ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2025.1833.569

FEATURE SELECTION METHOD USING HYBRID SWARM WITH IMPROVED FUZZY C-MEANS CLUSTERING IN DATA MINING FOR DISEASE DETECTION

M. Birundha Rani[1], Dr.A. Subramani[2]

[1]*Research Scholar, Mother Teresa Women's University, Dindigul, Tamil Nadu, India.*
e-mail: birundhaphd2021@gmail.com, orcid: https://orcid.org/0009-0003-4968-6234
[2]*Assistant Professor, Department. of Computer Science, M.V. Muthiah Govt. Arts College for Women, Dindigul, Tamil Nadu, India. e-mail: Subramani.appavu@gmail.com, orcid: https://orcid.org/0000-0002-8303-8770*

Received: June 27, 2025; Revised: September 11, 2025; Accepted: October 06, 2025; Published: October 30, 2025

SUMMARY

A crucial method for reducing the dimensionality issue in DM (Data Mining) tasks is FS (Feature Selection). Conventional techniques for FS do not scale well in vast spaces. The HPSO-IKM approach has a rather long processing time, so future studies will keep enhancing the technique's stages to reduce the duration of detection. PSO's poor local search capability and lagging convergence in the refining search phase prevent it from mitigating the effects of poor initialization by reducing the greatest number of IC (Intra-Clustering) faults. This paper suggests a novel approach to the dimensionality issue, in which a good feature subset is produced through combining the correlation metric using clustering. Following Z Score Normalization (ZSN) for pre-processing, a computational model is constructed to identify the pertinent features based on pertinent constraints, and a structure is developed by extracting features via Principal Component Analysis (PCA). Next, utilizing Multi-Objective Glowworm Swarm Optimization using Improved Fuzzy C-Means Clustering (MOGWO-IFCM), unnecessary features are removed, and non-redundant features are chosen from every cluster based on correlation measures. This approach employs the IFCM technique for optimizing the initial clustering center after receiving the optimal solution as an initial clustering center with the GSO (Glowworm Swarm Optimization) technique. Utilizing the Modified Long Short-Term Memory (MLSTM) classifier, the suggested approach is tested on UCI datasets, and the outcomes are contrasted with those of other well-known FS methods. Percent-wise criteria are employed to confirm the accuracy of the suggested technique with varying numbers of pertinent features. The suggested technique's accuracy and efficiency are demonstrated by the outcomes of the experiment.

Key words: *data mining, swarm optimization, z score normalization.*

INTRODUCTION

High-dimensional (HD) data, particularly with several features, is becoming more and more common in ML (Machine Learning) challenges [1][2][19][37]. To address these issues, a lot of researchers concentrate on their studies. In addition, significant characteristics from these highly dimensional variables and information should be drawn. The amount of noise and redundant data was reduced by

using statistical approaches. FS is crucial since you shouldn't use every feature for training an algorithm; instead, employ correlated and non-redundant features to enhance the system [32].

Additionally, it helps to train the model more quickly while simultaneously reducing its complexity, making it simpler to grasp, and enhancing its accuracy, precision, and recall metrics. The importance of FS is demonstrated by four key factors. Reduce the number of parameters by initially limiting the framework. The next steps are to shorten the training period, prevent the dimensionality curse, and improve generalization to lessen overfilling. Large datasets with numerous characteristics or attributes that affect the data's applicability and usefulness are common in the area of DP (Data Processing) and analysis [3][4][5]. Moreover, balanced and imbalanced data present a categorization issue [34]. Determining the optimal framework with high accuracy in prediction and low error rates is a further motivation [6][33].

FS is the process of reducing the original feature set to a smaller one while keeping the pertinent data and eliminating the duplicate data [6][7]. A reduced quantity of training samples is required to resolve this problem. The salient point of this example would be the application of FS and FE (Feature Extraction) methods. FS techniques are frequently employed to boost a classifier's capacity for generalization [8][9][10]. In order to obtain the best accuracy, via examining the dataset's outcomes with and without the identification of Important Features (IF) using PCA techniques. In order to generate predictions with a high degree of accuracy, ML fundamentally needs a large amount of data, features, and variables. Creating the prediction algorithm is not as critical as FS. Furthermore, the prediction's outcome will only deteriorate if the dataset is applied without any pre-processing.

In keeping with earlier studies, [36] assigns feature significance to classification models for Indonesian phenotypes of the CRC (ColoRectal Cancer) cases. Furthermore, the significance of these traits for EC (emotion classification) and ESS (Emotional Speech Synthesis) will be investigated in future genetic investigations of CRC [11] when they are included as covariates. Additionally, [12][13] conduct FIA (FI Analysis) with encouraging outcomes for the IRS (Industrial Recommendation System). In this study, researchers demonstrate the importance of the UCI dataset's FS. This study examined whether utilizing several approaches has any impact on classification performance and whether using EFS (Ensemble FS) methods can produce more robust FS procedures.

The area of EL (Ensemble Learning), where several (unstable) classifiers are integrated to produce a more stable and effective ensemble classifier, provides the justification for this theory. Similarly, combining single, less stable feature selectors may result in more robust FS methods. The present research emphasizes strategies in the big feature/small sample size domains, where this problem is particularly crucial.

This is the way the remaining portion of the study is organized. The methods utilized to test and analyze the techniques' robustness are explained in Section 2. The strategies for ESF are then covered in section 3, and section 4 presents the study's outcomes. Section 5 deals with the conclusion and framework.

RELATED WORK

In a distributed setting, [14] examined the selection of sub-features from every feature utilizing fuzzy techniques while protecting participant anonymity. Conditional expectation, which generates 2 fuzzy sets utilizing the Borel set to assist in selecting the sub-feature inside a particular interval, depends on fuzzy random variables. Nonetheless, the process offers a more effective means of choosing sub-features in many scenarios. To that end, this research [35] introduces FSS-FWNN, an intelligent feature subset selection method for large data categorization.

In order to identify businesses that engage in financial statement fraud, Ravisankar et al. [15] utilized DM methods like Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic Programming (GP), Group Method of Data Handling (GMDH), Logistic Regression (LR), and Probabilistic Neural Network (PNN). A dataset comprising 202 Chinese enterprises is employed to evaluate and compare each of these methods with and without FS. PNN fared better than all other

methods without FS, and GP and PNN performed better than other methods with FS and a similar accuracy. But in this instance, human prejudice is unavoidable, and the decisions are frequently biased. The method of using PSO to locate the centers of a user-specified number of clusters is described in [16]. The first swarm is seeded using K-means clustering, an extension of the technique. To improve upon the clusters established by K-means, this second approach primarily employs PSO.

In order to tackle the consequences that result from OFS (Optimal FS), [38] employed a novel lightweight mechanism as a feature selection approach. In large DM, the Accelerated Flower Pollination (AFP) technique is utilized for FS. This technique shortens processing times while increasing FS accuracy. However, mining through these HD data results in the creation of an ideal feature subset, which causes an unmanageable computational demand to develop rapidly.

When compared to other techniques like Direct Discriminative Pattern Mining (DDPMine) and Iterative Sampling Based Frequent Itemset Mining (ISbFIM), where enumeration of whole feature combination was completed, Devi & Sabrigiriraj [17] enforced weighted entropy frequent pattern mining (WEFPM) for FPM to achieve better computation time. Therefore, the WEFPM technique that is being used attempts to determine only those particular, frequent patterns that the user needs to detect.

The new hybrid FC technique dubbed SCAGA, which combines the Genetic Algorithm (GA) and the Sine Cosine Algorithm (SCA), was suggested by Abualigah & Dulaimi [18]. The 2 primary search strategies used in optimization techniques are usually exploration of the search space and exploitation to find the best answer. A fundamental SCA and other comparable techniques from published literature, such as Ant Lion Optimization (ALO) and Particle Swarm Optimization (PSO), were also used to compare the findings. By eliminating repetitive, noisy, and irrelevant variables from the original dataset, these strategies suggested ways to create a predictive model that reduces the classifier's prediction errors through choosing useful or significant features.

Fong and associates [31] A unique lightweight FS is suggested to address this issue, which is mostly related to the high-dimensionality and streaming style of data feeds in BD (Big Data). Specifically, the FS procedure is made to mine real-time streaming data utilizing Accelerated PSO (APSO) swarm search, which improves analytical accuracy in a manageable amount of computational time. In this study, we evaluate the novel FS approach for the evaluation of performance on a set of BDS with a very high degree of dimensionality. But the search space from which an ideal feature subset is created expands exponentially in size when mining over HD data, creating an unmanageable computational demand.

A novel multisource latent feature selective ensemble (SEN) modeling technique was presented by Tang et al. [20]. Initially, based on the properties of the modeling data, the input features are separated into various subgroups. Second, the multi-layered selection methods that produce the extracted multisource latent features are defined by the combined data value orderly for every subgroup, the feature reduction ratio, and the feature contribution ratio. A hybrid FS technique called FAFS_HFS was presented by Gong et al. [21] and depends on feature subsets produced by factor analysis. First, using factor analysis, this technique creates feature subsets based on each feature's highest load (maximum explanatory power). Subsequently, every feature subset's redundancy is eliminated using Sequential Forward Selection (SFS) and minimal Redundancy and maximal Relevance (mRMR). But the majority of research on FS techniques focuses on the single feature or the entire feature subset, ignoring the impact of correlation and redundancy on the feature subset's ability to classify data.

For unsupervised non-Gaussian FS in the framework of finite generalized Dirichlet (GD) mixture-based clustering, Fan et al. [22] suggested a variational inference approach. Then, simultaneously estimate all the involved parameters in a closed form under the suggested principled variational structure, in order to ascertain the complexity (i.e., both model and FS) of the GD mixture.

ReliefF-BSTA, a simple and effective hybrid FS technique, was presented by Huang et al. [23] and is based on the binary state transition technique and ReliefF. The filter phase and the wrapper phase are the 2 stages of this procedure. This approach offers 3 distinct benefits.

The (FRS) Fuzzy Rough Set-based bi-selection problem for data reduction was studied by Zhang et al. [24]. To be more precise, representative cases are chosen before key features using the unified ideas of fuzzy granules' importance degrees. Finally, a bi-selection strategy incorporating instance and FS techniques is provided for data reduction, based on Fuzzy Rough Sets (BSFRS). However, the simultaneous FS and instance BSFRS has received little attention.

Inference: Huge or Large-scale datasets are currently very frequent, and there has been a lot of interest in the advancements of DR (Data Reduction) strategies for these types of information. The clustering technique, along with swarm intelligence, is a potent tool for managing uncertainty in real-valued information. It has been extensively utilized in DR, such as FS techniques and instance selection techniques. However, the simultaneous feature and instance selection, which depends on the clustering approach, has received little attention. The optimal cluster-based problem for DR was examined in this paper.

PROPOSED METHODOLOGY

The techniques for Disease Prediction (DP) with 2 benchmark datasets that are readily available and it will be addressed in this part. There are several stages to this research, from data collection to DP. Data transformation techniques and feature scaling can be utilized to pre-process data in the initial stage. The next stage is to build the suggested framework with several ML methods. The following step of the procedure uses an ensemble technique to improve the accuracy of the model. An extensive workflow architecture design is displayed in Figure 1.

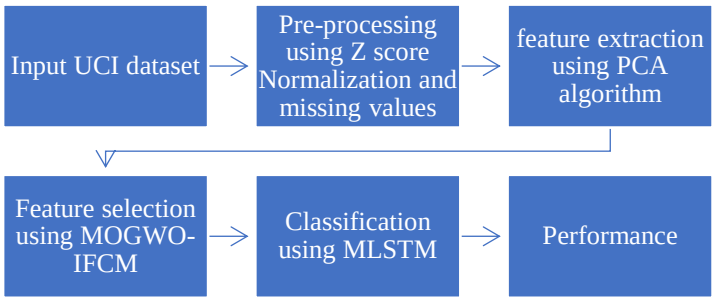

Figure 1. Workflow architecture design

Data Pre-processing

Missing Values: Determining and identifying duplicate or missing values in the dataset is the initial step in the pre-processing of the information. The statistical importance of the outcomes derived from the analysis of models may be impacted by the insufficient information implied by missing data in a dataset. The general validity and correctness of the study may be affected when there are missing data. Consequently, instead of totally suppressing the observations, it is advised to substitute the missing values via either a user-defined constant or the dataset average in order to optimize the study's efficacy.

Z-score Normalization

Preprocessing the training dataset was necessary before generating an algorithm from it. ZSN was the preprocessing method used in the current study. This method was said to be possible to increase the model's accuracy in the conducted study [25]. The dataset was converted into the UV (Unit-Variance), and a zero-mean distribution is employed in this method. There were six features in the original dataset, and every feature had various ranges. Consequently, the normalization procedure was used to convert the dataset into a predetermined range. The two datasets' performances were compared in this study. Two datasets were preprocessed by ZSN; the first dataset was the original, unaltered by preprocessing. Equation 1 displays the ZSN formula.

$$\mathbb{x}_i' = \frac{\mathbb{x}_i - \mu}{\sigma} \quad (1)$$

The normalized data is denoted by, whereas the original data is represented by. The Mean and standard deviation of the data are denoted by and, respectively.

With the purpose of selecting appropriate features, the suggested work develops a system that integrates FE via PCA and FS by employing IGR (Information Gain Ratio). This is made using a methodical approach. Eliminating the outliers is the initial stage. The observed data that varied significantly from the observed data are known as outliers. Noise is another term for it. Either attribute noise or data noise may be present. To eliminate the outliers, the first stage in the DM procedure is DC (Data Cleaning). Utilizing PCA, the IF, that is, the most relevant features is extracted in the second stage, which is FE.

Principal Component Analysis (PCA)

A prerequisite for solving ML issues is FE. By combining the original dimensions, FE generates novel dimensions. This can be broadly divided into 2 methods. For unsupervised learning, these include the projection approach, which employs PCA, LDA (Linear Discriminate Analysis), etc. For SL (supervised learning, the compression technique utilizes shared data as well as data theory. There are two broad categories for DM. The two types of DM are descriptive and predictive. Using the data provided, the former builds one or more algorithms and creates a model for prediction. The latter is employed to generate the data's report and label features.

The rationale behind the use of PCA is its excellent performance in handling linked attributes [26]. It finds similarities and variances among every feature and recognizes patterns within the data set. It serves as an effective tool for data analysis. The UCI repository is used to select the illness data collection. The initial data are selected, as is the original data average. One computes the covariance matrix. The (EV) Eigen values and Eigen vectors are then selected from the covariance matrix. The primary component of the illness data set is determined by selecting the eigenvector with the greatest EV. The relationship among the data elements is most essential when it is revealed. Ascending order is employed to sort the EVs. The most important data are selected, and the least important ones are eliminated or thrown away. This reduces the HD data to lower-dimensional data.

Step 1: The following formula is applied to compute the mean. [26].

$$\text{mean} = \frac{\sum_{i=1}^{n} \mathbb{x}_i}{n} \tag{2}$$

Here, the summation of all the values in the sample can be represented as, is the sample size, and the th observation of the random variable x can be denoted as

Step 2: Finding the data spread in the disease dataset is done by computing the variance [17]. To determine the data variation in the sample dataset, the variance is calculated using the formula given below.

$$Var(\mathbb{x}) = \frac{\sum_{i=1}^{n} (\mathbb{x} - \mathbb{x}')^2}{n-1} \tag{3}$$

Step 3: To determine the relationship between two classes, covariance is computed. Positive numbers imply that the two dimensions increase and decrease simultaneously, while negative values indicate that one-dimension increases while the other decreases. A zero value suggests that the two dimensions are unrelated.

$$\text{Covar}(\mathbb{x}, \mathbb{y}) = \frac{\sum_{i=1}^{n} (\mathbb{x} - \mathbb{x}')(\mathbb{y} - \mathbb{y}')}{n-1} \tag{4}$$

Step 4: Calculating the covariance matrix is the next step.

$$\text{Covar}(\mathbb{x}) = \begin{bmatrix} \Sigma\, \mathbb{x}_{12}/n & \sum \mathbb{x}_1 \mathbb{x}_2/ & \sum \mathbb{x}_1 \mathbb{x}_c/n \\ \sum \mathbb{x}_2 \mathbb{x}_1/n & \sum \mathbb{x}_{22}/n & \sum \mathbb{x}_1 \mathbb{x}_c/n \\ \cdots & \cdots & \cdots \\ \sum \mathbb{x}_c \mathbb{x}_1/n & \sum \mathbb{x}_c \mathbb{x}_2/n & \sum \mathbb{x}_{c2}/n \end{bmatrix} \tag{5}$$

Here, the number of scores in each of the data sets can be denoted as and the covariance matrix can be referred as Covar (x). A deviation score from the th data set is denoted by. The variance of elements from the th data set is denoted as. The covariance for elements in the and data sets is denoted by.

Step 5: Determine the covariance matrix's EVs and Eigen vectors.

Step 6: Apply the PCA assumption to the covariance matrix's diagonalization. The most significant qualities are assumed to have the greatest variance, and the ordered set of principal components, denoted by p, is the variable. By diagonalizing, the covariance value is maximized, and the redundancy is decreased. The variance linked to the p is discovered and arranged according to the calculated principal variance.

Step 7: In order to locate the pertinent vector, reduce the dimensionality. From lowest to highest, the calculated EVs are sorted. As the primary component, the vector having the highest values is selected. The less significant value is disregarded. Because the value is calculated utilizing mathematical models, it is more accurate.

Step 8: Make a matrix where the columns are the Eigen vectors. To extract the pertinent features, the vector with the highest significance is selected.

Feature Selection Using MOGWO-IFCM

The recommended technique for feature selection is MOGWO-IFCM. The FCMI clustering algorithm computes the center points of normal and abnormal data and groups homogenous values of data in the feature space into clusters in an unsupervised manner. The suggested approach yields a fuzzy clustering with C clusters for a C-class problem. The quality of every feature subset was assessed by contrasting the outcome with the desired outputs. Thus, the method that is being provided is a part of filter techniques that employ MOGWO and a random search technique.

Improved Fuzzy C-Means Clustering Algorithm: In addition to optimizing the sample's features, the kernel function's clustering approach can translate the sample's input space to an HD feature space and cluster within it. In terms of performance, the kernel function clustering technique outperforms the standard clustering technique. With nonlinear mapping, it can identify, extract, and enhance beneficial attributes to produce faster and more accurate grouping. The OF (Objective Function) of FCM, the given sample set can be written as follows [27]:

$$J_m(\mathbb{U},\mathbb{V}) = \sum_{i=1}^{N}\sum_{j=1}^{C} \mathbb{u}_{ij}^{m}\left\|\mathbb{x}_i - \mathbb{v}_j\right\|^2 \quad (6)$$

When the following conditions must be met by,

$$\forall i, \sum_{j=1}^{C} \mathbb{u}_{ij} = 1; \forall i,j, \mathbb{u}_{ij} \in [0,1]; \sum_{i=1}^{N} \mathbb{u}_{ij} > 0 \quad (7)$$

represents the number of clusters in Formula (2.2). is the membership value of the sample relating to the, the number of samples can be denoted as, and the fuzzy partition matrix can be represented as =. An exponential weight that affects the fuzzy degree of the membership matrix can be denoted as, and the cluster, a collection of C cluster centers, can be represented as. The following describes the manner in which the OF may be created using the LM (Lagrange Multiplier) technique:

$$\bar{J}(\mathbb{U},\mathbb{V},\lambda) = J_m(\mathbb{U},\mathbb{V}) + \sum_{i=1}^{N}\lambda_i\left(\sum_{j=1}^{C}\mathbb{u}_{ij} - 1\right) \quad (8)$$

The sample distance in the feature space can be defined as follows by inserting a non-linear mapping

$$\|\phi(\mathbb{x}_i) - \phi(\mathbb{v}_i)\| = K(\mathbb{x}_i,\mathbb{x}_i) = K\left(\mathbb{v}_j,\mathbb{v}_j\right) - 2k\left(\mathbb{x}_i,\mathbb{v}_j\right) \quad (9)$$

The OF of the IFCM is as follows: If K is the kernel function in Formula (2.4), then

$$J_\phi = \sum_{i=1}^{N} J_i = \sum_{i=1}^{N}\sum_{j=1}^{C} \mathbb{u}_{ij}^{m}\left\|\phi(\mathbb{x}_i) - \phi\left(\mathbb{v}_j\right)\right\|^2 \quad (10)$$

The (GF) Gaussian function was selected as the K in this research in the following manner:

$$K(\mathbb{x},\mathbb{y}) = \exp[-(\mathbb{x}-\mathbb{y})^2/\sigma^2] \quad (11)$$

Formula (10) can be changed into: Formula (11) can be substituted into Formula (9).

$$\mathrm{N}\ J_\phi = 2\sum_{i=1}^{N}\sum_{j=1}^{C}\mathbb{w}_{ij}^{m}\left[1-K\left(\mathbb{x}_i,\mathbb{v}_j\right)\right] \quad (12)$$

To find the update formulas for the membership matrix and the novel clustering center for ϕ, estimate the partial derivatives of, , respectively.

$$\mathbb{v}_j = \frac{\sum_{i=1}^{N}\mathbb{w}_{ij}^{m}K\left(\mathbb{x}_i,\mathbb{v}_j\right)\mathbb{x}_i}{\sum_{i=1}^{N}\mathbb{w}_{ij}^{m}K\left(\mathbb{x}_i,\mathbb{v}_j\right)} \quad (13)$$

$$\mathbb{w}_{ij} = \frac{\left(1-K(\mathbb{x},\mathbb{v})\right)^{-1/(m-1)}}{\sum_{j=1}^{C}\left(1-K\left(\mathbb{x}_i,\mathbb{v}_j\right)\right)^{-1/(m-1)}} \quad (14)$$

To determine the clustering of the dataset, IFCM uses Formula (11), which is used to construct the kernel function. Formulas (13) and (14) are then iterated. By using the kernel approach to the FCM technique, IFCM increases the difference among pattern classes, translates input space data to the HD feature space non-linearly, and accomplishes linear aggregation in the HD feature space. IFCM works well with data that has many dimensions. It can, to some extent, overcome the sensitivity of noise data, accomplish accurate clustering of data with varying shape patterns of distribution, and mitigate the dependence of FCM on the data type. However, IFCM suffers from the same drawbacks as FCM: the clustering outcome is dependent upon the beginning value and is susceptible to falling into the local minimum.

MOGWO: A Glow Worm (GW) in the wild will use a special light signal to locate itself and draw in the other sexGWs use this type of flash signal, which is a brief, rhythmic fluorescence, to carry out various activities such as feeding, alertness, relationships, and reproduction. A stochastic optimization technique known as the "GW process" was developed by mimicking the glowing behavior of real-world GWs. The procedure for optimization is made possible by the algorithm, which locates peers in the vicinity of the best firefly by utilizing the brightness properties of GWs.

GWs, each carrying a specific quantity of fluorescence, are distributed around the problem space by the GW algorithm [28]. While is the greatest perceptual radius, every GW has its own perceptual range, . In an attempt to reach maximum efficiency, a GW will hunt for higher fluorescence values in other GWs within its field of view and will proceed accordingly. The GW updates its perceived radius after it moves. The perceptual radius will grow or shrink in accordance with the number of neighbors. Ultimately, one or several extremes will host the majority of the GWs.

Fluorescein of Glowworms

The GW's fluorescent light brightness was determined by GW algorithm based on the present positions OF value. A higher brightness indicates a better location, which in turn indicates a higher OF value. The GWs' brightness indicates their orientation and weighs the advantages and disadvantages of their placement. A GW's luminosity is directly correlated with its fluorescence index. Within its perceived range, the GW is more beautiful the greater its fluorescent value. Formula (15) presents the GW fluorescein update formula:

$$I_i(t+1) = (1-\rho)*I_i(t) + \gamma^* f\left(\mathbb{x}_i(t+1)\right) \quad (15)$$

Among these are the following: is the fitness value of the i GW individual at the time; the fluorescein value of the i GW individual at the time can be denoted as; the volatile coefficient of the fluorescence value can be denoted as ρ , and the enhancement coefficient of the fluorescein value can be denoted as γ.

Collection of neighbours

Although the perceived range influences the attractive action, every GW individual can be drawn to an individual with a high fluorescence value. Within their perceptual range, GWs can only locate other GWs with high fluorescein readings. Formula (16) displays the exact formula of the neighborhood collection inside the perceptual range of the i GW:

$$N_i(t) = \{j: d_{i,j}(t) < R_d^i, (t) < I_j(t)\} \quad (16)$$

Here, the spatial distance between and is represented by the formula.

Transfer probability:

High fluorescein individuals can attract other GW individuals; however, a GW can only search for and move toward other GWs with high fluorescein values that are within its perceptual range. The GW individual uses formula (17) to determine the individual probability of each in the collection of its neighbors, and based on the computation, the GW with the bigger probability value is set as the target traveling direction.

$$P_{ij} \frac{I_j(t)-I_i(t)}{\sum_{m\in N_i(t)} I_m(t)-I_i(t)} \quad (17)$$

Update of the location

In order to update their position in accordance with formula (18), GW individual selects individuals that possess elevated fluorescein values in the perceptual range based on the transition probabilities.

$$\mathbb{x}_i(t+1) = \mathbb{x}_i(t) + Step * \frac{\mathbb{x}_j(t)-\mathbb{x}_i(t)}{\|\mathbb{x}_j(t)-\mathbb{x}_i(t)\|} \quad (18)$$

Here, the GW individual current location can be denoted as and its new location following movement can be represented as, is the mobile step.

Update of neighbourhood radius

The GW's perspective must be adjusted after it moves. Updates to the perceptual radius are associated with the group of neighbors of GW individual at that moment. The perceptual radius should be suitably lowered if a large number of GW individuals having elevated fluorescein values in the perceptual range are present. The perceptual radius must be magnified suitably if the number of GW individuals having elevated fluorescein values in the perceptual range is limited. Formula (19) indicates the GWindividual ‘s neighborhood radius.

$$R_d^i(t+1) = min\{R_s, \max\{0, Range_d^i(t), \beta(n_t - |N_i(t)|)\}\} \quad (19)$$

Is the number of ideal individuals within the perception range of GW individuals, t n is the total amount of exceptional individuals within the perceptual range, and is the maximum perceptual radius of all individuals. β is the variation coefficient of the perceptual radius.

Fitness Function: Let the sample space for the MOGWO-IFCM algorithm, and the dimension vector can be denoted as. Individual GWs are a cluster center, denoted as. Among these, the vectors and have the same dimension. The individual fitness function in GWs is defined as follows for the analysis of every solution (clustering center):

$$f(\mathbb{x}_i) = \frac{1}{1+J_\phi(\mathbb{U},\mathbb{V})} + accuracy \quad (20)$$

The OF found in Formula (6) is represented by in Formula (20). The stronger the clustering effect, the higher the value and the smaller the value.

MOGWO-IFCM Algorithm: The MOGWO-IFCM technique's core concept is as follows: the optimal solution is acquired by first optimizing the initial clustering center of the IFCM technique utilizing the MOGWO technique, and then the ideal solution is produced by employing the IFCM technique as the initial clustering center. The particular steps in the technique are as follows:

Step 1: Set the following initial values: Domain threshold, fluorescence value volatility ρ, variation coefficient of the perceptual radius β, fluorescence value enhancement factor γ, number of clusters, allowed error ε, fuzzy index m, Glowworm Swarm population N, moving step size, and fluorescein concentration.

Step 2: Initiate the GW population. The GW population v_j is a collection of randomly created clusters. Create the GW Swarm population at random.

Step 3: The kernel matrix must be computed.

Step 4: Determine the fitness value of each glowworm and the membership matrix for each glowworm.

Step 5: Determine every GW ‘s fluorescein value,

Step 6: Create neighborhood collection.

Step 7: Next, choose the glowworms that satisfy.

Step 8: Update the glowworms' position.

Step 9: Update glowworms' decision-making radius.

Step 10: Update the glowworms' membership matrix.

Step 11: The population's clustering center is updated. The difference between the two neighboring membership matrices is calculated; if it is $<\varepsilon$, the process ends; if not, skip to Step 10.

MLSTM Classifier

A cell, an input gate, an output gate, and a forget gate make up a typical LSTM [29] unit. The three gates control the flow of data into and out of the cell, and the cell remembers values across arbitrary time intervals. The layout of a conventional LSTM cell and an illustration of the manner in which the gates work is indicated in Figure 2. The input, forget, and output gates are the 3 gates that make up the basic LSTM cell. Each gate contains a point-wise multiplication operation and a sigmoid AF. Hochreiter and Schmidhuber suggested the concept of LSTM. It is made up of a memory cell and a computational unit that takes the role of traditional methods that utilized neurons in the network's hidden layer. The network overcomes some of the obstacles encountered throughout the training stage by using these memory cells. The BRNN design, which utilizes data from the past and future to determine the output at any given time, is then presented in BiRNN (Bidirectional RNNs) by Schuster and Paliwal.

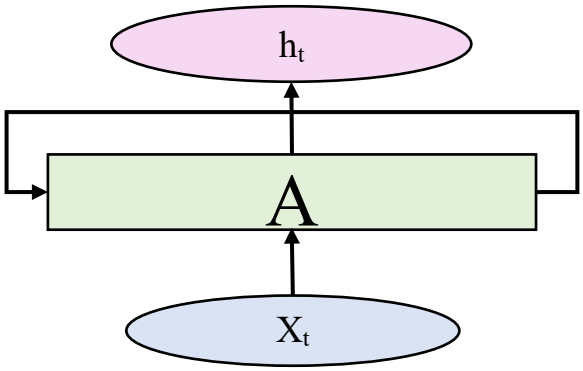

Figure 2. LSTM cell

Figure 3. Traditional LSTM memory cells take the place of neurons in the NN. It is applied to other RNNs to solve the GVP (Gradient Vanishing Problem) [30]. Short dependencies and a distinct set of weights for remembering and forgetting outputs are employed to train the RNN, which is incapable of remembering longer sequences.

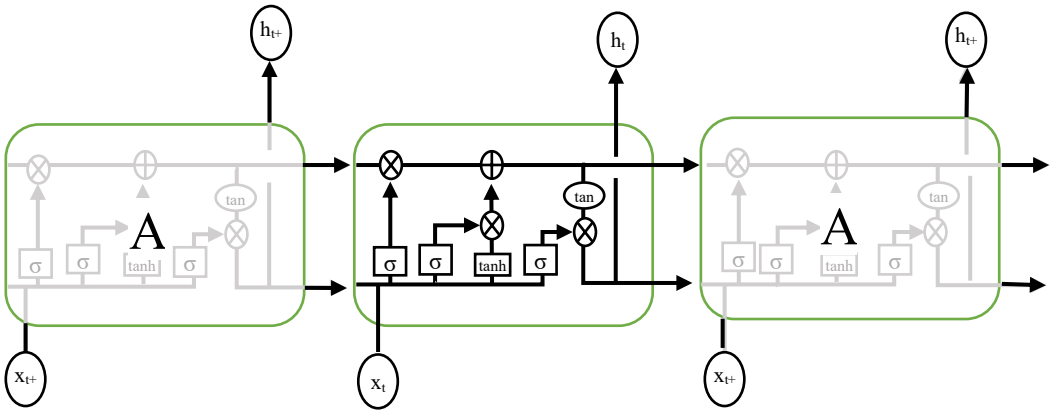

Figure 3. LSTM model

The output of an LSTM unit is after it has read an input and depends on a previous output. It consists of an input gate, an output gate (ot), a forget gate, and a memory cell. The following functions are carried out by each LSTM cell:

To determine which data should be removed from the memory vector, use the current input and the prior hidden state. This may be expressed as follows: where is a collection of weights and is a bias ate forget gate.

A matrix that allows for the updating of a particular piece of information in is created using t and. When using a bias gate forget gate, is bias

To obtain the necessary information, utilize and;

To sum up, use the formula to combine the new and old data.

It is evident that this algorithm will be trained to distinguish between data that has to be remembered, conserved, or retained utilizing SGD (Stochastic Gradient Descent). In Figure. 4, the MLSTM cell is displayed. Describe the LSTM network's Forward Propagation (FP) mechanism. The forgetting threshold must be calculated as the initial step in the LSTM network's FP. This threshold establishes which input data will be ignored and not have an impact on subsequent time steps. In specifics, as indicated in equation (21) the time vector is utilized as an input parameter of the forget gate after the time delay from the time step t-1 and the time step t has been reduced to produce a 3-dimensional vector.

$$f_t = \sigma(W_f[\mathbb{h}_{t-1}, \mathbb{x}_t] + bias_f) \quad (21)$$

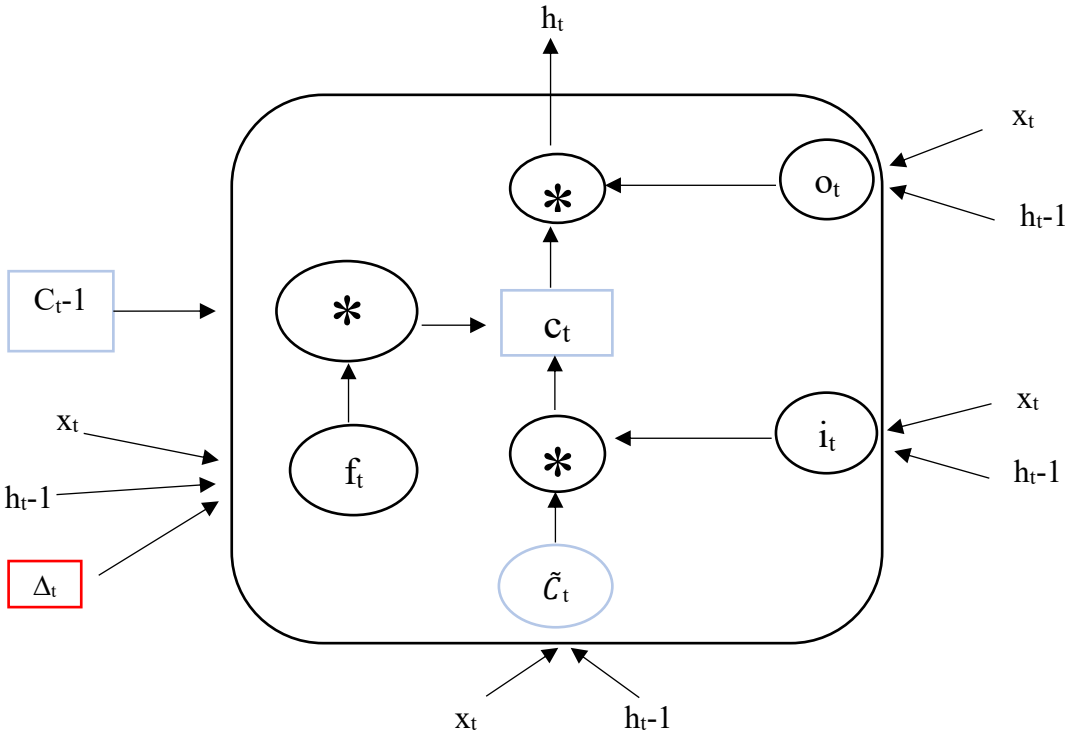

Figure 4. MLSTM model

$$f_t = \sigma(W_f[\mathbb{h}_{t-1}, \mathbb{x}_t] + P_{f_{\Delta_{t-1:t}}} + bias_f \quad (22)$$

After the time gap between time slices is smoothed, a vector is represented in equation (22), where is found. Equation (23) provides the smoothing formula.

$$\mathrm{p}_{\Delta_{t-1:t}} = \left(\frac{\Delta_{t-1:t}}{60}, \left(\frac{\Delta_{t-1:t}}{180}\right)^2, \left(\frac{\Delta_{t-1:t}}{365}\right)^3\right) \quad (23)$$

The time interval is represented in days in equation (23), using as its unit. So selected two months as the denominator, followed by 1⁄2 a year and one year as patients rarely rehospitalize in the same month. This puts the vector $\Delta t{-}1{:}\ t$ within a suitable range. In order to manage the memory effect caused by the irregular time interval, the connection weight parameter that corresponds to the time interval vector can be denoted as, must be improved during training. The data stored in the cellstate is determined by the second stage of FP. The previous cell state must be updated after creating a temporary state. Equations (24) and (25) display the formula.

$$\tilde{C}_t = tan\mathbb{h}\,(W_C[\mathbb{h}_{t-1}, \mathbb{x}_t + bias_c) \quad (24)$$

$$C_t = f_t * C_{t-1} i_t * \tilde{C}_t \quad (25)$$

Here, the temporary state's connection weight and offset are expressed. There are new candidate values in the temporary state. The status data from the previous time step is represented by. The time step following the update is represented by. Equation (26), which illustrates the final network output, is determined by the third stage of FP.

$$\mathbb{h}_t = o_t * tan\mathbb{h}\,(C_t) \quad (26)$$

here

$$o_t = \sigma(w_f.\mathbb{x}_t + bias_i + W_{ho}\mathbb{h}_{t-1} + bias_h) \quad (27)$$

Here, the current hidden state is denoted by, and the input for the subsequent time step is made up of and

Experimental Results and Discussion

The findings of experiments employing the Hybrid Feature selection MOGWO-IFCM method are presented. Two datasets, such as WDBC and Hepatitis from UCI's dataset, were used to assess the proposed MOGWO-IFCM and HIPSO-KM with existing methods such as MBACO and MOPSO.

Dataset Description

Wisconsin Diagnostic Breast Cancer (WDBC): When cells in the breast proliferate uncontrollably, it can lead to BC (Breast Cancer). BC comes in various forms. BC kind is determined by which breast cells develop cancerous cells. Tumors are classified as either benign or malignant using the WDBC dataset, which was acquired from the University of Wisconsin Hospital. Field 2 prediction is as follows: B = benign, M = malignant, and all 30 input features can be utilized for separating the sets linearly.

The WDBC Data Set (wdbc.data, wdbc.names) from the UCI ML

Repository https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic).

Hepatitus Disease Dataset: The UCI ML repository provided the data set. After deleting the observations with missing values, this data set is made up of 80 observations out of a total of 155 observations. Thirteen of the nineteen traits and qualities are binary, and the remaining six are discrete-valued. The observations are divided into two categories: classes that survive and those that die. There are "67" live observations and 13 "die" observations in the class.

This data set is available on: https://archive.ics.uci.edu/ml/datasets/Hepatitis.

Performance Evaluation

The True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) rates were first determined and then used to construct various performance indicators. The following parameters are evaluated in this work, 1. Precision, which is the fraction of relevant retrieved observations, 2. Recall, which is the proportion of relevant instances recovered, 3. F-measure is acquired by integrating precision and recall, and 4. Accuracy, which is a fraction of accurately predicted instances in comparison to all predicted instances.

The accurately detected positive instances to all of the expected positive instances are called precision.

$$\text{Precision} = TP/TP+FP \tag{28}$$

Recall is the accurately detected positive instances to the overall observations.

$$\text{Recall} = TP/TP+FN \tag{29}$$

F - measure is the weighted average of Precision as well as Recall. Therefore, it considers both FPs and FNs.

$$\text{F1 Score} = 2*(\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \tag{30}$$

Accuracy is assessed as below:

$$\text{Accuracy} = (TP+FP)/(TP+TN+FP+FN) \tag{31}$$

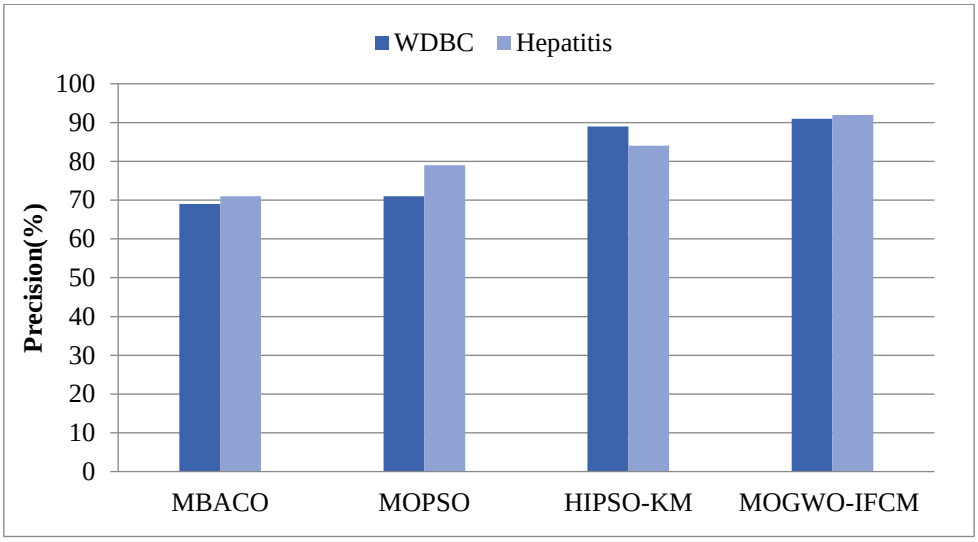

Figure 5. Precision comparison outcomes among the suggested and current methods for classifying the health data

The figure 5. shows the precision outcomes of the newly suggested approach and the existing techniques for categorizing health information. The findings indicate that the MOGWO-IFCM (91%-WDBC and 92%-Hepatitis) is effective in identifying the health information, and also, the important characteristics do not impact the accuracy of the combined features transformation. In the proposed algorithm, the effects of outliers and missing data are reduced in determining optimal cluster centres using MOWO. In each iteration, the MOGWO-IFCM procedure is employed for estimating the "goodness" of the subsets developed by ants as solutions. The suggested system is effective by the execution, speed and anticipating trapping in local minima.

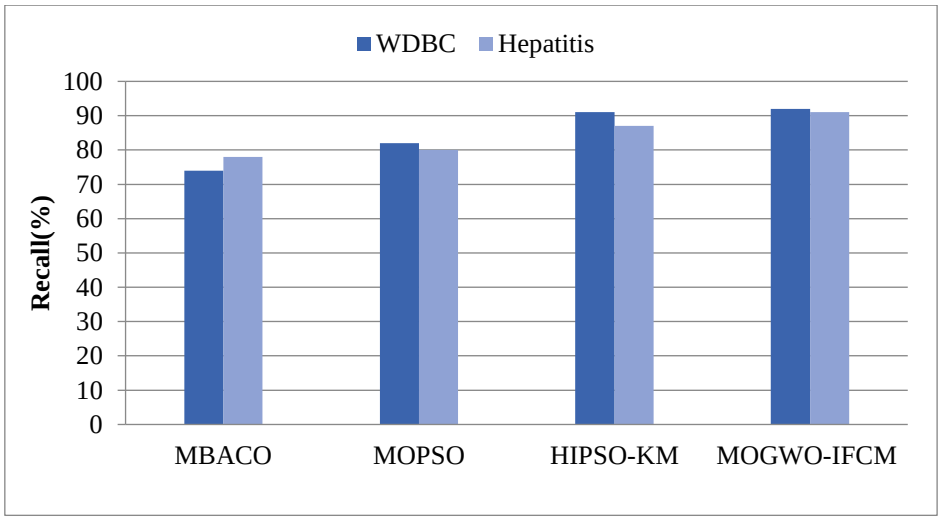

Figure 6. Recall results of the newly suggested and traditional approaches for classifying the health data.

The figure 6. depicts the results of the recall of the newly suggested approach and the conventional approaches. From the above outcomes, the suggested technique is highly efficient in all applications. In addition, the findings indicate that the suggested approach has higher recall rates of 92% and 91% for WDBC and Hepatitis data, respectively. In contrast, the traditional approaches such as HIPSO-KM, MOPSO, and MBACO approaches yield the recall rate of only 91%, 82% and 74% for WDBC data and 87%, 80% and 78% for Hepatitis data, respectively.

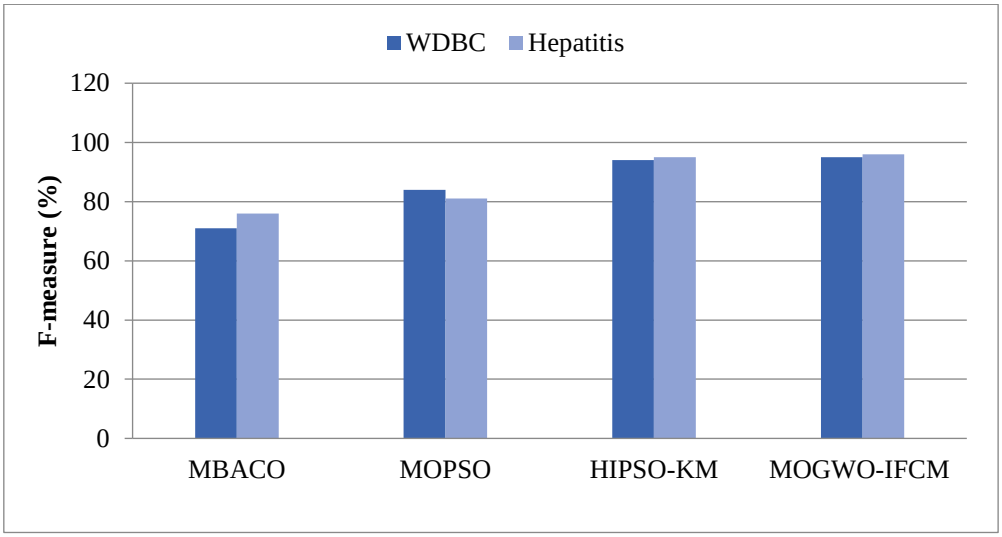

Figure 7. F-measure results of the newly suggested and traditional approaches for classifying the health data

The figure 7. represents the results of F-measure values of the newly suggested approach and the conventional approaches. The findings indicate that the suggested MOGWO-IFCM (95%-WDBC and 96%-Hepatitis) and HIPSO-KM method yield high F-measure values than the traditional approaches. It combined supervised IFCM algorithm and MOGWO technique for FS that will detect the relative significance of the IFCM error rate and the amount of FS. It is efficient, and very faster than the algorithms that evaluate feature subsets with classification accuracy.

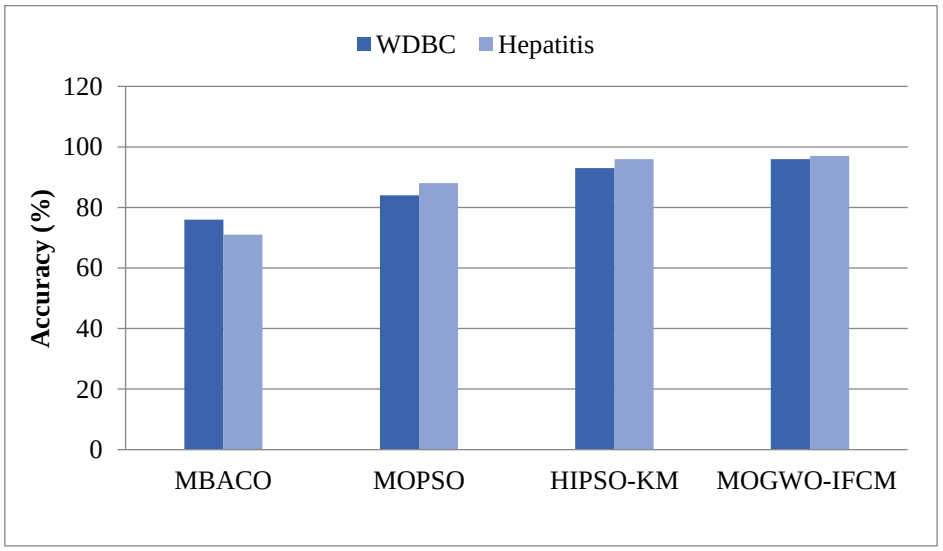

Figure 8. Accuracy results of the newly suggested and traditional approaches for classifying the health data

The figure 8. depicts the results of accuracy values of the newly suggested approach and the conventional approaches. In addition, the proposed method's average accuracy of classification (in 97%-Hepatitis and 96%-WDBC) compared to existing approaches using MOPSO and MBACO classifiers over 10 separate runs are shown. As the results show, the suggested method has higher accuracies than existing methods. In addition, the rank of feature selection approaches is detected based on achieved accuracy for a specific dataset. MOGWO is quicker in finding optimal solution. When a sub-optimal solution is found for a HD dataset, the PSO and ACO are unable to locate a better one; however, the ACO can continue searching in the feature search space until the optimal solution is found. In addition, GA doesn't have recall but the MOGWO system has recall. In ACO, the ants that were optimal in the past, are pulled to return towards, so data of good solutions is reserved, that it is an advantage of MOGWO-IFCM.

CONCLUSION AND FUTURE WORK

This study created the hybrid FS MOGWO-IFCM approach, that combines the FCM technique and chaos theory with the CSO technique. The global optimization technique is employed by the suggested MOGWO-IFCM techniques to prevent local minima from becoming trapped, and chaos theory is used to address the MOGWO technique's inability to converge, which caused the random variables to behave in a chaotic manner rather than according to a Gaussian distribution. Additionally, the FCM is applied for assessing crows in HDspace and manage uncertainty. The suggested MOGWO-IFCM technique made use of a variety of unique chaotic map types in order to improve effectiveness, accelerate convergence, and maintain a balance among the rate of exploration and exploitation. One of the hard FS challenges in clinical diagnostics is employed to test the MOGWO-IFCM.

The best features are given as input to a classifier like MLSTM to classify normal and abnormal. In comparisons, selected features yield better result in MLSTM with the high accuracy. Two separate medical data sets (WDBC and hepatitis) were employed for assessing the performance, and each set was examined using a distinct set of evaluation metrics (precision, recall, f-measure, and accuracy). MOGWO-IFCM operates better than alternative methods like HIPSO-KM, MOPSO, and MBACO, according to the experimental data. Furthermore, the outcomes demonstrate that the MOGWO-IFCM technique's classification performance may be greatly improved by combining singer, Gauss, and tent maps. Future research will concentrate on using the chaotic version in conjunction with parallel bioinspired optimization for a variety of applications. In order to guarantee stability, the MOGWO-IFCM will also assess on real-world issues.

REFERENCES

[1] Wang XD, Chen RC, Yan F, Zeng ZQ, Hong CQ. Fast adaptive K-means subspace clustering for high-dimensional data. IEEE Access. 2019 Mar 22; 7:42639-51. https://doi.org/10.1109/ACCESS.2019.2907043

[2] Nizam MU, Zaneta SA, Basri FA. Machine Learning based Human eye disease interpretation. International Journal of Communication and Computer Technologies (IJCCTS). 2023;11(2):42-52.

[3] Jaiswal JK, Samikannu R. Application of random forest algorithm on feature subset selection and classification and regression. In2017 world congress on computing and communication technologies (WCCCT) 2017 Feb 2 (pp. 65-68). Ieee. https://doi.org/10.1109/WCCCT.2016.25

[4] Singhal P, Yadav RK, Dwivedi U. Unveiling patterns and abnormalities of human gait: a comprehensive study. Indian Journal of Information Sources and Services. 2024;14(1):51-70.

[5] Snousi HM, Aleej FA, Bara MF, Alkilany A. ADC: Novel Methodology for Code Converter Application for Data Processing. Journal of VLSI circuits and systems. 2022 Sep 20;4(2):46-56. https://doi.org/10.31838/jvcs/04.02.07

[6] Blum AL, Langley P. Selection of relevant features and examples in machine learning. Artificial intelligence. 1997 Dec 1;97(1-2):245-71. https://doi.org/10.1016/S0004-3702(97)00063-5

[7] Lakkaraju H, Kamar E, Caruana R, Horvitz E. Identifying unknown unknowns in the open world: Representations and policies for guided exploration. InProceedings of the AAAI Conference on Artificial Intelligence 2017 Feb 13 (Vol. 31, No. 1). https://doi.org/10.1609/aaai.v31i1.10821

[8] Soofi AA, Awan A. Classification techniques in machine learning: applications and issues. Journal of Basic & Applied Sciences. 2017 Jan 5; 13:459-65.

[9] Chen RC, Dewi C, Huang SW, Caraka RE. Selecting critical features for data classification based on machine learning methods. Journal of Big Data. 2020 Jul 23;7(1):52.

[10] Alhassan S, Abdul-Salaam G, Micheal A, Missah YM, Ganaa ED, Shirazu AS. CFS-AE: Correlation-based Feature Selection and Autoencoder for Improved Intrusion Detection System Performance. Journal of Internet Services and Information Security. 2024 Mar;14(1):104-20. https://doi.org/10.58346/JISIS.2024.I1.007

[11] Ying S, Xue-Ying Z. Characteristics of human auditory model based on compensation of glottal features in speech emotion recognition. Future Generation Computer Systems. 2018 Apr 1; 81:291-6. https://doi.org/10.1016/j.future.2017.10.002

[12] Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas MJ. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore geology reviews. 2015 Dec 1; 71:804-18. https://doi.org/10.1016/j.oregeorev.2015.01.001

[13] Zhu F, Jiang M, Qiu Y, Sun C, Wang M. RSLIME: an efficient feature importance analysis approach for industrial recommendation systems. In2019 International Joint Conference on Neural Networks (IJCNN) 2019 Jul 14 (pp. 1-6). IEEE. https://doi.org/10.1109/IJCNN.2019.8852034

[14] Bhuyan HK, Kamila NK. Privacy preserving sub-feature selection in distributed data mining. Applied soft computing. 2015 Nov 1; 36:552-69. https://doi.org/10.1016/j.asoc.2015.06.060

[15] Ravisankar P, Ravi V, Rao GR, Bose I. Detection of financial statement fraud and feature selection using data mining techniques. Decision support systems. 2011 Jan 1;50(2):491-500. https://doi.org/10.1016/j.dss.2010.11.006

[16] Rahmat A, Nurrahman AA, Pramono SA, Ahmadi D, Firdaus W, Rahim R. Data Optimization using PSO and K-Means Algorithm. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications. 2023;14(3):14-24. https://doi.org/10.58346/JOWUA.2023.I3.002

[17] Devi SG, Sabrigiriraj M. Swarm intelligent based online feature selection (OFS) and weighted entropy frequent pattern mining (WEFPM) algorithm for big data analysis. Cluster Computing. 2019 Sep;22(Suppl 5):11791-803. https://doi.org/10.1007/s10586-017-1489-9

[18] Abualigah L, Dulaimi AJ. A novel feature selection method for data mining tasks using hybrid sine cosine algorithm and genetic algorithm. Cluster Computing. 2021 Sep;24(3):2161-76. https://doi.org/10.1007/s10586-021-03254-y

[19] Muralidharan J. Wideband Patch Antenna for Military Applications. National Journal of Antennas and Propagation. 2020 Jan 1;2(1):25-30. https://doi.org/10.31838/NJAP/02.01.05

[20] Tang J, Zhang J, Yu G, Zhang W, Yu W. Multisource latent feature selective ensemble modeling approach for small-sample high-dimensional process data in applications. IEEE Access. 2020 Aug 11; 8:148475-88. https://doi.org/10.1109/ACCESS.2020.3015875

[21] Gong L, Xie S, Zhang Y, Wang M, Wang X. Hybrid feature selection method based on feature subset and factor analysis. IEEE Access. 2022 Nov 17;10:120792-803. https://doi.org/10.1109/ACCESS.2022.3222812

[22] Fan W, Bouguila N, Ziou D. Unsupervised hybrid feature extraction selection for high-dimensional non-gaussian data clustering with variational inference. IEEE Transactions on Knowledge and Data Engineering. 2012 May 15;25(7):1670-85. https://doi.org/10.1109/TKDE.2012.101

[23] Huang Z, Yang C, Zhou X, Huang T. A hybrid feature selection method based on binary state transition algorithm and ReliefF. IEEE journal of biomedical and health informatics. 2018 Sep 28;23(5):1888-98. https://doi.org/10.1109/JBHI.2018.2872811

[24] Zhang X, Mei C, Li J, Yang Y, Qian T. Instance and feature selection using fuzzy rough sets: A bi-selection approach for data reduction. IEEE Transactions on Fuzzy Systems. 2022 Oct 25;31(6):1981-94. https://doi.org/10.1109/TFUZZ.2022.3216990

[25] Muthulakshmi P, Parveen M. Z-Score Normalized Feature Selection and Iterative African Buffalo Optimization for Effective Heart Disease Prediction. International Journal of Intelligent Engineering & Systems. 2023 Jan 1;16(1). https://doi.org/10.22266/ijies2023.0228.03

[26] Shah SM, Batool S, Khan I, Ashraf MU, Abbas SH, Hussain SA. Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis. Physica A: Statistical Mechanics and its Applications. 2017 Sep 15; 482:796-807. https://doi.org/10.1016/j.physa.2017.04.113

[27] Sefidian AM, Daneshpour N. Missing value imputation using a novel grey based fuzzy c-means, mutual information-based feature selection, and regression model. Expert Systems with Applications. 2019 Jan 1; 115:68-94. https://doi.org/10.1016/j.eswa.2018.07.057

[28] Aljarah I, Ludwig SA. A new clustering approach based on glowworm swarm optimization. In2013 IEEE congress on evolutionary computation 2013 Jun 20 (pp. 2642-2649). IEEE. https://doi.org/10.1109/CEC.2013.6557888

[29] Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. Neural computation. 2019 Jul 1;31(7):1235-70. https://doi.org/10.1162/neco_a_01199

[30] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena. 2020 Mar 1; 404:132306. https://doi.org/10.1016/j.physd.2019.132306

[31] Fong S, Wong R, Vasilakos AV. Accelerated PSO swarm search feature selection for data stream mining big data. IEEE transactions on services computing. 2015 Jun 1;9(1):33-45. https://doi.org/10.1109/TSC.2015.2439695

[32] Wang D, Nie F, Huang H. Feature selection via global redundancy minimization. IEEE transactions on Knowledge and data engineering. 2015 Apr 28;27(10):2743-55. https://doi.org/10.1109/TKDE.2015.2426703

[33] García-Escudero LA, Gordaliza A, Matrán C, Mayo-Iscar A. A review of robust clustering methods. Advances in Data Analysis and Classification. 2010 Sep;4(2):89-109. https://doi.org/10.1007/s11634-010-0064-5

[34] Chen RC. Using Deep Learning to Predict User Rating on Imbalance Classification Data. IAENG International Journal of Computer Science. 2019 Mar 1;46(1).

[35] Varshavardhini S, Rajesh A. An Efficient Feature Subset Selection with Fuzzy Wavelet Neural Network for Data Mining in Big Data Environment. Journal of Internet Services and Information Security. 2023 May;13(2):233-48. https://doi.org/10.58346/JISIS.2023.I2.015

[36] Cenggoro TW, Mahesworo B, Budiarto A, Baurley J, Suparyanto T, Pardamean B. Features importance in classification models for colorectal cancer cases phenotype in Indonesia. Procedia Computer Science. 2019 Jan 1; 157:313-20. https://doi.org/10.1016/j.procs.2019.08.172

[37] Camgözlü Y, Kutlu Y. Leaf image classification based on pre-trained convolutional neural network models. Natural and Engineering Sciences. 2023 Dec 1;8(3):214-32. https://doi.org/10.28978/nesciences.1405175

[38] Venkatasalam K, Rajendran P, Thangavel M. Improving the accuracy of feature selection in big data mining using accelerated flower pollination (AFP) algorithm. Journal of medical systems. 2019 Apr;43(4):96. https://doi.org/10.1007/s10916-019-1200-1