can be selected for applications such as state estimation accuracy, node monitoring and maximum coverage values. In IEEE 15 bus system, there are five PMUs are required to observe the whole system. The SLnO algorithm finds more set of locations with different coverage values. Based on the highest coverage value, the best locations optimize by the proposed algorithm. Therefore, the best locations are 2, 4, 9, 11 and 13. The highest coverage or SMRI value found as 20 for the above said optimal locations. Similarly, the optimal placements of PMU with lesser quantity and its coverage value for the IEEE 33, 69 bus systems given in Tab. 2. The PMU placement in the IEEE 85 bus system is maiden attempt and its results obtain with a minimum number of PMU, best locations and high coverage values. In addition, the total PMU channel values also obtained for the all-test systems shown in Tab. 2 that claimed as part of the uniqueness of the proposed work

Tab. 3 shows, the number of PMU, SMRI and PMU channels achieved by the proposed algorithms with existing algorithms of the distribution test systems. In case of IEEE 33 bus system, 11 PMUs are required to observe the system with the coverage value of 34. The AGA, CES, NSGA and MST methods also required to 11 PMUs to observe the system. Nevertheless, the coverage values of the existing methods very low compared to the proposed method. Therefore, the SLnO algorithm affords effective results. Even though, Greedy Algorithm and ACO algorithm achieved high coverage value such as 38 and 36 respectively. It should not take as optimal because the number of PMUs required in that method is more. Since the proposed method is reduced the number of PMU from 14 (Greedy Algorithm), 12 (ACO) to 11. Hence, the proposed method is delivering the better results against existing methods.

Tab. 2 Best PMU Locations with Maximum Coverage Obtained by SLno Algorithm

Test System	No. of PMUs	PMU Location	SMRI	Total PMU channel	Execution Time in sec
IEEE 15	5	2, 4, 9, 11, 13	20	11	0.25
IEEE 33	11	2, 5, 8, 11, 14, 17, 21, 24, 26, 29, 33	34	25	0.48
IEEE 69	23	1,3,5,8,12,15,18,21,24,27,30,33,38,41,44,48,50,52,55,58,61, 64,69	81	68	0.96
IEEE 85	32	2,4,6,8,10,12,13,15,17,19,22,24,26,27,29,32,35,37,41,45,47, 50,53,55,58,62,64,67,70,73, 81,84	102	88	1.12

Tab. 3 Comparisons of PMU Location and SMRI

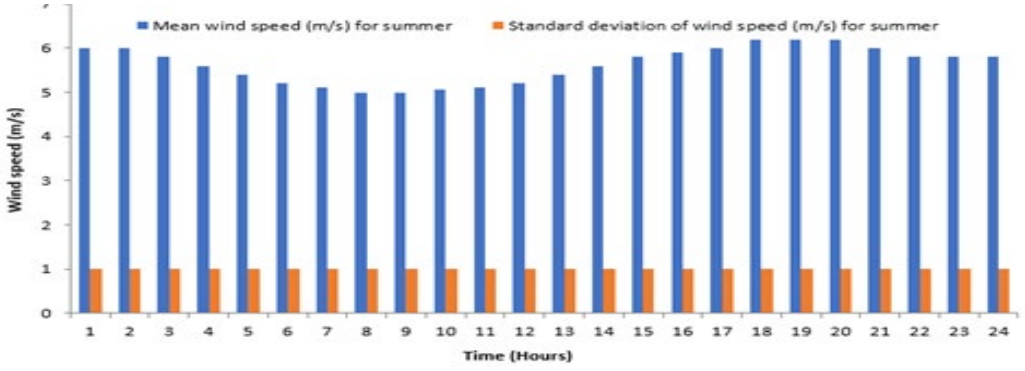
Methods	IEEE 15 Bus			IEEE 33 Bus		
	No. of PMU	SMRI	PMU channel	No. of PMU	SMRI	PMU channel
Proposed method	5	20	11	11	34	25
AGA [26]	-	-	-	11	33	-
Greedy Algorithm [27]	7	22	-	14	38	-
CES [28]	-	-	-	11	33	-
NSGA [29]	5	19	-	11	32	-
ACO [10]	6	21	-	12	36	-
MST [2]	5	19	-	11	32	-

Methods	IEEE 69 Bus			IEEE 85 Bus		
	No. of PMU	SMRI	PMU channel	No. of PMU	SMRI	PMU channel
Proposed method	23	81	68	32	102	88
AGA [26]	26	84	-	-	-	-
Greedy Algorithm [27]	27	85	-	-	-	-
CES [28]	25	82	-	-	-	-
NSGA [29]	23	80	-	-	-	-
ACO [10]	24	82	-	-	-	-
MST [2]	23	80	-	-	-	-

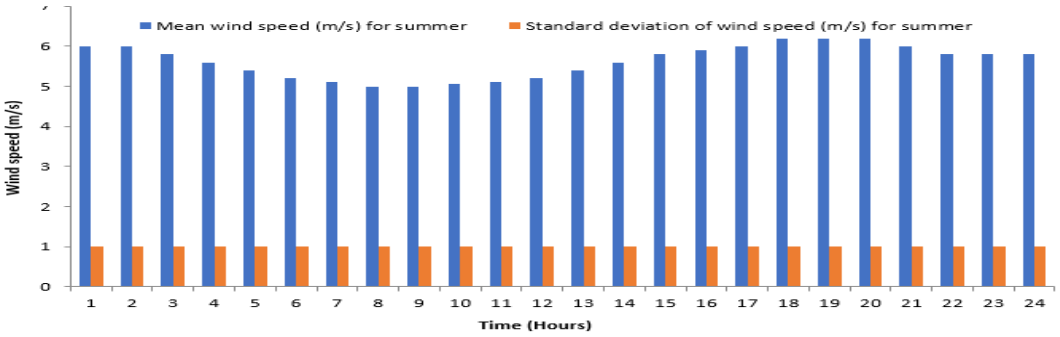
4.2 Energy losses cost with and without wind integration

The integration of the windmill into the distribution network carried out in the IEEE 33 bus system. IEEE standard radial distribution system data acquired as defined in [32-33]. The base power of IEEE 33 is 100MVA and the base voltage is 12.66KV. The cumulative active power load is 3.72 MW with a reactive power load of 2.30MVAR. Realistic load flow (RLF)for the feeder system shall be performed using forward/reverse load flow method [24]. The exact measurements taken by the PMU and n are compared with the load flow values. Then the error values were downplayed. The windmill selected for this analysis have the following characteristics: Vci = 2.5 m/s, Vr = 13 m/s, Vco = 25 m/s and Pr=0.6 MW [21].

Ten wind turbines planned to install in the test distribution grid. The 24-hour daily data were used for the two seasons of the year, winter and summer. The average wind speed, standard deviation for the 24-hour data (Cape Comorin, a southern Indian location) are obtained for January and May. Average and standardized wind speed deviations for the summer, winter seasons are monitored by the PMU indicated in Fig. 4.



(a) summer



(b) winter

Fig.4 *Mean and Standard Deviations of wind speed – (a) summer and (b) winter*

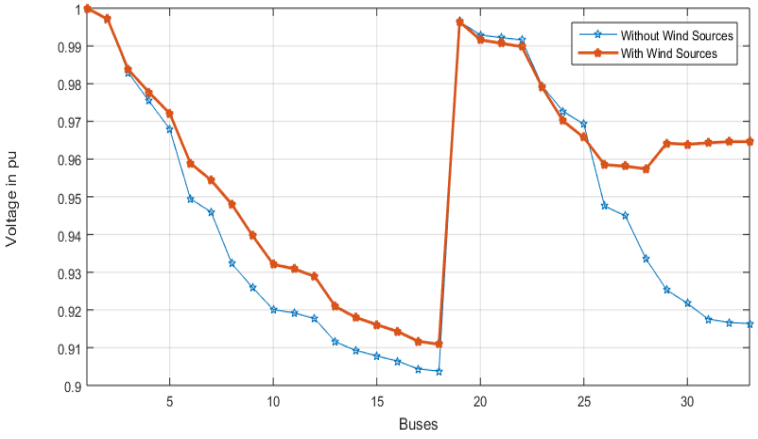
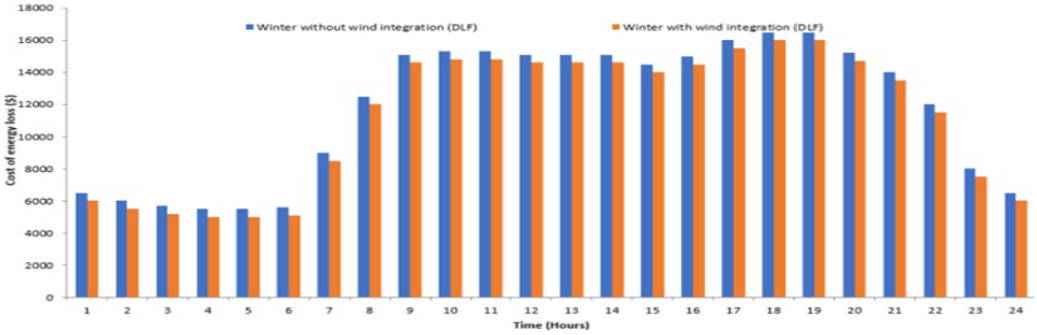
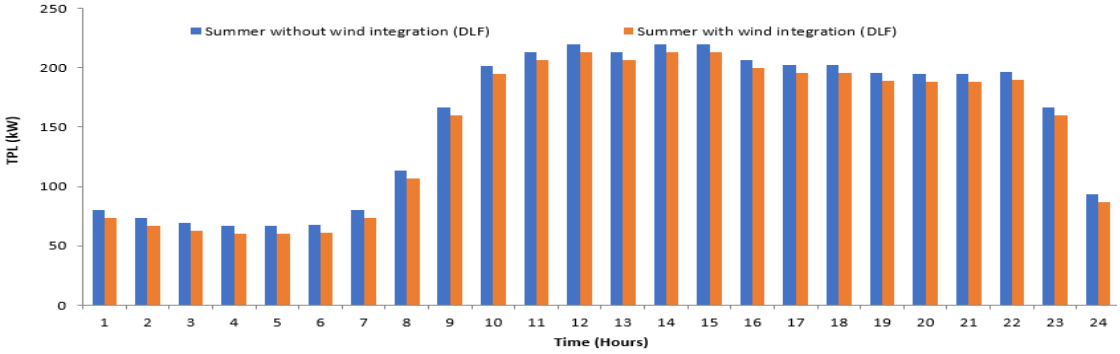


Fig. 5 *System Voltage measured by PMU*

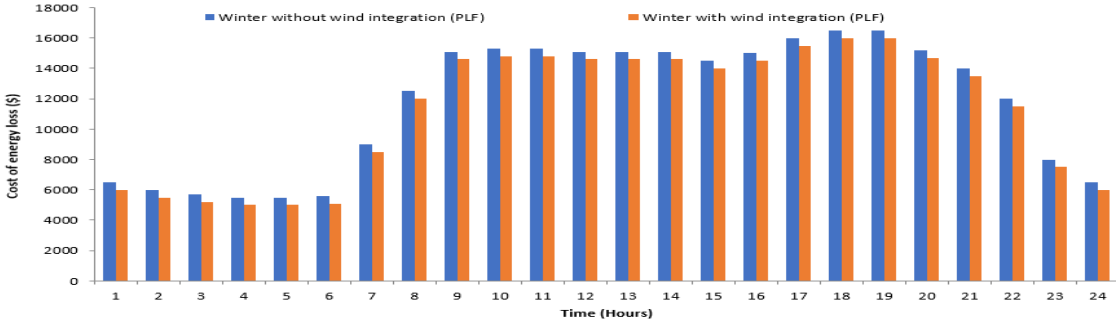
There are approximately 10 wind turbines installed in the distribution network. Initially, the forward / backward load flow was executed for the IEEE 33 bus utility grid. From load flow result, the voltage magnitude angle, real power, reactive power, losses obtained. Based on the voltage values, the integration of the wind turbine implemented. The node 18, 17, 16, 15, 14, 13, 33, 32, 31 and 12 having the minimum voltage values. Theses nodes considered as the weakest node and it may outage quickly if any fault occurs in the distribution network. Therefore, Wind turbines placed in these nodes to expand the voltage profile of the particular nodes and its whole system. Fig. 5 clearly shows the system voltage with and without wind turbine. The system voltage monitored by the PMU very precisely.



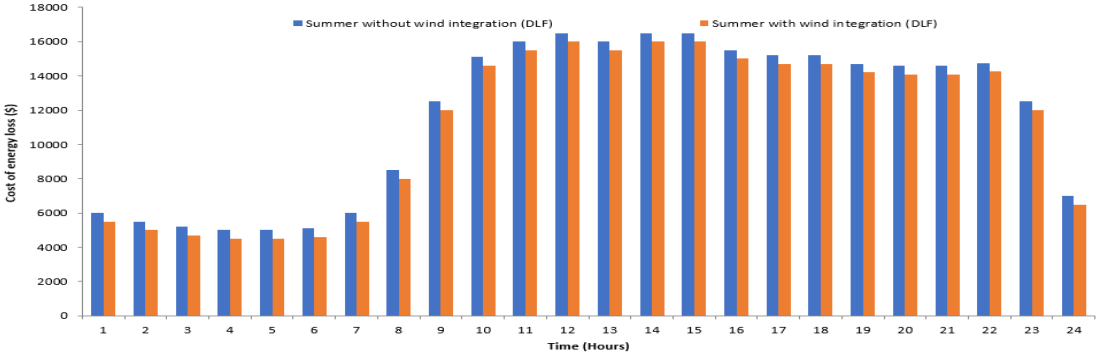
(a) Summer-RLF



(b) Winter– RLF



(c) Summer– SLF

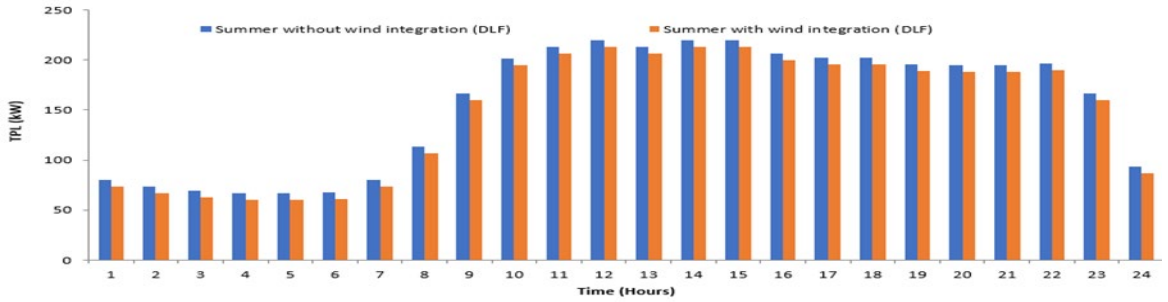


(d) Winter – SLF

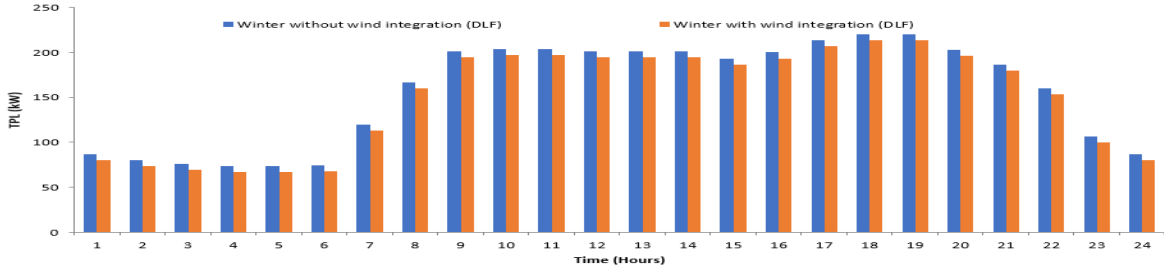
Fig. 6 *Cost of energy losses ($) of 33 bus RDS – (a) Summer – RLF, (b) Winter – RLF, (c) Summer – SLF and (d) Winter – SLF*

4.3 Energy losses saving for summer and winter conditions

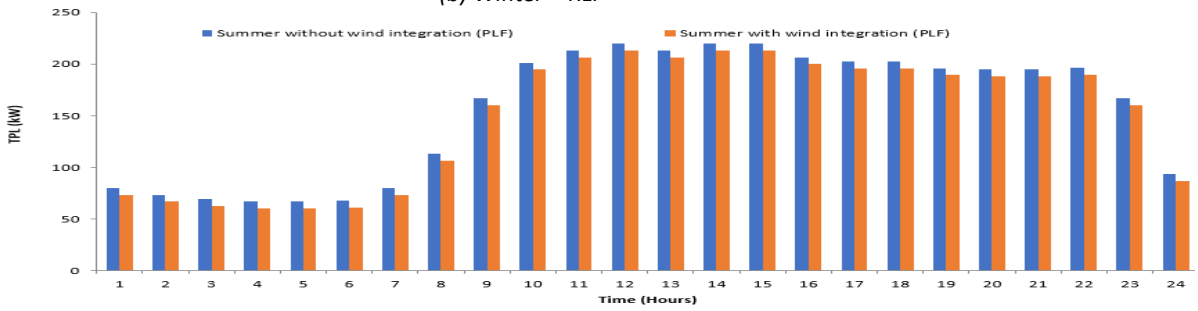
Fig.6 shows the energy losses cost in $ based on RLF and SLF estimates for the summer season and winter seasons of 33 RDS buses. Fig. 7 provides a summary of overall actual power losses or values of TPL that are derived from RLF and SLF measurements for both the seasons summer, winter of 33 RDS bus. It is clear that from Fig.6, the cost of grid energy losses decreases considerably in the midsummer and midwinter following the integration of windmill.



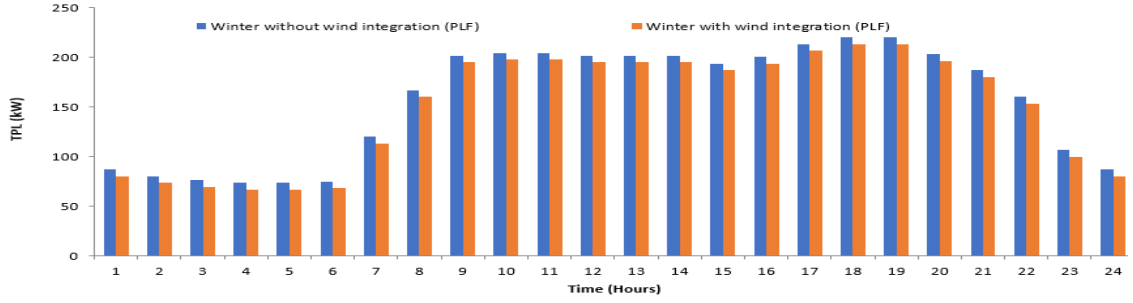
(a) Summer-RLF



(b) Winter – RLF



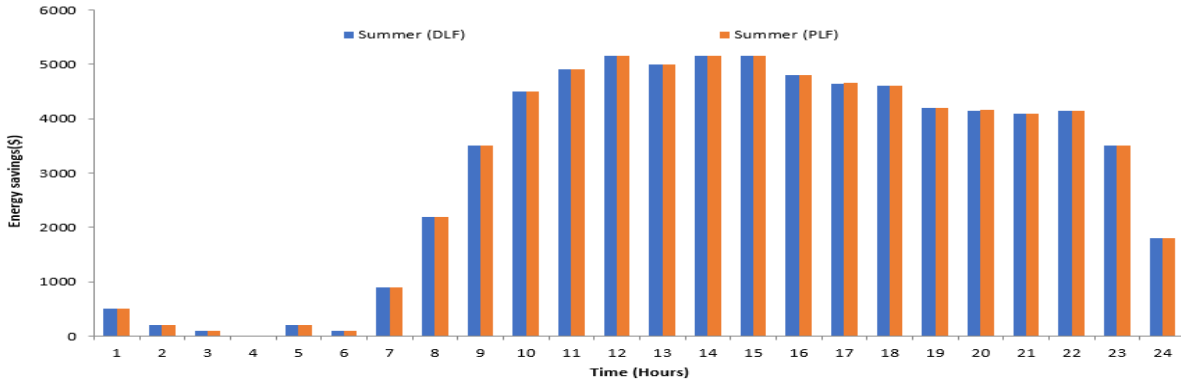
(c) Summer – SLF



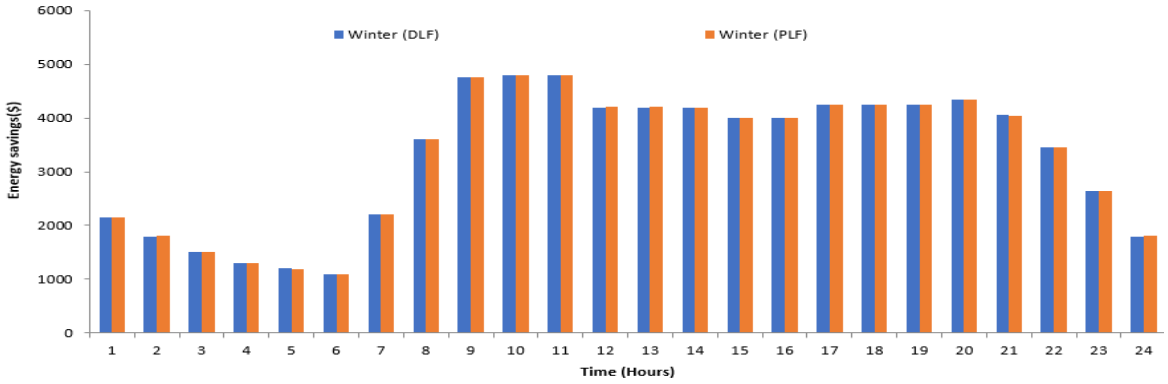
(d) Winter – SLF

Fig. 7 *Total real power loss (TPL) for the 33 bus RDS – (a) Summer – RLF, (b) Winter – RLF, (c) Summer – SLF and (d) Winter – SLF*

Fig. 7 demonstrates that the lowest cost of energy lost per hour5, during the summer season without the incorporation of wind turbines. The charge for the minimum hourly cost of energy losses is $4,877 from the RLF and $4,963 from the SLF. Both actual and reactive power losses values are 61.83 kW and 41.86 kVAR respectively.The minimum bus voltage for that hour is 0.9481p.u. The minimum load for summer energy losses following wind turbine deployment shall also be measured per hour 5. The minimum cost per hour for energy losses is $4,719 RLF and $4,727 SLF. The actual and associated power losses are $57.77 and $45.03 respectively.



(a) Summer



(b) Winter

Fig.8 *Energy loss savings for summer and winter seasons of 33 bus RDS*

Seasonal energy conservation is depicted in Fig.8. It indicates a substantial difference in energy conservation for the test method during the winter and summer seasons. Peak energy reserves for the summer season were reached at 3 pm. It is $5,209 based on SLF estimates and $5,209 based on RLF calculations. For the same time, winter electricity reserves total $4,013.38. The maximum energy reserves for the 10-winter hour reached this value of $4,772 based on SLF calculations and $4,779 based on RLF calculations. Energy reserves for the summer amount to $4,523 per hour. A minimum electricity savings of $25 was achieved at the time of 4 for the summer. Minimal energy reserves of the winter at the time of 6 and benefit for it are $1,065. Minimum energy reserves of the winter outweigh the minimum energy reserves for the summer season. Fig.8 reveals that energy reserves were greater in the winter from 1 to 10 hours and in the summer from 11-19 hours and from 21-24 hours. Disparity of wind speed and load patterns influences net conservation of energy.

5 Conclusion

In recent years, improvements to distribution networks have exacerbated the distributed power line problem. This increases the number of PMUs to be monitored and replaces the data with an increased sampling rate. This paper presents a helpful justification for the optimisation of PMU placement technique through sea lion optimisation. This technique provides perspectives such as optimum locations, maximum visualisation, minimum PMU and total PMU channel. In addition, this article describes the effect of RES integration on the distribution network. Energy reserves of the generation were resolute by computing Realistic and stochastic methods of energy loss before and after the integration of the wind turbine as regards the achievement of energy reserves. A demonstration of the proposed technique was conducted for standardised distribution systems. Based on the results, SLnO is proved as a promising tool for solving such a kind of constrained objective optimisation problems. As a result, it can conclude that SLnO solution to the precise problem is the best one so far.

5.1 Policy implications

For SG applications, many solutions needed to overwhelm the challenges of PMU-placement restrictions, topology observability and maximum coverage, as they provide planning engineers with additional options in selecting the solution that best encounters their requirements.

Acknowledgements

The research work titled as “Distribution System State Estimation Considering Renewable Energy Sources with Optimized PMU Placements” by A. Thamaraiselvi and Dr. S. Subramanian was supported **by Tamil Nadu State Council for Higher Education (TANSCHE)**

References

[1] Guo XC, Liao CS, Chu CC. Enhanced optimal PMU placements with limited observability propagations. IEEE Access. 2020 Jan 17; 8:22515-24.

[2] Devi MM, Geethanjali M. Hybrid of Genetic Algorithm and Minimum Spanning Tree method for optimal PMU placements. Measurement. 2020 Mar 15; 154:107476.

[3] Almunif A, Fan L. Optimal PMU placement for modeling power grid observability with mathematical programming methods. International Transactions on Electrical Energy Systems. 2020 Feb;30(2): e12182.

[4] Su H, Wang C, Li P, Liu Z, Yu L, Wu J. Optimal placement of phasor measurement unit in distribution networks considering the changes in topology. Applied Energy. 2019 Sep 15; 250:313-22.

[5] Babu NP, Babu PS, Sivasarma DV. Binary cuckoo search based optimal PMU placement scheme for united Indian grid-A case study. International Journal of Engineering, Science and Technology. 2018 Jun 6;10(2):10-24.

[6] Basetti V, Chandel AK. Optimal PMU placement for power system observability using Taguchi binary bat algorithm. Measurement. 2017 Jan 1; 95:8-20.

[7] Maji TK, Acharjee P. Multiple solutions of optimal PMU placement using exponential binary PSO algorithm for smart grid applications. IEEE Transactions on Industry Applications. 2017 Feb 8;53(3):2550-9.

[8] Tran VK, Zhang HS. Optimal PMU placement using modified greedy algorithm. Journal of Control, Automation and Electrical Systems. 2018 Feb;29(1):99-109.

[9] Dalali M, Karegar HK. Optimal PMU placement for full observability of the power network with maximum redundancy using modified binary cuckoo optimisation algorithm. IET Generation, Transmission & Distribution. 2016 Aug 15;10(11):2817-24.

[10] Mouwafi MT, El-Sehiemy RA, Abou El-Ela AA, Kinawy AM. Optimal placement of phasor measurement units with minimum availability of measuring channels in smart power systems. Electric Power Systems Research. 2016 Dec 1; 141:421-31.

[11] Venkatesh T, Jain T. Optimal PMU placement using best first search algorithm with pruning. In2014 Eighteenth National Power Systems Conference (NPSC) 2014 Dec 18 (pp. 1-5). IEEE.

[12] Wu Z, Du X, Gu W, Liu Y, Ling P, Liu J, Fang C. Optimal PMU placement considering load loss and relaying in distribution networks. IEEE Access. 2018 May 29; 6:33645-53.

[13] Telukunta V, Pradhan J, Agrawal A, Singh M, Srivani SG. Protection challenges under bulk penetration of renewable energy resources in power systems: A review. CSEE journal of power and energy systems. 2017 Dec 25;3(4):365-79.

[14] Jain R, Zhang Y, Hodge BM. Investigating the impact of wind turbines on distribution system stability. In2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) 2016 Sep 6 (pp. 1-5). IEEE.

[15] Syahputra R, Robandi I, Ashari M. Performance analysis of wind turbine as a distributed generation unit in distribution system. International Journal of Computer Science & Information Technology. 2014 Jun 1;6(3):39.

[16] Shafiullah GM, Oo AM, Ali AB, Stojcevski A. Influences of wind energy integration into the distribution network. Journal of Wind Energy. 2013 Nov 11;2013.

[17] Vlachogiannis JG. Probabilistic constrained load flow considering integration of wind power generation and electric vehicles. IEEE Transactions on Power Systems. 2009 Sep 25;24(4):1808-17.

[18] Tande JO. Grid integration of wind farms. Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology. 2003 Jul;6(3):281-95.

[19] Dugan RC, Mcdermott TE. Distributed generation. IEEE industry applications magazine. 2002 Aug 7;8(2):19-25.

[20] Ramachandran R, Karthick S. Optimal PMU Placement for Tamil Nadu Grid Under Controlled Islanding Environment. Procedia Technology. 2015 Jan 1; 21:240-7.

[21] Abdelsalam HA, Abdelaziz AY, Mukherjee V. Optimal PMU placement in a distribution network considering network reconfiguration. In2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014] 2014 Mar 20 (pp. 191-196). IEEE.

[22] Eminoglu U, Hocaoglu MH. A voltage stability index for radial distribution networks. In2007 42nd International Universities Power Engineering Conference 2007 Sep 4 (pp. 408-413). IEEE.

[23] Dukpa A, Duggal I, Venkatesh B, Chang L. Optimal participation and risk mitigation of wind generators in an electricity market. IET renewable power generation. 2010 Mar 1;4(2):165-75.

[24] El-Fergany AA. Involvement of cost savings and voltage stability indices in optimal capacitor allocation in radial distribution networks using artificial bee colony algorithm. International Journal of Electrical Power & Energy Systems. 2014 Nov 1; 62:608-16.
[25] Ranjan R, DAS. Simple and efficient computer algorithm to solve radial distribution networks. Electric power components and systems. 2003 Jan 1;31(1):95-107.
[26]
[27] Zhou X, Sun H, Zhang C, Dai Q. Optimal placement of PMUs using adaptive genetic algorithm considering measurement redundancy. International Journal of Reliability, Quality and Safety Engineering. 2016 Jun 1;23(03):1640001.
[28] Abdelsalam H A, Abdelaziz A Y, Osama R A, Salem R H. Impact of distribution system reconfiguration on optimal placement of phasor measurement units. In Proceedings of the 2014 Clemson University Power Systems Conference.11–14 March 2014: 1–6.
[29] Chen X, Chen T, Tseng KJ, Sun Y, Amaratunga G. Customized optimal μPMU placement method for distribution networks. In2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC) 2016 Oct 25 :135-140.
[30] Kong X, Wang Y, Yuan X, Yu L. Multi objective for PMU placement in compressed distribution network considering cost and accuracy of state estimation. Applied Sciences. 2019 Jan;9(7):1515.
[31] Vijayan J, Radharamanant T, Sridharant R. Sea lion with enhanced exploration phase for optimization of polynomial fitness with SEM in lean technology. Evolutionary Intelligence. 2020 Mar 2:1-8.
[32] Masadeh R, Mahafzah BA, Sharieh A. Sea lion optimization algorithm. Sea. 2019;10(5): 388–395.
[33] Das D, Kothari DP, Kalam A. Simple and efficient method for load flow solution of radial distribution networks. International Journal of Electrical Power & Energy Systems. 1995 Oct 1;17(5):335-46.
[34] IEEE PES AMPS DSAS Test Feeder Working Group. (2019, 02 September). Online Available: http://sites.ieee.org/pes-testfeeders/.

AWOA-RP SCHEME FOR IMPROVED COMMUNICATION OF VANET USING NETWORK PROTOCOLS

Abstract-

VANET offers a secure and safe transportation over the vehicle to vehicle communication meant for the application of industrial wireless network. The vehicular communication in turn enhances the easy convenience between source and destination nodes. The BS and RSU too facilitate knowledgeable for attaining the field of vehicular data communications. Once activating the communication of VANET, this in turn offers new energy efficiency in the V2V connection and entire vehicular nodes to be enhancing by considering minor time. Consequently, the efficient routing path, lower delay, uncongested path, minimum consumption of energy, less execution time, and classification accuracy will create the VANET dependent competent communication model. The AODV protocol performance needs to enhance the stability parameters, less cost, and energy in relation to the communication of VANET. In this approach, the Adaptive Whale Optimization Algorithm-Routing Path (AWOA-RP) is presented for enhancing the AODV protocol in relation to the energy-efficient, delay enabled dynamic routing and path evaluation, stability and there is some indispensable variation that allows the preeminent routing path devoid of less delay, any link breakage for employing the stability of route. For congested path identification, the projected Clustering-dependent Congested Path Detection Algorithm (CCPDA) is employed. The CCPDA offers the transmission delay, PDR over the speed of vehicle through the trust-dependent sufficient residual energy levels. For the detection of malicious node in VANET, the presented Reputation value-based Prominence Algorithm (REPA) along with the identification of the malicious node should be considered. The PA improvements offer the reduced End-to-End delay, the minimum overhead enhanced PDR, routing problem in the network. The performance of VANET communication on the whole is analyzed to have improved performance of the malicious node identification. As a result, this approach offers secured and safe wireless communication appropriate for the industrial wireless networks.

Key Words: Adaptive Whale Optimization Algorithm-Routing Path, Vehicular Adhoc network, safety, Clustering-dependent Congested Path Detection Algorithm, security, Reputation value-based Prominence Algorithm.

I. INTRODUCTION

In recent years, safety and secure communication between the vehicles as necessary. The efficient routing path among the vehicles without any traffic, delay of time, dynamic routing between the source and destination nodes. Wireless Networks will play major role in the industry 4.0. The vehicular fields provide wireless data communications and also capable of communicating with one another. i.e., roadside unit infrastructure with the respective trusted authorities. VANET has split into two concerned methods that are Vehicle to Vehicle (V2V), and Roadside to Vehicle (R2V) communication. The primitive technology provides multiple applications like traffic accidents, vehicle to vehicle communication, vehicle traffic levels, drivers, passengers, and pedestrians discussed in [1]. The tremendous characteristic features, such as active topology changes and high vehicle mobility. The high mobility is the essential factor that is describing the VANETs from other wireless networks and Ad hoc networks. The VANET area in the vehicle node density is not identical which exhibits spatiotemporal variation in [2]. The ancient trajectories of data may be solving reduced road congestion in VANET. The fuel wastage to be increases, capital losses, and fatal with the prediction of congestion control discussed in [3]. VANET is helping to generate every device with the information it necessary to route traffic timely. The routing protocols in VANET based on various internal and external factors like as vehicle node mobility, signals, obstacles, and road topology to enable the scalable characteristics in VANET shown in [4]. The challenges in VANET shown in figure 1.

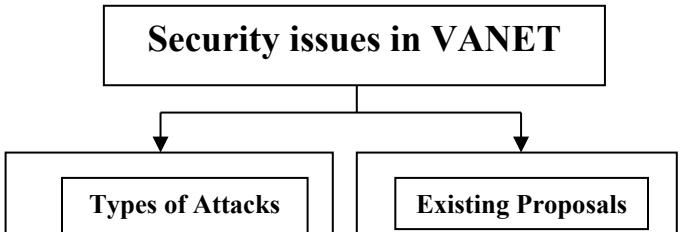


Figure 1 Challenging factors in VANET

AODV routing will not hold all routes at all times.Whenever packet is ready to sent automatically the route process will start{Yahiabadi, 2019 #855}. The goal is to maximize the vehicle and passenger safety and comfort also. In VANET, every vehicle capable of connecting to VANET and make a network with a wide range of applications. If any vehicle to be dropped out of the range of transmission, the network also dropped in [5]. The data packet in the system provides efficient routing protocols. The routing algorithm gives less delay, overhead routing, dynamic PDR and End-to-End delay in the network. The vehicles in the network are working the better routing algorithms to information delivering to the neighboring vehicle position at any time of instance. If the positioning system fails, this provides high cost on routing, high bandwidth usage, and routing overhead shown in (Shafi & Ratnam (2019)). The vehicle mobility, speed, converged the huge calculation overheads to vehicles to detriments of the performance of communication when they exploit the largest part of eminent degree algorithm to decide the CH selection. By consider the mobility of vehicle parameters, to utilizes each vehicle's hand-off capacity and the handoff time defer according to the packet loss rate (Valayapalayam Kittusamy, Elhoseny & Kathiresan 2019). The various applications in VANETs function the multiple connections such as Vehicle to Infrastructure transmission and Vehicle to Vehicle transmission. The information services as the prerequisites to efficacy and long life multi-hop routing design approach. The VANET's characteristics give the dynamic technique, high mobility, road layout limitations, easy access wireless system, and severe channel conditions are the demerits to design a routing scheme (Chen, Liu, Qiu, Wu & Ren (2019)). VANET has some inability characteristics to attain efficient routing among the destination node of vehicles and the source node of vehicles. The problem infers that

1. Inefficient routing path, which makes the traffic and delays time to improve the AODV protocol for energy, delay time, and cost enabled vehicle to vehicle communication.
2. Congested paths infer the inefficient of residual energy
3. The malicious node attacks may lead to the routing overhead problem, End-to-End delay in the network.
4. To implement a secured framework applicable for industrial wireless network (IWN).

Here, to overcome all these inabilities and satisfies the efficient vehicle to vehicle connection, Vehicle to Infrastructure connection as well as enable the safety and secured connection in VANET.

II. LITERATURE SURVEY

[6] proposed the VANET based routing protocols define the location of the vehicles, but never rely on route entries of predestination that solves the challenging factors such as inaccuracy, broadcasting overheads. **[7]** proposed multi-agent reinforcement learning through the trajectory rejection method to regularize the traffic congestion also minimizes the travel time, and travel distance. The road network-based congestion index provides the best traffic efficiency with the adaptive learning method. **[8]** proposed a VANET-Cloud layer to enable network performance and traffic management. The traffic services performed through the data exchange mechanism help of the fuzzy aggregation technique. The VANET cloud layer allows enhanced traffic safety. **[9]** proposed the Multimodal Nomad Algorithm associative of communication scheme to assist the multiple drones of VAN. The enhanced performance produces a specific space for the global application, and its performance like hop count, throughput, and PDR values are better. **[10]** proposed the safety and security based fifth-generation network to outperforms the applications and business models along with VANETs and SDR. The new drive of security threats and the vulnerabilities function through the different entities and the various architectural

components. The critical security services and enhanced Intelligent Transport Systems (ITS) that incorporate the proposed SDVN techniques. [11] stated the Vehicular Cloud (VC) computing environments enable through V2V and V2I to optimize the traffic intersections along with the signal control strategy. The effect of the traffic patterns functions the probability-based traffic signal control parameters. **[12]** proposed the statistical Network Tomography (NT) based VANETomo for congestion identification. The Connected Vehicle (CV) technology-based Dedicated Short Range Communication evolves vehicular communications. To improve the accurate QoS through the Statistical Network Tomography that concludes the transmission delays over the network nodes. **[13]** proposed the framework of distributed collaborative detection to tracking the collection of data. The malicious nodes detection evolves through dynamic behavior analysis technology. The Stochastic Petri Net used for security-based detection of higher detection rates. **[14]** proposed the vehicular traffic management provides the vehicular traffic congestion paths through the congestion routes and weighted graphs. To identify the congested routes without traffic jams to improve the traffic distance by using Re-RouTE. Re-RouTE used to minimize traffic jams and enhanced travel time, distance, speed, and the traffic flow of the road networks. [15] proposed that cloud computing enabled VANET in IoT environments to construct control on traffic and data processing systems through a cloud-based vehicular network. The machine learning and data mining techniques how performs the heterogeneous cloud platforms. **[16]** proposed the DisTraC for traffic congestion control parameters based on less communication overhead problems to be satisfied with the minimum travel time and all other external infrastructures. Nowadays many people's eye on Industrial wireless networks (IWNs) than academic communities and other domains. **[17]** proposed the optimization algorithm for Dynamic Interior Point Method congestion path identification through driver rerouting. DIPM outperforms in the knowledgeable real-time deployment of user-based optimal approaches. **[18]** proposed the refined congestion control problem may elongate through the traffic-based, resource-based classical techniques. The optimization of traditional methods used for soft computing-related approaches. [19] proposed the Base Station (BS) and Road Side Unit (RSU) over the vehicular medium with the extensive adaptation. To mitigate security-based threats and privacy to be evolved. **[20]** proposed the Vehicle-to-Vehicle (V2V) communication models provides efficient access, reliability, and availability through the concept of spatial intelligence. V2V enables the Intelligent Transport Systems (ITS) for the exchange of information and intelligent decisions on roads.

III. PROPOSED METHODOLOGY

To establish the improvements in AODV protocol with **t**he clustering enabled route selection process for optimal routing path (VANET Communication) to decrease overhead and enhance the route stability in the proposed Adaptive Whale Optimization Algorithm (AWOA-RP) to be considered. The algorithm enables the better performance of packet delivery ration and the link stability while improvements on energy, delay time, least congestion, malicious node estimation are the parameters that get to consider. The schematic diagram of the proposed AWOA-RP methodology shown in figure 2. Thus, a secured communication path has been implemented with the use of this approach which is greatly applicable in industrial wireless network.

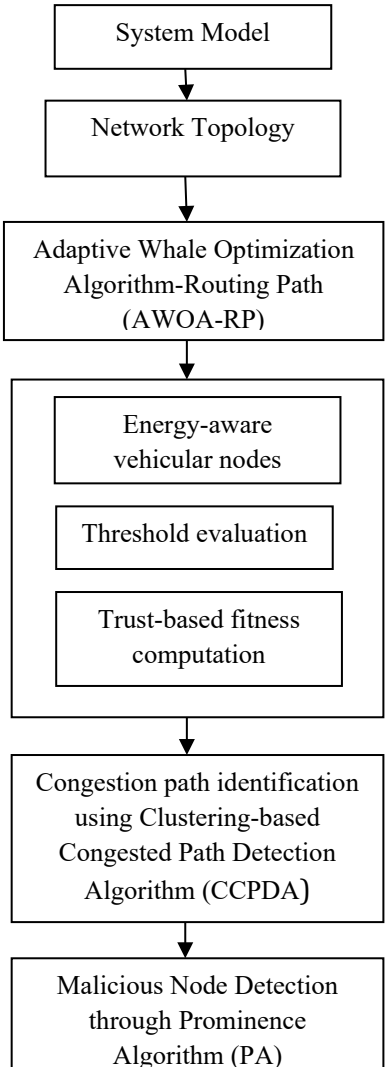


Figure 2 Schematic diagram of the proposed method

Normally, the Whale Optimization Algorithm depends on the bubble net feeding behavior of Whales.

4.1 System Model

The vehicular nodes to be initialized by considering the effective routing path among the vehicular nodes through the topology of the vehicular network for V2V, and V2I communication. Initially, to construct the clustering formation, CH selection before the routing process and the source and destination of the vehicular nodes to be initialized first and its details explain in further process.

To imagine that the vehicles are equipped with Road Side Units (RSUs), On-Board Units (OBUs), Locations and sensors. For effective vehicle communication, the RSU works as an intermediate node which will exchange the data between the vehicles. The network model is to be considered for the establishment of the road model structure with N average number of vehicle nodes.

4.2 Adaptive Whale Optimization Algorithm

The better path of selecting a route is usually to predict traffic between the source and the destination points in VANET routing. The direction of a low traffic density is favored for optimum route choice, as a high traffic density increases the flow of vehicles on the road. An active prediction method is necessary to choose the best path. Route exploration is the central process in the current implemented optimization method. The optimization algorithm finds the right route, from source to destination, which meets the entire considered multiple limitations. For optimizing multipath, an algorithm can be used to provide reliable data delivery. During the route selection, the data transfer to the destination can be postponed if the nodes are not energy efficient and overloaded because of high traffic. In VANETs, the vehicle nodes travel along with fixed paths alone because of some restriction on the roads and streets.

The Whale Optimization Algorithm (WOA) is a new optimization technique which is used to solve optimization problems. The WOA consists of three operators to simulate the search for, encircling prey, prey and bubble-net foraging behavior of humpback whales to reduce the clusters and also achieves the reduced end-to-end delays. Road Side Units (RSU) will allow the planners to deploy effective communication without any link breakage. The QoS in the network is measured using various parameters like end-end delay, energy consumption, packet loss, and throughput.

Initially, clustering the vehicles is non-linearly set into motion through the own fitness value of own and their neighbors thus imitate efficient neighbors by attaining the optimal path between the vehicles. The clustering of vehicles uses for continuous herds or each vehicle can belong to a single global flock. Clustering starts with collecting arbitrary particles to improve the cluster stability and also to reduce the cluster regeneration timing. The wireless-enabled vehicle nodes get clustering according to the encircling prey, Let the optimal path for clustering the vehicle nodes considered as,

$$\vec{V} = \left|\vec{D}.w^{better}(t) - w(t)\right| \quad (1)$$

$$w(t+1) = w^{better}(t) - \vec{B}.\vec{V} \quad (2)$$

Where, V as vehicle nodes

t as time is taken during clustering

w (t+1) as optimal path

In the routing table with no path to destination node, the route discovery phase will start when a source node likes to propel the packets to the destination node. A Route Request (RREQ) packet has broadcasts by the source node to its neighbors. The RREQ packets received by neighbors are classified into three categories: the receiver node has a path to the destination; the receiver node is the destination node and none of both. In this condition, the RREQ packet has

forward to its neighbors by the receiver node until the RREQ packet has received by the destination node or the destination node has routed to the node, and the process repeats with the reduced overhead problem.

To achieve the optimal routing path, the constant number of clusters in a vehicle nodes gets enabled for AODV protocol improvement. For the Trust-based routing path by making the cluster creation, a Trust-based rule manager never considered for the clustering process, and all vehicle nodes imagined as trust nodes. Moreover, the routing processes enable each trust-based vehicle nodes. The large differences in speed of any 2 vehicles might result in lesser link time taken and the links are break easily. Based on the previous knowledge, the proposed AODV protocol explained in detail the trust-based route calculation, link breakage among the vehicles, and the threshold assignment according to the vehicle nodes in the wireless network.

Now, the trust-based routing enables through the threshold assignment among the vehicle nodes with the optimal path estimation. The Threshold assignment depends on the residual energy mean value and overall Trust based score mean value for each vehicular nodes. By using the overall trust scores to find the mean trust-based score value for each vehicular node

$$TM_i = \frac{(\frac{TS_i + \sum_{j=1}^{n} TN_i}{n-1})}{n}, j! = i \quad (3)$$

Where,

TM_i = Trust Mean value (i_{th} node).

TS_i = Summation of trust scores

TN_j = Trust Value of the Neighbor,

n = Number of nodes in the network

Threshold value that helps for routing path calculation, i.e., Mean value >Threshold to participate The DV for vehicular nodes selection through the formula as,

$$\text{Decision value}_i = \frac{(W1*TM+W2*RE*W3*(\frac{1}{D}))}{(W1+W2+W3+} \quad (4)$$

Where TM > Threshold

RE is residual energy

D is Distance from the neighbor vehicle nodes

The threshold values assigned to the vehicle nodes and stable to enable the trust-based calculation over the neighbor nodes without any link breakage. The trust-based calculation among the vehicle nodes used to find the Trust based score for each vehicle nodes considering the two limitations.

1. The vehicle nodes are sending an ACK to neighbor vehicle nodes when it receives the packets.

2. The vehicle nodes that drop more packets.

$$TSC_{i1} = (\frac{ACK}{RP}) * 100 \quad (5)$$

Where TSC_i = The percentage of the initial trust-based score for the i_{th} node. ACK = Acknowledgements sent to the nearby vehicle nodes. RP = Total number of received packets.

$$TSC_{i2} = 100 - ((\frac{DP}{TDP}) * 100) \quad (6)$$

Where, DP = Dropped packets, TDP = Overall dropped packets.

$$TSC_i = \frac{(TSC_{i1} + TSC_{i2})}{2} \quad (7)$$

After trust-based score evaluation, the Cluster Head Selection will do for the effective nest preparation over the nodes. The selection of Cluster Head algorithm to calculate the high residual energy optimization and Trust-based score values, but with the reduced distance from all participated vehicular nodes for distance evaluation. The distance of the vehicle nodes as,

$$D_i = |avg\ dist\ of\ nodes\ (v1) - \max dist\ of\ nodes| \quad (8)$$

Where, D_i as the average distance of vehicle nodes

The traffic system on each vehicle reduced and the average distance helps to enable the Cluster Head (CH) Selection. The sensor nodes enable the efficient CH selection to predict the path and less-cost system by addition of distance between the each vehicle.

The sum of distance among the vehicle nodes provides the essential message for the feasible

communication that utilized to detect the distance of sensor node. The sensor node distance is,

$$S_i = \frac{Number\ of\ sensor_j}{Maximum\ sensors\ used_j} \quad (9)$$

The sensor nodes help to establish the optimal position of the vehicle using the vehicle's sensor and to enable the AODV protocol gets improved based on the least traffic occurrence and energy efficiently. The area of the best hunt used to estimate the clustering enabled CH selection with the updating of adjusting the proximate distance over the vehicle nodes. The fitness value calculated by using the formula as,

Fitness = {Max_stability of cluster} (9)

The fitness value helps to improve the channel capacity among the vehicle nodes. An Adaptive Whale Optimization is to increase the capacity of the channel with lesser time by choosing the suitable necessities of communication and radar sensing and the equation as follows,

$$CC = \max_{cc} \frac{N_c\left(T_g\right|+T)_{f*M*S_c}}{T} * \log_2(1 + \frac{\delta TP}{N_0.S_c}) \quad (10)$$

Where,

N_c as spacing of subcarrier nodes

S_c as number of subcarrier nodes

T_g as yclic prefix

f as frequency among the vehicle nodes

Here, the Whale Optimization Algorithm Encircling prey, Bubble net attacking model used to enable the shortest path among the vehicle nodes, without any traffic, improved data loss, the shortest path to be estimated.

Encircling prey

It is depicted by

$$I = \left|\vec{x}.\vec{b}(t) - \vec{b}x(t)\right| \quad (11)$$

$$b(t+1) = b^{better}(t) - H.\ x\ \vec{1} \quad (12)$$

where, t as recent CH selection

H, x as coefficient vectors

$\vec{b}(t)$ as better CH selection arrangements with respect to the position of vectors.

$x\ \vec{1} = 2x\bar{2}.\vec{s}$ and

$x = \bar{2}.r$

where x2→ randomly reduced from two to zero all through the iterations r→ is a non-linear vector in [0, 1].

Bubble net attacking model:

Updating the new Cluster Head values is completed by two steps with the help of vehicle clustering model. The steps used are Spiral updating position and Shrinking encircling mechanism.

Both the methods are used to enable the better CH selection with the accordance of vehicle nodes. The position of humpback whales seems chaotic as indicated by the position of each other for optimal shortest path estimation as given below:

$$\vec{N} = |x.c_{rand} - \vec{c}| \quad (13)$$

$$C(t+1) = c_{rand} - y1.\vec{H}\ ]\text{correlation} \quad (14)$$

This function is done by the decreasing the value linearly an y2 from zero to two; the vector k1 with random value in the ranges are fixed between −1 and 1.

The Correction Factor(CF) which stimulate the whales to move in lesser steps in the direction of the prey to explore the search space resourcefully for better and efficient shortest path selection. To find the trust-based degree of vehicle nodes with the respective shortest path equation based on the RSU system as,

$$D_{RSU\text{-}U} = \sqrt[2]{(U_{RSU} - U_i) + (V_{RSU} - P_t)^2} \quad (15)$$

The proposed algorithm considers the trust aware score values, distance, and residual energy for the unique routing path, which is optimum. The limitation of CH is that the speed and mobility < average mobility

speed of all the vehicular nodes. Finally, the vehicle communication established the efficient routing path with the improvements and modification on AODV protocol according to the vehicle's stability and the primary and essential parameters such as energy consumption, delay time reduction, and the least traffic also attained.

4.2 Congested path in VANET

The predicted congestion never specifies the actual coordinates also for the congestion time. For congested path identification, the Clustering-based Congested Path Detection Algorithm (CCPDA) to be used. This algorithm detects congestion due to traffic, location, and congestion time.

For path clustering, path algorithm enables one to cluster the path. The algorithm performs the longitudinal paths clustering, and it follows 3 steps. Firstly, it describes the features. Secondly, to select a subset of data through the factor analysis. Finally, the performance of cluster evaluation to find the cluster path and allocate each cluster in all ways. The path clustering method consists of path range, SD, Mean change/time.

The path distance calculated through the Harvesine formula such as, For any two points of vehicles, the central angle among the two vehicles as,

$$\theta = \text{hav}(\phi 1 - \phi 2) + \cos(\boldsymbol{\phi} 1) + \cos(\boldsymbol{\phi} 2)\text{hav}(\lambda 1 - \lambda 2) \quad (16)$$

Where θ referred to as Harvesine function (θ):

$$\theta = \sin^2\left(\frac{\theta}{2}\right) = \frac{1-\cos(\theta)}{2} \quad (17)$$

After the distance of each path to be found through the Havesine formula to calculate the distance and path traveling time. The time to travel the path distance and the time estimation as,

$$t = t_i - t_{i-1} \quad (18)$$

The formula for cluster speed and path for least congestion of traffic path that can be found using below as;

$$S = \frac{d}{t} \quad (19)$$

By considering each cluster paths, the average speed/cluster computed by using;

$$Cs = \frac{T_S}{n} \quad (20)$$

The congestion depends on the threshold value with its bandwidth, and residual energy If the averages speed of a cluster < the speed of the threshold, it attain less than 1 km/h and the time is important than a threshold time. The duration of stopping congestion event < the threshold duration, which is not congestion, it has a stoppage of traffic. If cluster speed < 1 km/h and the event duration > threshold duration, the vehicles to stop for while a time, then that are locked. If stoppage duration $\geq$ threshold duration, which is not a traffic stoppage or passengers to be picked, which is a congestion occurrence. The results of congested events performance conducted in terms of the positions like longitudinal and latitude of time occurrence.

4.3 Malicious node Detection

For malicious node attacks in VANET, the proposed REputation value-based Prominence Algorithm (REPA). PA helps for malicious nodes detection. The Prominence Algorithm necessarily depends on the node's prominence value. The prominence values based on the Packet Forwarding Ratio (PFR), i.e., the packets are received and the packets are ahead to the neighboring nodes. The malicious node connects; it receives the default prominent value and the performance of a specific node. The prominent value describes the perceiver node. In PA, the selection of the perceiver node based on various specifications. The selection of a perceiver node never has the best prominence value and less computation load parameters.

Malicious node "s" (perceiver node) needs to find the prominence of node "t" depends on the calculation of Packet Forwarding Ratio (PFR), and node "t".

Pseudocode for Malicious node detection:

Packet Forwarding Ratio (PFR) value (P_b) as an input
Status of the malicious node "t" as an output.

Step 1: The node "s" finds the PFR value of node "t" described as P_b:

$P_b = (u/v)$

Where u as = the data packets need to forward by 's' to 't'.

v as = the data packets forwarded by the 't' node

Step 2: Depends on the value of Packet Forwarding Ratio, it decides the node "t" of prominence value(Direct prominence)

If

i) $P_b \geq T_h$ and $P_b \leq 1$; then prominence value

$R_t^s = 1$; Else

ii) $R_t^s = 0$;

Step 3: Node "s" asked other neighbor nodes to send the prominence value of node "t":

$R_b = \sum_{i=1}^{i=n} R_t^i$

If

i) $R_t \geq n/2$; then reputation value $R_t = 1$; Else

ii) $R_t = 0$;

step 4: Node "s" computes the final prominence of node "t":

$R_f = R_t^s + R_t$

If

i) R_t^s and R_t are 0; then final prominent value $R_f = 0$;

Else

ii). $R_f = 1$;

Step 5: For final prominence R_f, value declares prominence of node "t":

If

i) $R_f = 0$, the 't' node is a malicious;

ii) To remove the "t" node

Else

iii) $R_f = 1$; the 't' node is trusty node

Step 6: The message delivers to all the other nodes about the node's status as 't'.

The Prominence Algorithm that detects the attacks in V-2-V and V-2-I communication in VANET networks. The attack detection range of the network status is improved and compare to the other existing algorithm.

V. RESULTS AND DISCUSSION

The proposed work for an efficient routing path has done through MATLAB. For malicious node based attack detection in VANET, the prominence algorithm that represents the attack detection. AWOA-RP has a trustworthy secured and optimized routing algorithm. AWOA-RP used for direct Trust to evolve the trust based on multiple nodes. AODV protocol based optimized routing process enhances the performance of routing overhead issues. The optimized AODV protocol routing algorithm provides the addition of efficient shortest path identification with the trust-based secured malicious node detection.

To calculate direct trust values and indirect trust values with the help of final prominence value. The enhanced results like Routing Overhead PDR, and End-to-End Delay by compared with the existing performance. The malicious nodes have the dropping in the packet that gets affected and PDR value is worst. To calculate the PFR with the help of the perceiver nodes, to estimate the direct as well as the indirect prominence values of the node. For better performance, the average of End-to-End Delay could be as low. When the highest path breakages occurred because of the presence of malicious nodes, so the overhead routing issues to be improved by the concern of frequent routes and its measurement. Thus, a secured communication path has been implemented with the use of this approach which is greatly applicable in industrial wireless network.

The results of malicious nodes detection in VANET shows as follows: Figure 3 represents the packet Delivery Ratio Vs Time (ms). Figure 4 describes the End-to-End delay Vs Time (ms). Figure 5 depicts the routing Overhead Vs Nodes

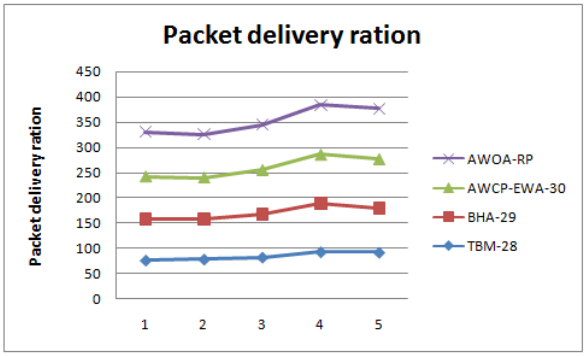


Figure 3 Packet Delivery Ratio Vs Time (ms)

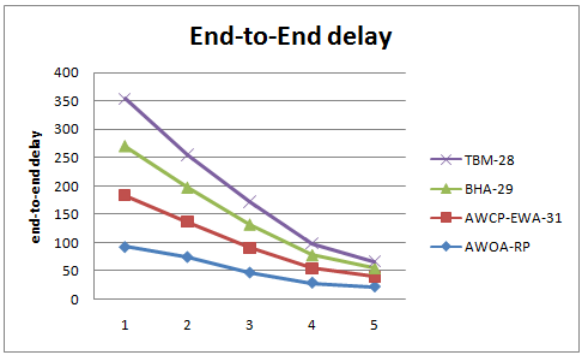


Figure 4 End-to-End delay Vs Time (ms)

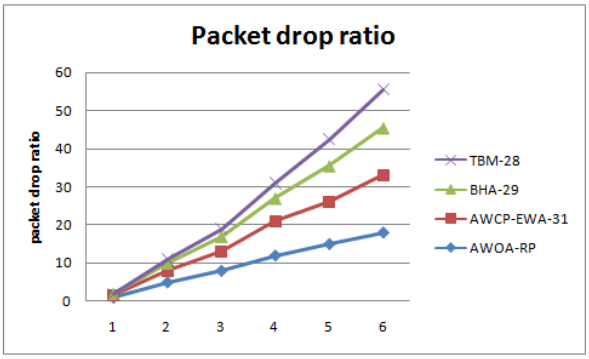


Figure 7 Packet drop ratio Vs Time (ms)

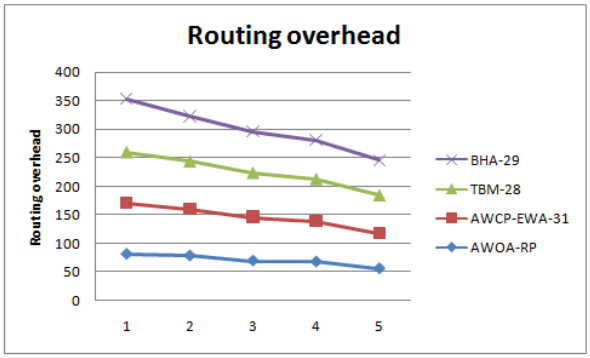


Figure 5 Routing Overhead Vs Nodes

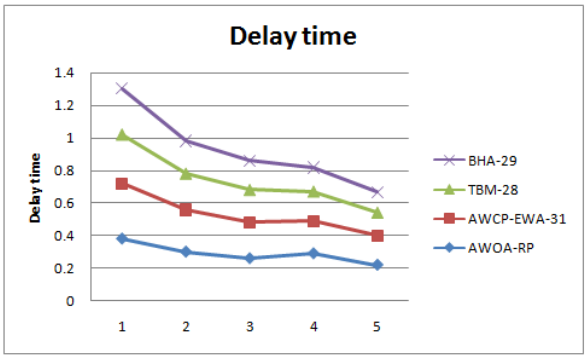


Figure 8 Detection time Vs Nodes

Figure 6 represents the lifetime duration. Figure 7 shows the packet drop ratio Vs Tie (ms). Figure 8 illustrates the detection time Vs Nodes. The reduced energy consumption shown in figure 9 with the number of nodes in the network.

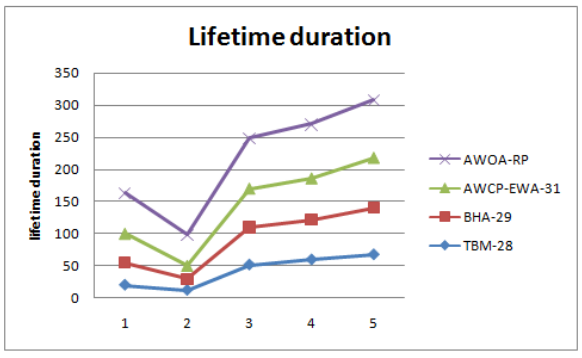


Figure 6 Lifetime duration

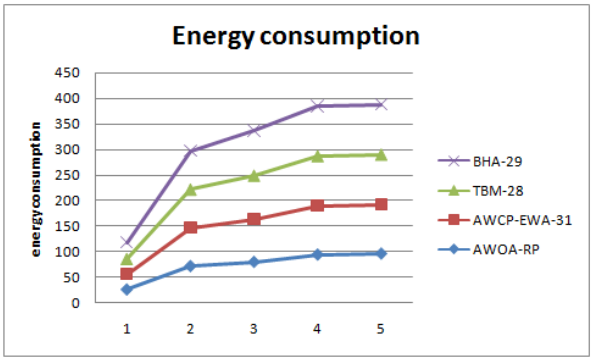


Figure 9 Energy consumption Vs Nodes

VI. CONCLUSION

In VANET, the enhanced road traffic conditions was employed for safe and secure communication,. The vehicles safety like traffic jam that were uncongested, with no delay time should be considered in order to offer effective communication. The improvement on efficient routing path was attained from these characteristics, identification of uncongested

path, and attack detection also. The AWOA-RP was employed for enabling enhanced routing path by adapted the AODV protocol in accordance with its channel capacity and the stability by the energy efficient, less cost and reduced traffic with the less delay, less packet loss ratio ,high throughput and condensed residual energy consumption with the improved performance. For improved packet delivery ratio, Trust based efficient residual energy levels, and the standard transmission delay, the proposed CCPDA should be well utilized. On behalf of the attacks from the malicious node identification and detection, the REPA performed well, i.e., the enhanced PDR value minimum overhead routing, reduced End-to-End delay in the network. The vehicular nodes and malicious nodes identification provided better communication with RSU, BS for enabling competent VANET. Hence, a secured communication path has been implemented among the usage of this approach which is applicable significantly in industrial wireless network.

Acknowledgement

This work is partially funded by Centre for System Design, Chennai Institute of Technology vide funding number is CIT/CSD/2021/012

REFERENCES

[1] A. K. Malhi, S. Batra, and H. S. Pannu, "Security of vehicular ad-hoc networks: A comprehensive survey," *Computers & Security,* vol. 89, p. 101664, 2020.

[2] O. Senouci, S. Harous, and Z. Aliouat, "Survey on vehicular ad hoc networks clustering algorithms: Overview, taxonomy, challenges, and open research issues," *International Journal of Communication Systems,* p. e4402, 2020.

[3] B. K. Chaurasia, W. S. Manjoro, and M. Dhakar, "Traffic Congestion Identification and Reduction," *WIRELESS PERSONAL COMMUNICATIONS,* 2020.

[4] A. Bhati and P. Singh, "Analysis of Various Routing Protocol for VANET," in *Cybernetics, Cognition and Machine Learning Applications*, ed: Springer, 2020, pp. 315-325.

[5] R. Kolandaisamy, R. M. Noor, M. R. Z'aba, I. Ahmedy, and I. Kolandaisamy, "Adapted stream region for packet marking based on DDoS attack detection in vehicular ad hoc networks," *The Journal of Supercomputing,* pp. 1-23, 2019.

[6] A. Srivastava, A. Prakash, and R. Tripathi, "Location based routing protocols in VANET: Issues and existing solutions," *Vehicular Communications,* p. 100231, 2020.

[7] C. Tang, W. Hu, S. Hu, and M. E. Stettler, "Urban Traffic Route Guidance Method With High Adaptive Learning Ability Under Diverse Traffic Scenarios," *IEEE Transactions on Intelligent Transportation Systems,* 2020.

[8] S. Abdelatif, M. Derdour, N. Ghoualmi-Zine, and B. Marzak, "VANET: A novel service for predicting and disseminating vehicle traffic information," *International Journal of Communication Systems,* vol. 33, p. e4288, 2020.

[9] N. Lin, L. Fu, L. Zhao, G. Min, A. Al-Dubai, and H. Gacanin, "A Novel Multimodal Collaborative Drone-assisted VANET Networking Model," *IEEE Transactions on Wireless Communications,* 2020.

[10] W. B. Jaballah, M. Conti, and C. Lal, "Security and Design Requirements for Software-Defined VANETs," *Computer Networks,* p. 107099, 2020.

[11] P. Sun and N. Samaan, "A Novel VANET-Assisted Traffic Control for Supporting Vehicular Cloud Computing," *IEEE Transactions on Intelligent Transportation Systems,* 2020.

[12] A. Paranjothi, M. S. Khan, R. Patan, R. M. Parizi, and M. Atiquzzaman, "VANETomo: A congestion identification and control scheme in connected vehicles using network tomography," *Computer Communications,* 2020.

[13] M. Zhou, L. Han, H. Lu, and C. Fu, "Distributed collaborative intrusion detection system for vehicular Ad Hoc networks based on invariant," *Computer Networks,* p. 107174, 2020.

[14] D. L. Guidoni, G. Maia, F. S. Souza, L. A. Villas, and A. A. Loureiro, "Vehicular Traffic Management Based on Traffic Engineering for Vehicular Ad Hoc Networks," *IEEE Access,* vol. 8, pp. 45167-45183, 2020.

[15] A. Musaddiq, R. Ali, R. Bajracharya, Y. A. Qadri, F. Al-Turjman, and S. W. Kim, "Trends,

Issues, and Challenges in the Domain of IoT-Based Vehicular Cloud Network," in *Unmanned Aerial Vehicles in Smart Cities*, ed: Springer, 2020, pp. 49-64.

[16] R. S. de Sousa, A. Boukerche, and A. A. Loureiro, "A distributed and low-overhead traffic congestion control protocol for vehicular ad-hoc networks," *Computer Communications,* 2020.

[17] C. Guo, D. Li, G. Zhang, X. Ding, R. Curtmola, and C. Borcea, "Dynamic Interior Point Method for Vehicular Traffic Optimization," *IEEE Transactions on Vehicular Technology,* 2020.

[18] D. Pandey and V. Kushwaha, "An exploratory study of congestion control techniques in Wireless Sensor Networks," *Computer Communications,* 2020.

[19] M. Obaidat, M. Khodjaeva, J. Holst, and M. B. Zid, "Security and Privacy Challenges in Vehicular Ad Hoc Networks," in *Connected Vehicles in the Internet of Things*, ed: Springer, 2020, pp. 223-251.

[20] K. J. Bwalya, "Spatial Intelligence and Vehicle-to-Vehicle Communication: Topologies and Architectures," in *Connected Vehicles in the Internet of Things*, ed: Springer, 2020, pp. 19-36.

IMPLEMENTING WIRELESS SENSORS FOR MONITORING COVID USING VARIANCE DETECTOR

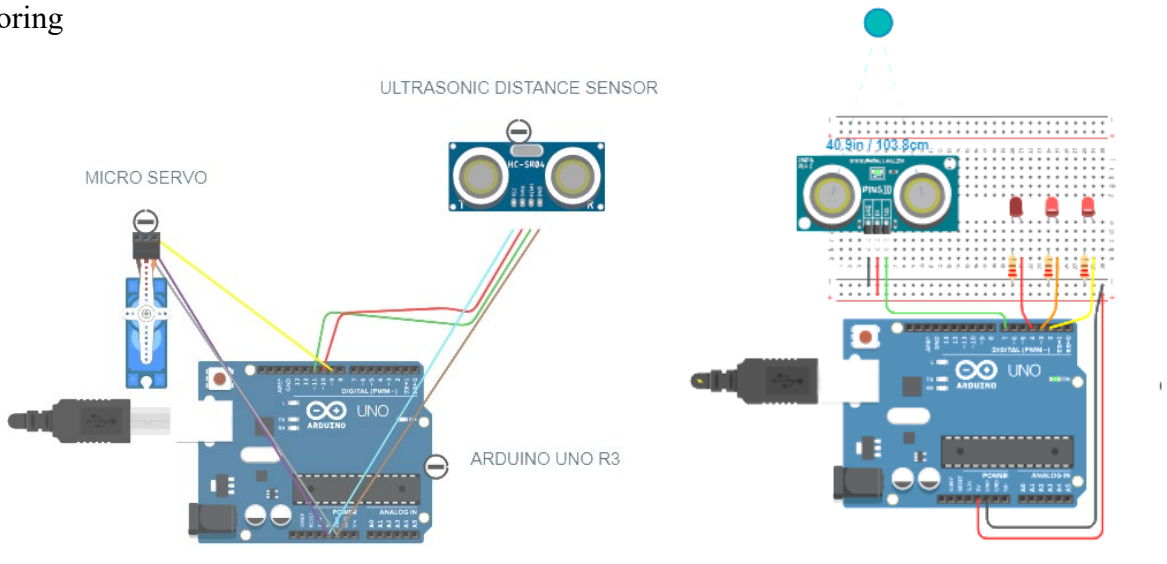
Abstract

This article focuses on implementing wireless sensors for monitoring exact distance between two individuals and to check whether everybody have sanitized their hands for stopping the spread of Corona Virus Disease (COVID). The idea behind this method is executed by implementing an objective function which focuses on maximizing distance, energy of nodes and minimizing the cost of implementation. Also, the proposed model is integrated with a variance detector which is denoted as Controlled Incongruity Algorithm (CIA). This variance detector is will sense the value and it will report to an online monitoring system named Things speak and for visualizing the sensed values it will be simulated using MATLAB. Even loss which is produced by sensors is found to be low when CIA is implemented. To validate the efficiency of proposed method it has been compared with prevailing methods and results prove that the better performance is obtained and the proposed method is improved by 76.8% than other outcomes observed from existing literatures.

Keywords: Wireless sensors; COVID; Energy consumption; Angle of inclination; Internet of Things (IoT).

1. Introduction

Recently, Internet of Things (IoT) has been emerging as a useful platform for medical applications. This helps the doctors all over the world to take care of their patients in the place where they are residing. This type of developments is possible only by using intelligent monitoring



devices like sensors. In line with above concern, sensors can also be designed and used for handling the current pandemic Corona Virus Disease (COVID). It is well know that no medicine is available for stopping COVID and only way is to maintain distance among individuals. Even the sensors can be designed for monitoring whether hands of individuals are sanitized or not. Therefore, an ultrasonic distance sensor can be integrated with high performance boards with a micro server as shown in Figure 1a. In addition, these sensors are designed for installing in all public places which even includes small shops. Therefore, in this case different LEDs can be connected for determining distance as shown in Figure 1b. This method is not a complete remedial method for COVID but by implementing this technology the people can preserve themselves from COVID and the monitoring station will try to stop the anomaly behavior of each individual.

(a) (b)

Figure 1. Implementation of proposed method for monitoring COVID (a) With micro server (b) Without micro server

1.1 Literature works

For integrating sensors in general public it is necessary that basic parameters and design of sensors have to be surveyed. Therefore, this section focuses on literature works by examining different parameters that have been discussed by several authors. The basic parameter for monitoring distance with basic principles of infrared illumination in mobile robots is discussed [1]. The authors have designed a low cost sensor but the major disadvantage is that the sensors are designed for monitoring very low distance which is up to 1 meter. Also, the angle of inclination for installing sensors is much hard task in public which is also discussed with necessary formulations. But after installation it is observed that energy consumption is much lesser and therefore it needs to be solved. For solving the problem of energy consumption high power wireless nodes have been deployed [2] and after deployment it is observed that even additional energy can be saved. However, there is advantage of much higher energy saving the method suffers from disadvantage that error rate is much higher where, exact parametric values cannot be obtained [2]. In addition, the distance between individuals and walls should be

maintained and in this case ultrasonic sensors have been designed with exact angle of inclination [3].

Even though a low cost sensor for moving environment is designed, straight angle measurement which is much necessary for monitoring in epidemic situation has been inapt. If both straight and cross angle measurements are made then, it would be a major advantage to the society [3]. To overcome the drawback in [3] a new technology using robotic application have been developed [4]. This kind of robotic environment will be mostly used in applications like vehicle monitoring where, exact distance of other vehicles can be informed to all individuals. Additionally, an alarm gust will also be deployed and if two vehicles are getting closer and any indication of accident is present then the alarm will be automatically switched on. But, installing the sensors of this type will be much higher and public cannot be able to afford much higher cost [4]. Consequently, for reducing the cost of sensor which needs to be installed in highways an ultrasonic pulse with a new method on selecting ground based points with necessary threshold parameters have been designed [5]. If this case [5] is considered then, cost of installing sensors in important points will reduce but the pulse will produce only less energy where information cannot be processed in a correct way between transmitter and receiver [5]. Furthermore, the sensors designed [1-5] cannot be able to monitor any one necessary parameter therefore, the authors [6] have combined two different techniques such as ultrasonic and infrared sensors for measuring distance. Even if accurate distance is monitored and the values are correct but when two methods are combined it is apparent that cost will be much higher [6]. From examining [6] it is observed that instead of using two different methods, if thermal sensors are implemented then, it is easy to predict the position of each individual [7]. Still, this virtuous method suffers from disadvantage that exact values can only be obtained if sensors are installed only on ceiling of all buildings [7] and angle of inclining sensors is missing. After some years a unique application on door step have been developed for computing group of peoples and exact distance between them which is denoted as smart construction monitoring [8]. In this application a lot of security measures have been updated from existing platforms and also an array of sensors has been implemented in grid-based structure. But the problem with array structure is that when sensors are grouped the data that is sent from transmitter to receiver have to be arranged in order for saving the energy consumption which is not possible with this type of IR sensors [8]. For implementing sensor arrays another method that is diverse from [8] have been introduced with photodiodes and bridge routes [9].

When photodiode is introduced exact measurement accuracy will be much higher and values will be obtained exactly within 2.5ms. But this creates an uncertainty where, bandwidth supplied by source should be much higher [9] where, it is not possible in all situations. Even for better installation a comparison has been made between ultrasonic and infrared sensors for detecting obstacles with exact detachment rate [10]. The aforementioned model is examined using vehicular applications and a time of flight method is enabled which makes the entire system to provide more accurate results. But unreasonably more energy have been used which makes the cost to rise beyond exact economic value [10]. When sensors are tested for integrating in medical application mainly for Corona Virus Disease (COVID) a new method using neural network have been applied. When this neural network is introduced it can able to detect the affected patients immediately with ten days. A common drawback in neural network is more data needs to be specified at same time and during each stage separate operations will be performed [11]. To introduce a new method that differs from neural network, a block chain technology has been integrated [12]. When this block chain is introduced then, it is very easy to store large amount of data. This creates may positive advantage over other methods because these systems can be introduced in mobile environments where correct of inclination will be specified. But the authors have not explained about security measures that are involved during this risk process on storing large amount of data [12]. To have complete exploration on security an energy cost behavioral model have been implemented [13] where, different position of sensor is acquired in accurate way and complete analysis for integrating in public places have been made with help of EC curve. However, uniform distribution of data is necessary for implementing in low cost condition [13].

1.2 Research gap and motivation

All the existing work which provides base information [1-13] fails to detect any one basic parametric value like energy consumption or cost of installation. Moreover, the existing articles have not focused on detecting any pandemic situations like COVID. But only basic parametric values have been measured to some extent. So the gap on monitoring the distance with high node energy at low cost has been missing in all deliberated literatures. Therefore, the authors have formed a base work by maximizing distance, energy and minimizing the cost of installation with

proper angle of inclination of sensors by formulating the objective function (Equation (7)) and integrating it with deep learning model for building the research gap.

1.3 Objectives

The main objective of this research work is to protect all people and stopping the spread of COVID by monitoring the distance between them using an online monitoring system (Things speak). In addition, the objective on minimizing the cost of installation with high energy for transferring the information from source to destination is also fulfilled. The main purpose of considering this objective is that no medicine is available for COVID and all people have to maintain distance between them with proper sanitizing lifestyle. This is possible only by implementing a monitoring device like sensors. Therefore, a cost-efficient sensor will be installed in all public places with correct angle of inclination for monitoring the distance between each individual.

2. Problem formulation

In this section the sensors will be designed according to the requirements by taking necessary parameters. Since the sensors are designed for pandemic situation all necessary parameters must be integrated together.

2.1 Calculation of distance

It is well known that for preventing the pandemic COVID situation no remedy is available and the only way is to maintain distance. But most of the people are not maintaining distance properly in all places. Therefore, for monitoring the distance between individuals a distance monitoring sensor will be placed and it will be formulated as given in Equation (1).

$$D_i = \sum_{i=1}^{n} CT_i \tag{1}$$

Where,

C represents distance between each individual

T_i denotes the time that both individuals are standing close to each other

Equation (1) represents the mutual formulation of sensors whereas, it differs in each situation based on selective criteria as shown in Equation (2).

$$D_i = \frac{3*V(i)}{10} \tag{2}$$

Where,

$V(i)$ represents the output voltage obtained from each sensor

Equation (2) will be implemented if the distance is calculated using output voltage. The value 3 indicates angle of inclining each sensor and 10 represents distance between each obstacle (In 360 degree view). In addition if both equations are implemented for designing the sensor then, it should provide full accuracy and it can be observed from Equation (3).

$$A_i = 100 - D_i \tag{3}$$

Where, A_i represents the accuracy of sensor for calculating the distance (For both Equation (1) and (2))

2.2 Angle of inclination

Angle of inclination for implementing sensors in outside environment is much important. If the sensors are not mounted properly in objects then, necessary parametric value cannot be evaluated properly and exact value will be much different where, Equation (3) will not be satisfied. Therefore, the position for mounting the sensors can be formulated in a collective way as given in Equation (4).

$$\emptyset_i(D_i) = \frac{\rho_i}{D_i^2} + \gamma_i \tag{4}$$

Where,

$\emptyset_i(D_i)$ indicates that angle of inclining sensors will vary depending on distance

ρ_i represents the replication of i^{th} target source

γ_i denotes the radiation of integrated sensors (0 or 1)

Equation (4) cannot be implemented if necessary constraints in Equations (1) and (2) are satisfied. Also, two different values will be indicated when radiation occurs through sensors. Therefore, the value will change at different times where, exact value will be obtained in all circumstances.

2.3 Energy consumption

Once the accuracy of designed sensor is much higher then, the next parameter for valuation is consumption of energy by all nodes that are connected with sensors. Always the sensors should be designed in a correct way that it should maximize the energy consumption which indicates that power transmitted form source to destination should be much higher. Calculation of power that is consumed by nodes (Both transmitter and receiver) can be expressed as,

$$E_i^{Tx,Rx} = TP_i^{A+D} + TP_i^{amplifier} \tag{5}$$

Where, TP_i^{A+D} denotes the total power that is supplied to all paths inside the sensors

$TP_i^{amplifier}$ represents the total power of amplifier that is present at transmitter side that includes peak-to-average power ratio

2.4 Total cost

It is necessary that even in small industries and in other yards the sensors needs to be installed. Therefore, it should be fabricated in a way with much less cost as given in Equation (6). When considering the cost of sensor units distance, angle of inclination and width of sensors will also be included.

$$C_i = (w_i * D_i) + \emptyset_i \tag{6}$$

Where, w_i represents the width of installed sensors

2.5 Objective function

By considering Equations (1) to (6) the objective function on implementing sensors for distance monitoring and hand sanitizing can be modeled as follows,

$$O(i) = max \sum_{i=1}^{n} D_i, E_i \ , min \sum_{i=1}^{n} C_i \tag{7}$$

If the objective function in Equation (7) is implemented then sensors will be fabricated in right way and exact results will be predicted in online monitoring system.

3. Optimization algorithm

To handle the pandemic COVID situation each and every individual should follow some consistency between them and all should regularly sanitize their hands. This is the only way for reducing the death rate across all countries. Therefore, for detecting the regularity of all people in public places an improved deep learning optimization algorithm which is denoted as Controlled Incongruity Algorithm (CIA) have been integrated with proposed model. One major advantage on choosing CIA is that it can able to provide solutions only if the problem is convex in nature and it can able to converge earlier [14, 15]. For integrating the deep learning model with proposed objective function some basic integration formulations are essential. One important

function in CIA is used for detecting losses of sensors using a conjugate function as given in Equation (8).

$$loss(i) = con(D_i - |l(D_i)|) \tag{8}$$

Where,

$|l(D_i)|$ represents the lagrangian function of distance

If this kind of deep learning algorithm is introduced then, loss which is produced by sensors will be much lesser and therefore exact langrangian controls can be made for distance measurements. For accurate measurements distance should always be measured from center of target and therefore, centric function can be given as,

$$t_i = \sum_{i=1}^{n} (\emptyset_i - m_i)\, 2\pi r \tag{9}$$

Where, m_i denotes the mid value of target

From Equation (9) it is clear that target centric function will be formulated only based on parameters like angle of inclination and area of circle. If sensors are not mounted at correct angle then the target cannot predict exact value which will result is wrong calculations. To avoid this situation area of circle with midpoint is also added for accurate calculations. Finally, the mean value of distance followed by different people will be calculated as given in Equation (10).

$$mean(i) = \frac{\max D_i - \min D_i}{2\pi r} \tag{10}$$

From Equation (10) the mean incongruity value will be calculated for all individuals. If any individual walks to maximum distance the sensor will be automatically switched off where, in this case energy loss will occur. To overcome this loss difference between maximum and minimum distance will be calculated in addition with area of circle. This makes the optimization process to converge earlier and within small values the sensor will be switched off and lot of energy can be saved. The flow chart of proposed CIA is shown in Figure 2.

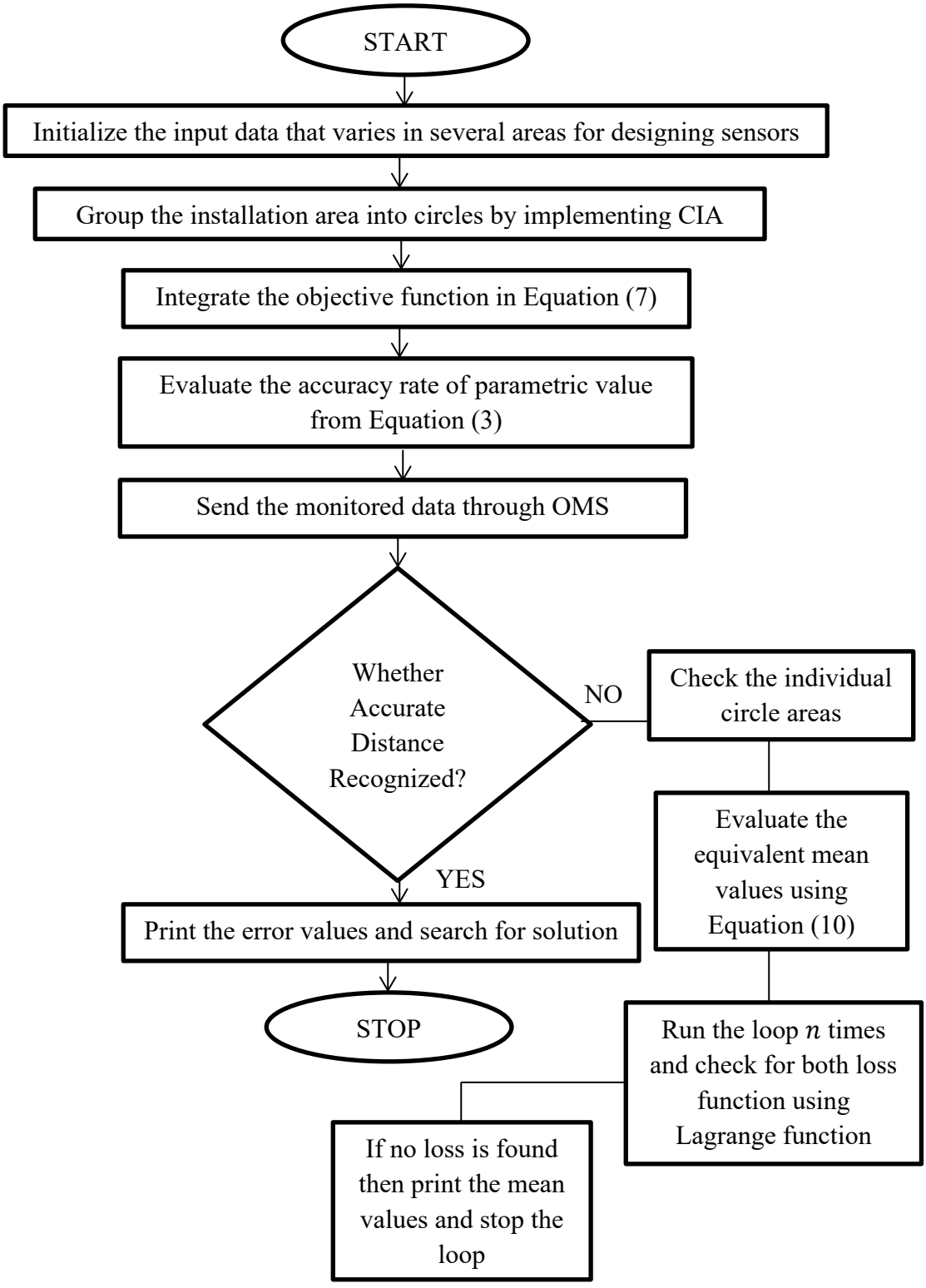


Figure 2. Flow chart of CIA for monitoring distance between entities

4. Results and discussion

The proposed method of installing sensors for being alert to prevent COVID have been monitored using an online monitoring system and it is executed using MATLAB where, different parametric values are monitored and they are listed as following scenarios.

Scenario 1: Angle of inclination

Scenario 2: Measurement of distance

Scenario 3: Calculation of energy consumption

Scenario 4: Cost of installation

Scenario 5: Proportion of gap

Scenario 1

For installing sensors one major prerequisite is inclining angle which is discussed in this scenario. For each public building the angle of inclination is different and it depends on radius of buildings. If correct angle of inclination is not provided then distance between two individual cannot be monitored properly which results in failure of installing sensors. Therefore, angle of inclination for mounting sensors in suitable places should be followed correctly. It is observed that angle of inclination for small radius will be tilted up to -50 degrees whereas, for large radius of buildings the angle of inclination will be 90 degrees. If in this angle the sensor is mounted then it can be able to detect each individual within the distance of 100 meters as shown in Figure 3.

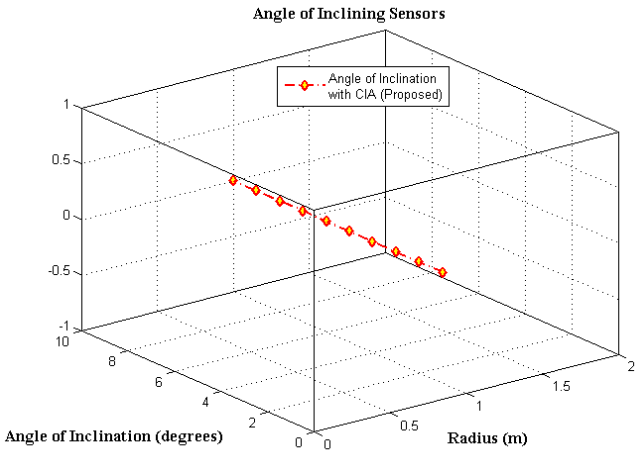


Figure 3. Angle of inclination (in degrees)

From Figure 3 it can be seen that sensors are inclined correctly at precise angles where; maximum and minimum limit will be -50 to 90 degrees. If radius of building is much higher then, angle of inclination can be calculated using Equation (4). This inclination value will be automatically fed as input to Things speak, an online monitoring system and if inclination angle varies an alert will be sent to the user and therefore, they have to re-install it. This parameter is not compared with any existing method because this is an unique parameter and no previous exists for comparison states.

Scenario 2

Once the sensors are mounted at correct angle, distance can be calculated which is conversed in this scenario. Generally, sensors should be able to monitor distance with less interpretation values. For monitoring long distance above 50 meters the interpretation values should be constant. By this monitoring station can detect whether each and every individuals are maintaining minimum distance between them. If any person is creating anomaly then interpretation will be stopped and alert will be sent to the users. In this case the data such as mobile numbers and names can also be fed as input and therefore, if any user violates guidelines then a message will be sent.

Figure 4 shows calculation of distance which is obtained in a small yard. The results are monitored for a single day using Thing speak and it is visualized in MATLAB. For the proposed method if the users are between 50-100 meters a constant voltage will be supplied because in this case there is no need for monitoring but if the users are closer between 10-40 meters then, more interpretation is needed which is greater than 3 volts. But in existing method [1] even if individuals are at long distance more interpretation is needed and more wastage will occur. Therefore, the proposed method proves to be more efficient then existing method for monitoring distance between individuals.

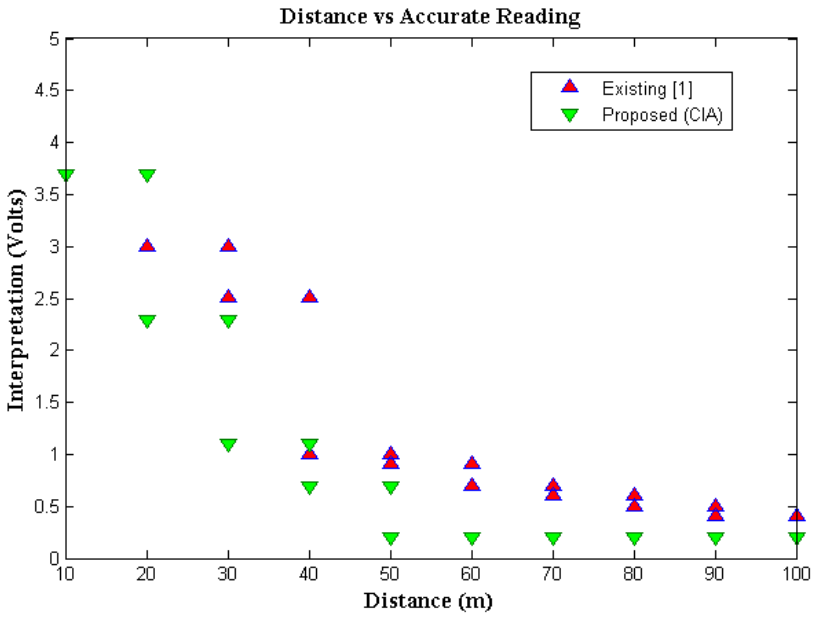


Figure 4. Distance vs accurate reading

Scenario 3

In this scenario the basic parameter that is necessary for sensors during transfer of information which is referred as energy consumption is determined. For transferring information from source to destination amount of energy supplied to nodes should be much higher. This is one important characteristic that is much needed for proposed method. Figure 5 portrays energy that is consumed by nodes for transferring information and it can be seen that more energy is supplied in proposed method because for monitoring this pandemic situation more energy is required as number of peoples are higher.

It can also be observed that the packets arrived at receiver will consume up to 8 mW for 100 nodes whereas, the existing method [2] consumes only 7.19 mW. In this case if energy is supplied much lower then information will not reach the receiver and therefore, the interpretation values will be much higher for existing method as shown in Figure 2.

Scenario 4

The sensors designed for this COVID situation have to be installed in all local places therefore, cost of installation should be much lesser. Here, the cost will be calculated by taking into account the width of current abode and angle of inclination. If these two bounds are considered then, automatically cost of installing sensors will be much lesser. Figure 6 displays the simulation results that is obtained after integrating Equation (6). From Figure 6 it can be observed that cost

of installing sensors through proposed method is much lesser than existing method [12]. For example if number of nodes is 60 then, cost of installation for proposed method will be 46.1K whereas, for existing method it is found to be 53.8K respectively.

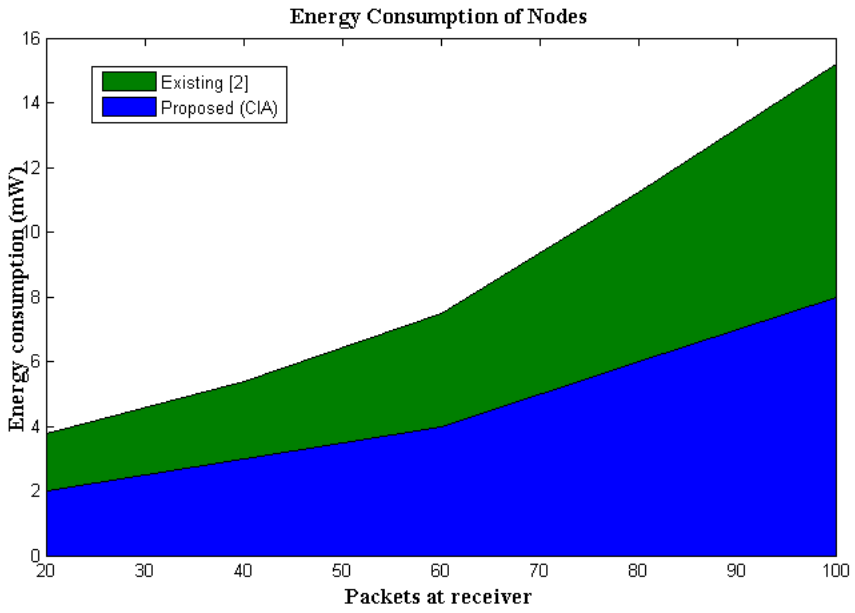


Figure 5. Comparison of energy consumption

This shows that if angle of inclination is correct then, cost of installation will be much lesser. This parameter will not be fed as input values to Things speak but for providing information to public it has been included. Even a small dealer should be in a position to buy this sensor because it is much necessary to monitor distance which is the only way to stop this pandemic situation. So, if the proposed method is implemented in real time all small dealers can buy this type of sensor and they can install it at much lower rate.

Scenario 5

In this scenario accurate measurement distance and gap will be calculated between each individual. This is treated as special case because once an individual arrives at a shop a small sanitizer magnum will be kept where, if the person arrives within a particular distance, they will be alerted to sanitize their hand. Next after sanitizing percentage of gap between each individual will be observed and if either hands are not sanitized and if percentage of gap is low then an

alarm message will be sent to particular person. This is monitored by Things speak and the results are visualized in MATLAB.

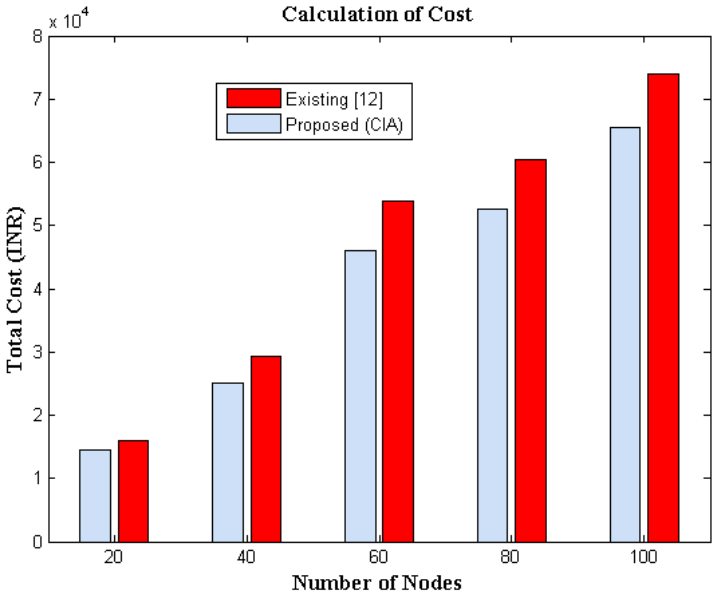


Figure 6. Calculation of total cost (INR)

Figure 7 shows the simulated results by taking distance in X-axis and percentage of gap in Y-axis. It can be seen in Figure 7 that a person is ready to go into s shop and when that person is outside at a distance of 50 meters then percentage of gap observed will be 36. But once the person enters into the shop and it the shop is much smaller to a distance of 10 meter then gap maintained by that person is 1.3 whereas, for existing method [6] more gap is specified and it is not much possible for small sized shops. This proves that proposed method is capable of monitoring exact values than other existing methods.

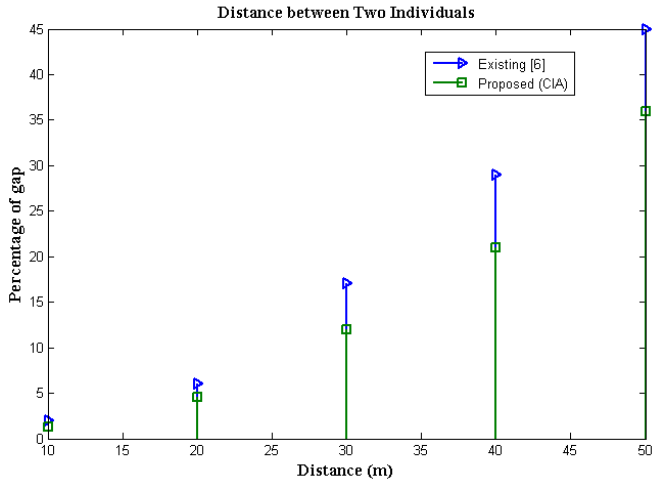


Figure 7. Accurate distance vs percentage of gap

4.1 Performance analysis of CIA

Beyond some real time factors it is much important to analyze the simulation time of implemented algorithm and it should be compared with existing one for understanding the performance and accuracy of parametric values. Therefore, when the proposed method is visualized in MATLAB simulation time is calculated by varying the number of rounds and is shown in Figure 7.

From Figure 7 it can be seen that simulation time for proposed CIA is much lesser than existing method [11]. For accurate checking the numbers of rounds have been considered as 100 and even for all rounds simulation time is much lesser and it is considered as foremost advantage of integrated algorithm. For example, if the number of round is equal to 60 then simulation time for proposed method is 3.81 seconds whereas for existing method [11] it is equal to 5.16 seconds. This proves that proposed method is capable of converging solutions at earliest and all the parametric values will be obtained within fraction of seconds.

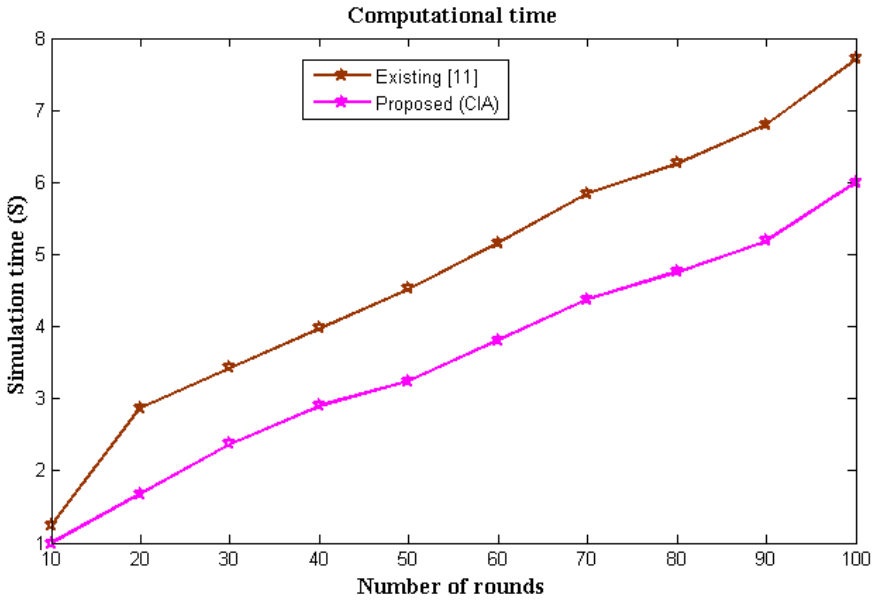


Figure 8. Computational time (seconds) of CIA

5. Conclusion

Many countries around the world are suffering from COVID and day-by-day the count of people affected by this disease keeps on increasing. Even though preventive measures can be taken by staying at home it is not possible for all working people to stay at home all the time. Therefore, only way of preventing this pandemic situation is to maintain distance between two individuals which is monitored by wireless sensors and it has been addressed in this article. Most of the people are not maintaining even minimum distance between them when they are going in public places. To stop this sensors are designed in a way that it can monitor the movement of people using Things speak, an online monitoring system which will predict exact values when wireless sensors are connected.

Moreover, the sensors are designed where, even small dealer can able to afford it and it can provide high energy by passing the information from transmitter to receiver. In addition, for effective functioning CIA algorithm have been implemented which can be treated as other advantage of proposed method. The projected method is simulated and compared with existing methods and in future it can be extended for detecting the humans by integrating it in their apparels with necessary energy consumption.

Acknowledgement:

The author (VT) gratefully acknowledges the Deanship of Scientific Research, King Khalid University (KKU), Abha-Asir, Kingdom of Saudi Arabia for funding this research work under the grant number RGP.2/58/42.

References

1. Denkena B, Möhring HC, Litwinski KM (2008) Design of dynamic multi sensor systems. Prod Eng 2:327–331. https://doi.org/10.1007/s11740-008-0102-8
2. Ali A, Abo-Zahhad M, Farrag M (2017) Modeling of Wireless Sensor Networks with Minimum Energy Consumption. Arab J Sci Eng 42:2631–2639. https://doi.org/10.1007/s13369-016-2281-5
3. Ishihara M, Shiina M, Suzuki S (2009) Evaluation of Method of Measuring Distance Between Object and Walls Using Ultrasonic Sensors. J Asian Electr Veh 7:1207–1211. https://doi.org/10.4130/jaev.7.1207
4. Okugumo M, Kimura A, Ohki M, Ohkita M (2008) Development Research on High Performance Ultrasound Sensor System. IEEJ Trans Electron Inf Syst 128:55–61. https://doi.org/10.1541/ieejeiss.128.55

5. Carullo A, Parvis M (2001) An ultrasonic sensor for distance measurement in automotive applications. IEEE Sens J 1:143–147. https://doi.org/10.1109/JSEN.2001.936931
6. Mustapha B, Zayegh A, Begg RK (2014) Ultrasonic and infrared sensors performance in a wireless obstacle detection system. Proc - 1st Int Conf Artif Intell Model Simulation, AIMS 2013 487–492. https://doi.org/10.1109/AIMS.2013.89
7. Singh S, Aksanli B (2019) Non-intrusive presence detection and position tracking for multiple people using low-resolution thermal sensors. J Sens Actuator Networks 8:. https://doi.org/10.3390/jsan8030040
8. Mohammadmoradi H, Munir S, Gnawali O, Shelton C (2018) Measuring people-flow through doorways using easy-to-install IR array sensors. Proc - 2017 13th Int Conf Distrib Comput Sens Syst DCOSS 2017 2018-January:35–43. https://doi.org/10.1109/DCOSS.2017.26
9. Nemecek A, Oberhauser K, Zimmermann H (2006) Distance measurement sensor with PIN-photodiode and bridge circuit. IEEE Sens J 6:391–397. https://doi.org/10.1109/JSEN.2006.870164
10. Adarsh S, Kaleemuddin SM, Bose D, Ramachandran KI (2016) Performance comparison of Infrared and Ultrasonic sensors for obstacles of different materials in vehicle/ robot navigation applications. IOP Conf Ser Mater Sci Eng 149:. https://doi.org/10.1088/1757-899X/149/1/012141
11. Al-Qaness MAA, Ewees AA, Fan H, Aziz MA El (2020) Optimization method for forecasting confirmed cases of COVID-19 in China. Appl Sci 9:. https://doi.org/10.3390/JCM9030674
12. Peker YK, Rodriguez X, Ericsson J, et al (2020) A cost analysis of internet of things sensor data storage on blockchain via smart contracts. Electron 9:. https://doi.org/10.3390/electronics9020244
13. Yufu J, Tianlin D, Jian S (2005) Analysis on energy cost for wireless sensor networks. ICESS 2005 - Second Int Conf Embed Softw Syst 2005:144–151. https://doi.org/10.1109/ICESS.2005.30
14. Khaleghi A (2018) Based on A Deep Learning Method. 2018 8th Conf AI Robot 10th Rob Iranopen Int Symp 73–81
15. Görnitz N, Kloft M, Rieck K, Brefeld U (2013) Toward supervised anomaly detection. J Artif Intell Res 46:235–262. https://doi.org/10.1613/jair.3623