Enabling A Better Hybrid Learning Algorithm For Enhancing the Classification Accuracy of Datamining Applications Improve the Classification Accuracy

Dr.N. Elavarasan
Assistant Professor, Department of Computer Applications, Srinivasan College of Arts and Science, Perambalur. elavarasann1975@gmail.com

S. SIVARAJ
Assistant professor, Department of Electronics and Communication Engineering, Roever Engineering College, Perambalur.

Abstract

Data Analytics and Data Mining are two significant pillars mainly used in Data Science. The main objective of the data science domain is to provide an accurate output to the customer's requirements. It is an extensive process of analyzing the raw data and mining the user-required pattern from a large data set. Recent data mining applications need improvement, so the data mining industries are motivated to enhance the quality of data and the efficiency of data analysis methods. Several earlier research works have been proposed for improving the mining accuracy in various real-time applications. But their outputs are comparatively less due to inefficiency in clustering, classification, and mining. Thus, this paper makes the user choose a better hybrid algorithm, improving the mining efficiency by preprocessing, clustering, and type. It also mainly focused on reducing time and computation complexity by applying dimensionality reduction. The hybrid algorithm is designed by integrating MCA PCA with FCM, RBFN, and MLP. The best method is to compare error, accuracy, and time complexity.

Keywords: Data Mining, Hybrid Algorithm, FCM, MCA, PCA, RBFN, MLP.

Nomenclature

KNN: K-Nearest Neighbour
RF: Random Forest
FCM: Fuzzy C Means Clustering
MCA: Multiple Correspondance Analysis
PCA: Principal Component Analysis
RBFN: Radial Basis Function Networks
MLP: Multilayer Perceptron

DOI Number: 10.14704/nq.2022.20.9.NQ44642 **Neuro Quantology 2022; 20(9):5494-5504**

Introduction

The development of software and its usability has managed to gain a significant population who depends on it daily. It also eases several business models by developing online portals. These online portals and other websites can reach several people and advertise themselves. However, such a vast development is also the downside because of the management of the available big data. Managing a large amount of data is the most tedious job for software companies. While scaling the platform, it is necessary to know the sentiments and nature of the users. This vast data needs to be processed effectively, and the essential information needs to

be extracted from the big data. This extensive data needs to be extracted from the big data. For the extraction of such valuable data, data mining is used.

Data mining involves algorithms that analyze the data pattern and efficiently extract the data. Data mining uses machine learning algorithms to detect the patterns and logic behind the data. They are also used to sort and filter the data. Data mining algorithms provide the space for fine-tuning to suit specific business purposes to extract data. It also reduces the cost of holding such a massive amount of data, effectively handling business, and increasing sales. Apart from business applications, it is widely used for detecting fraud, mass behavior, and security threats. Data mining has enabled the conversion of raw data into useful information. Data mining helps online entities scale their business and automate their workloads. When data mining is combined with data analysis, one can envision the scopes of the data. However, developments in data mining are too slow and need relevant insights for extracting the data. Most of the online entities are using these methods to manage the data that is available with them. It also enables effective decision-making from raw and unusable data. Data mining depends on three essential criteria. One is the data collection, then data warehousing, and finally computation of the data. As said before, data mining looks for patterns and logic in a massive set of data. It uses these patterns and logic to extract meaningful data from the vast data. Data mining involves the following steps.

- Collection of data
- Loading the data to the databases
- Stored data is then managed locally or in the cloud. Several paid cloud services offer storage services.
- Professionals from that specific field then view the stored data and determine how to organize the data.
- The data mining algorithm sorts the data stored based on the instructions given by the professionals.

The professionals who make the decisions are data scientists and business analysts. Both these professions are different, but their combined input is needed for making definite rules and defining the parameters to sort the data. The data sequence is utilized to reduce the amount of data, and unwanted data can be excluded from the analysis. However, the data scientists maintain the key parameters that influence the quality of the data. The data patterns are identified by finding the relationship between the individual elements in the data. The association is sequential, serial, or fit into any arithmetic series or sequences. Data analysis becomes easy if they fit into any of those sequences. Otherwise, several other measures need to be followed, like supervised, unsupervised, or semi-supervised learning. These learning methods are incorporated through deep learning algorithms.

Supervised learning requires manual interventions, and unsupervised learning is fully automatic. Semi-supervised learning uses both manual and automatic functions for learning. These learning methods suit specific necessities. Mostly, all the criteria will fit into the category mentioned above. But the efficiency varies with every technique used. Finally, the sorted data can be presented in user-readable formats like graphs and spreadsheets. Data mining algorithms generally classify the data and group them into clusters. However, it needs to be cleaned before grouping the data. They may also contain repetitions, missing values, outliers, and corruption. These clusters have data collected from the parent dataset, which is connected through criteria. These connections or the homogeneity of the data are utilized to form groups. For example, several customers may come for specific foods at specific times in a restaurant. It may depend on the season or climate, timing, unique dishes, and several other criteria. These criteria mentioned above can be considered a connection between the customers coming for that criteria and converted into clusters. These clusters contain the details of the customers coming for that specific criteria. So, for each measure, a cluster can be created. Then, these clusters are compared with each other to have an overall view of their sales, customer needs, timings, and unique dishes. These data can be grouped to make practical conclusions like dishes with more sales at specific times and particular people. The age group of the people coming to the restaurant can be sorted out to find the majority of the age group of the customers. These are just a few conclusions based on the parameters mentioned. But, in real-world business, there will be a large no of criteria and parameters to be managed. It cannot be

controlled and analyzed manually. Data mining takes care of these parameters and can make effective solutions out of the garbage data.

Warehousing is also an essential process in data mining. Warehousing is nothing but centralizing the data available for data mining. Data needs to be stored in a single place to process it efficiently. The performance and time consumption may increase if the information is decentralized. So, it is essential to maintain a single and accessible area to maintain our data.

Another thing is the number of resources required to maintain the data. Because storing a large amount of data requires more storage and higher data transfer rates for storing and retrieval. Local storage can be handy for processing smaller amounts of data but has limitations over scaling. It is limited to our systems, and it isnot easy to access it from different areas at different times. So, the cloud is the only solution to this problem. Cloud provides access at any place, at any time. It also provides higher data transfer rates and heavy resources for the vast data. They also help in processing data as the programs and algorithms used can take advantage of the features provided by cloud computing services.

Data mining is best suited for social media companies because it involves a large user base that interacts with the world privately and publicly. But, it is tedious as the data obtained will be humongous and multifaceted. However, if the data is effectively processed, it can lead to a great understanding of the people's nature, preferences, necessity, and capacity. This information can increase sales. By knowing the people's specific needs, marketing targeted those belonging to that demography. It can also help effective marketing. They have a vast scope in terms of positively and negatively influencing people. They can make compelling predictions for social events like elections and manipulate people towards specific trends. It has led to several controversies between social media and the government using personal details for mass manipulation. However, leaving aside such disputes, from a productive standpoint, data mining has excellent scope in e-commerce and other social media platforms.

Contribution of the Paper

The paper's contribution is to provide an option to the developer to choose a better hybrid algorithm created by integrating various state-of-the-art methods. This paper discusseddifferent state-of-the-art techniques used in data mining applications to improve prediction accuracy. Some algorithms include RF, KNN, RBFN, MLP, MCA, PCA, and FCM to create hybrid algorithms to enhance data mining accuracy. The combination of MCA+FCM+MLP is selected as a better hybrid algorithm based on their performance comparison. The learning, clustering, and classification efficiencies are obtained from three advanced methods such as MLP, FCM, and MCA methods. The overall process carried out using all the deep learning methods is illustrated in Figure-1.

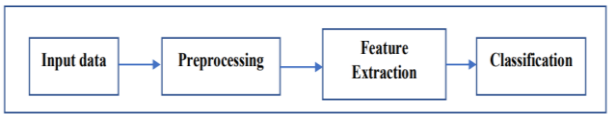


Figure-1. Overall Process Carriedout Using Learning Methods

Literature Survey

Many methods are used to diagnose and classify the data for heart diseases. Many authors combine classification and cluster methods to achieve above 75% classification accuracy.H. Hwang et al. (2006) proposed an extension of multiple correspondence analysis. The author created a unified framework by adjoining various correspondence analysis and k-means methods to improve efficiency.One of the authors in Abdi and Valentin (2007) has been extended the correspondence analysis (CA) to multiple correspondence analysis (MCA). It is used to identify the relationship patterns of various dependent variables. Here the variables belong to categorical. MCA is similar to the following methods: optimal scaling, optimal or appropriate scoring, homogeneity analysis, dual scaling, quantification method, and scalogram analysis. Hence, MCA is obtained with the help of formal correspondence analysis. It is represented on an indicator matrix.

Kurt et al. (2008) compared various classification techniques' performance to predict coronary artery disease (CAD). These authors also discuss the logistic regression (LR), multilayer perceptron (MLP), classification and regression

tree (CART), self-organizing feature maps (SOFM), and radial basis function (RBF). ROC curve, Hierarchical Cluster Analysis (HCA), and Multi-dimensional Scaling (MDS) are used to evaluate the performance of the classification techniques. For predicting CAD, the four classification techniques such as MLP, CART, LR, and RBF are better to predict than SOFM.

H. Hwang et al. (2010) introduced multiple correspondence analysis (MCA), a tool to explore the inter-dependencies between the variables. In addition, cluster-level heterogeneity is extended with MCA. Then, MCA merged with fuzzy k-means to obtain a map of both variable-level and cluster-level structures. Kumari and Godara (2011) investigated several data mining classification techniques are as follows: RIPPER classifier, artificial neural networks (ANNs), Support Vector Machine (SVM). For measuring the unbiased estimation, a 10-fold cross-validation method is proposed. The RIPPER, ANN, SVM, and Decision Tree outperform the other models in terms of accuracy. SVM helps detect cardiovascular disease with high precision and less error rate among the four classification models.

Bhatla and Jyoti (2012) aimed to analyze the different data mining techniques to predict heart disease. Several classifiers are introduced to detect heart disease. Hence, the neural networks and 15 attributes show 100% accuracy. Genetic algorithms and feature subset selection are used to achieve accuracy through decision trees. The decision tree yields 99.62% accuracy.

V. D. Hunt (2012) describes Artificial Intelligence (AI). AI includes Natural-Language Processing (NLP), problem-solving and planning, intelligent robots and machine vision, and Expert systems. An example of an Expert system programming tool is AGE. It contains two predefined components configurations such as blackboard and back chain framework. Blackboard is used to form hypotheses, whereas back chain framework utilizes the backward chaining production rules to build programs.

J.C. Bezdek (2013) suggested information and uncertainty in a new structure of mathematics in pattern recognition. The author introduced Fuzzy sets and their properties. Then it also partitioned the finite data sets and listed the differences between conventional and fuzzy partitions. Finite data sets include complicated and ambiguous relations. M. A. Jabbar et al. (2013) describe medical data mining. Associative classification combines both association rule mining and variety. For prediction, a set of high-quality rules are used. These authors proposed a decision support system to predict the risk score. Thus, a lazy associative classification helps predict heart disease to enhance a patient's lifetime.

Ziasabounchi implements a clustering approach, and Askerzade (2014) helps diagnose heart disease patients. It is based on Principal Component Analysis (PCA). With the help of the PCA reduction method, the original dataset is minimized. Both the fuzzy C-means and K-means algorithms are used to apply the reduced dataset to the clustering methods. Hence, the clustering algorithms and PCA are combined to diagnose heart disease patients more effectively. R. El-Bialy et al. (2015) obtained different data sets of the CAD disease from the input of the machine learning analysis. It will minimize the missing, incorrect, and inconsistent data problems. The resulted tree is compared with the fast decision tree and the pruned C4.5 tree. Thus, the collected dataset provides a classification accuracy of 78.06%.

R. D. F. Dalvi et al. (2017) deployed a new approach to classify the ECG heartbeats. It is widely used in telecardiology services. It is mainly focused on reducing the classification complexity. For removing the baseline, a preprocessing step is introduced. In addition, a QRS complex detection algorithm, which contains the basic information of the heartbeats, is designed. A set of features are used to segment the ECG. Then, principal component analysis helps to minimize the size of the feature vector. These reduced feature vectors were given as input to ANN. Hence, the classification performance obtains 96.97% accuracy.

Some neurodegenerative disorders are Parkinson's disorder (PD) and Lewy body dementia. V. Bhateja et al. (2018) modeled the 3D structure of SNCA. They are as follows: Deriving the 3D structures of Phyto-compounds from several online databases, virtual screening of Phyto-compounds against the SNCA receptor, and selecting the best ligand. A. U. Haq et al. (2018) deployed a machine-learning-based diagnosis system to predict heart disease with the help of heart disease. Every classifier computes optimistic receiver curves and the area

of the curves. A complete set of reduced features was validated to determine the classifier's performance. Hence, it helps to detect heart disease effectively.

M. Kumar et al. (2018) deployed four classification algorithms like Naïve Bayes, Multilayer Perceptron, Decision Table, and Random Forest. It depends on the integrated measurements of the dataset. These algorithms are constructive to determine the patient belongs to tested positive or tested negative. To achieve robustness, S. Chakraverty et al. (2019) implemented soft computing, which is reliable. It consists of fuzzy set theory and neural networks. Fuzzy sets represent vague or uncertain sets. It includes the membership function that is in between 0 and 1. For imitating the human brain, ANN is proposed with the help of learning processes. Fuzzy set, ANN, and interval arithmetic are implemented in C/ MATLAB programming codes.

To enhance diagnostic accuracy, T. L. Le (2019) developed a fuzzy c-means clustering interval type-2 cerebellar model articulation neural network (FCM-IT2CMANN). It adjoins both the IT2CMANN classifier and the fuzzy c-means algorithm. The classifier is used to predict whether the sample is either healthy or diseased after the data is divided and grouped using the classifier. The Lyapunov stability theory improves system convergence. Hence, the proposed classifier's performance has been effectively increased.

H. Riahi-Madvar et al. (2019) utilized soft computing methods. Five models were used to predict the location and dimension of the scour pit. Five models are as follows: multilayer perceptron (MLP) neural network, radial basis functions (RBF) neural network, adaptive neuro-fuzzy inference systems (ANFIS), multiple linear regression (MLR), and multiple non-linear regression (MNLR). Hence, the output of MLP and ANFIS outperforms the other models while predicting the scour geometry accurately. S. Deng (2020) used the clustering algorithmto improve the efficiency of the clustering process. It focused on handling the three common clustering challenges. Initial bias, long operational time, and noise problems are some clustering challenges. For addressing the clustering problems, meta-learning and adversarial learning are employed.

Wiharto and Suryani (2020) compare the k-means algorithm and fuzzy c-means (FCM). These algorithms are compared to analyze the retinal blood vessel segmentation performance. It consists of preprocessing, segmentation, and performance analysis. Preprocessing includes the green channel method, median filter, and Contrast-limited adaptive histogram equalization (CLAHE). Clustering, thresholding, and the region of interest (ROI) come under the segmentation. The mean and the median are used in the thresholding process. Finally, the performance analysis uses the performance parameters of the AUC.

Classification Methods

Artificial Intelligence algorithms provide a subset of machine learning and deep learning algorithms

Random Forest Algorithm

One of the popular classification methods is the Random Forest (RF) method. It comprises more decision trees to take the right decisions at each level of data analysis. The final result of the RF method compares the results obtained from more decision trees. The functionality of the RF method is illustrated in Figure-2. Thus the performance of the RF method is always better. The RF method combines the features of bootstrap aggregation and randomization methods. Each tree (**T**) has **N** number of leaves, segment the entire feature vector into **S** number of feature space. ($\boldsymbol{F_s}$), where

$$\mathbf{1} \leq \boldsymbol{s} \leq \boldsymbol{S} \qquad \mathbf{(1)}$$

The classification function f(x) is written as

$$\boldsymbol{f(x)} = \sum_{\boldsymbol{s=1}}^{\boldsymbol{S}} \boldsymbol{c_s} \prod(\boldsymbol{x}, \boldsymbol{F_s}) \qquad \mathbf{(2)}$$

The above equation $\boldsymbol{f(x)}$ is the function, $\boldsymbol{S}$ denotes the total number of feature space regions, $\boldsymbol{F_s}$ represents the corresponding feature space region to **s**, $\boldsymbol{c_s}$ represents the corresponding constant to **s**. Finally,the classification result is obtained using the following formula:

$$\prod(\boldsymbol{x}, \boldsymbol{F_s}) = \begin{cases} \mathbf{1}, \boldsymbol{if}\ \boldsymbol{x} \in \boldsymbol{F_s} \\ \mathbf{0}, \boldsymbol{Otherwise} \end{cases} \qquad \mathbf{(3)}$$

The above equation provides the maximum vote obtained from all the decision trees, considered a classification result.

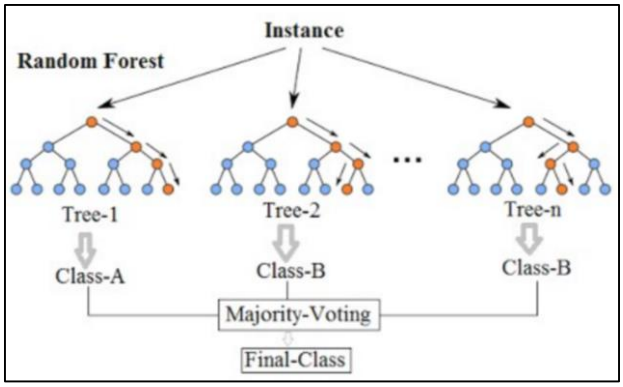
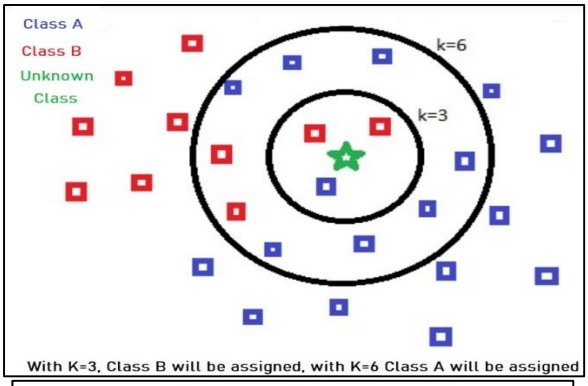


Figure-2. Random Forest Algorithm

K – Nearest Neighbour (KNN) Algorithm

In this algorithm, K – number of neighbors are compared and cluster the same instances in this single class. The term K denotes the number of cases used to comparethe same feature vector in the feature space for the training process. The duplicate instances are selected based on various measures like distance and frequency of usage. The distance between the instances is calculated using the following equation:

$$L\left(x_i, x_j\right) = \left(\sum_{i,j=1}^{n}\left(\left|x_i, x_j\right|\right)^2\right)^{\frac{1}{2}} X \in R^n \qquad (4)$$



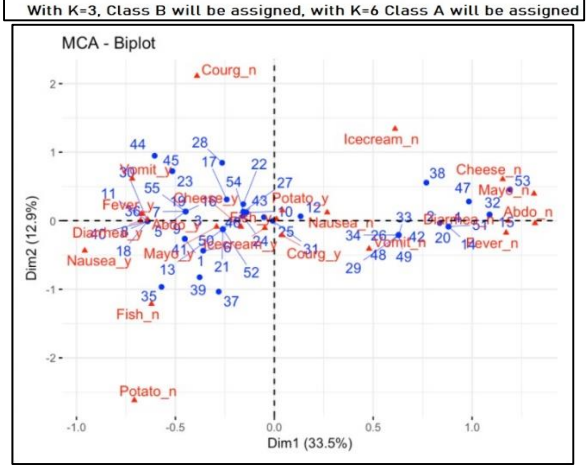


Figure-3. KNN Algorithm;**Figure-4.** Multiple Correspondence Analysis

The KNN process each data point x_i, x_j and calculates the distance to obtain the similarity-based data. In equation (4), the term L represents the length from x_i, to x_j. The entire functionality of KNN algorithm is illustrated in Figure-3.

Multiple correspondence analysis

It is a valuable tool that defines the connectivity between a set of variables. It is used to reduce the dimensions of complex data (see Figure-4). It reduces the dimensions by increasing the homogeneity of the variables. It looks for similar things among the variable and groups them. (Hwang et al. 2006). Object scores are also used in specific techniques at the initial stage. (Hwang et al. 2010). By applying MCA to an indicator matrix, which provides the scores for both the rows and columns. The obtained values are again used to the MCA, and the Burt matrix is obtained. The dimension with an eigenvalue less than one should be rejected. After rejection, a total of 6 dimensions are obtained. The values are measured by SPSS 16 and STATA 15. (Abdi and Valentin, 2007).

Analysis of Principle components

PCA is widely used as a dimension-reducing technique. It transforms the data into coordinates which can also be called uncorrelated variables. It reduces the dimensional space by combining the related variables. It also removes the unnecessary variables without affecting the data quality (Ziasabounchi and Askerzade, 2014).The steps for the PCA are,

Let,

$$\chi = \backslash\{\chi_1, \chi_2, \ldots \chi_n\backslash\} \qquad (5)$$

be the data set, and $\chi$is transformed and becomes a conventional variable. Either the correlation matrix is be calculated, or the covariant matrix is calculated. The eigenvalues and the eigenvectors are calculated. The principal components are computed. The feature vector is obtained after computation. Thus, the dimensions of the data are reduced.

Fuzzy C-means method

Among the Fuzzy clustering methods, fuzzy-c-means are widely used in the medical field. This method works based on limited integration followed by fuzzy partitioning. Let us assume that

$$\chi = \backslash\{\chi_1, \chi_2, \ldots \chi_n\backslash\} \qquad (6)$$

be the datasets and the number of clusters is $k$which is an integer between $\mathbf{2} \leq \boldsymbol{k} \leq \boldsymbol{n}$and its integration level be $\boldsymbol{u_{ij}}$. The integration level shows the intensity of the integration among the elements present in a fuzzy set. The integration level varies from **0** to **1**, and the value **0** means no integration, and the value **1** means that the elements are fully integrated. The values between **0**to **1** show the integration level among the factors (Hunt, 2012).

The FCM reduces the dataset $\boldsymbol{x}$ to $\boldsymbol{c}$ by division. The FCM identifies the center of each cluster and reduces and overall size of the objective function $\boldsymbol{J_M}$. It can be derived as,

$$\boldsymbol{J_M = \sum_{i=1}^{N} \sum_{j}^{C} U_{ij}\, m \|x_i - c_i\|^2, 1 \leq m < \infty} \quad \textbf{(7)}$$

$\boldsymbol{d_{ij}}$**-** Euclidean distance between the $\boldsymbol{\chi_i}$and $\boldsymbol{c_i}$.

$\boldsymbol{\chi_i}$- data point

$\boldsymbol{c_i}$- center

$\boldsymbol{u_{ij}}$- Integration level of $\boldsymbol{\chi_i}$ and $\boldsymbol{J}$.

$\boldsymbol{j}$- Cluster center in range $[\mathbf{0}, \mathbf{1}]$.

$\boldsymbol{M}$- Weighting exponent.

$\boldsymbol{\chi_i}$**-** a dataset with **d**-dimensional ($\boldsymbol{nxp}$).

$\boldsymbol{c_i}$**-** cluster center ($\boldsymbol{kxp}$)d-dimensional.

‖*‖ shows the correspondence of the data measured with the center. For reducing the objective function and arriving at the preceding equation, the integration level and cluster center should be maintained in the following way. (Bezdet, 2013).

$$\boldsymbol{c_i = \frac{\sum u_{ij} m X_i}{\sum u_{ij} m} forall i} \quad \textbf{(8)}$$

$$\boldsymbol{u_{ij} = \frac{1}{\sum \left(\frac{|x_i - c_i|}{|x_i - c_j|} \right)^{2/(m-1)}} forall k} \quad \textbf{(9)}$$

When, $\boldsymbol{\|U(k+1) - U(k)\| < \varepsilon}$ the iteration stops,$\boldsymbol{\varepsilon}$lies between **0** and **1**, which is the termination tolerance, and the iteration number$\boldsymbol{k}$. The stages involved in the FCM algorithm are, the partition matrix *U* is initialized, and it is shown as $\boldsymbol{[u_{jk}]}$matrix $\boldsymbol{U^0}$. The center vector $\boldsymbol{C(k) = (c_j)}$along with$\boldsymbol{U(k)}$ is calculated at $\boldsymbol{k}$ steps. The $\boldsymbol{U^k}$ is calculated with four steps. The $\boldsymbol{U^k}$is updated to $\boldsymbol{U^{k+1}}$. The iteration stops at $\boldsymbol{\|U_{(k+1)} - U_{(k)}\| < \varepsilon}$, otherwise it proceeds with step 2. (Liu et al., 2011).

Multiplayer Networks

Feed-forward neural networks mostly require multiple layers, and the multilayer networks suit well to the necessity. The multilayer network is trained using a backpropagation algorithm. The backpropagation depends on preparing a non-linear feed-forward neural network. The training is based on reducing the error by generalization. It reduces the error value between the targeted and web outputs, and it is done by adjusting their weight. The misconceptions spreading backward determine the algorithm and continuously adjusts the network weights. The backpropagation is done in 3 stages.The backpropagation algorithm involves the following steps:-

The weights are initialized by generating random values of low proportions. The

$$\boldsymbol{\chi = \backslash\{\chi(1), \chi(2), \ldots . \chi(N)\backslash\}} \quad \textbf{(10)}$$

are display inputs pointed towards the targeted outputs

$$\boldsymbol{d = \backslash\{d(1), d(2), \ldots . d(N)\backslash\}} \quad \textbf{(11)}$$

Where the $\boldsymbol{N}$ is the count of training patterns. The activation function is used to calculate the accurate outputs and its signals like

$$\boldsymbol{y = \backslash\{y1, y2, \ldots yN_M\backslash\}} \quad \textbf{(12)}$$

using the formula.

$$\boldsymbol{y_i = \varphi\left(\sum_{j-1}^{N_{M\text{-}1}} w_{ij}\,(M-1) x_j(M-1) + b_i(M-1)\right), i = 1, \ldots . , N_{M\text{-}1}} \quad \textbf{(13)}$$

The algorithm modifies the weight between the hidden layer and the output layer. Weight modification is $\boldsymbol{(w_{ij})}$and biases ($\boldsymbol{b_i}$).

$$\boldsymbol{\Delta w_{ij}(l-1)(n) = \mu x_j(n) \delta_i(l-1)(n),} \quad \textbf{(14)}$$

$$\boldsymbol{\Delta b_i(l-1)(n) = \mu \delta_i(l-1)(n),} \quad \textbf{(15)}$$

$$\boldsymbol{\delta_i(l-1)(n) = \varphi'\left(NET^{(l\text{-}1)}\right)[d_i - y_i(n)], 1} \quad \textbf{(16)}$$

$$\boldsymbol{\delta_i(l-1)(n) = \varphi'\left(NET^{(l\text{-}1)}\right) \sum w_{ki}\, \delta_k(l)(n), 1} \quad \textbf{(17)}$$

The node $\boldsymbol{j'}$soutput is $\boldsymbol{\chi_j(n)}$ in cycle $\boldsymbol{n}$and layer $\boldsymbol{l}$. $\boldsymbol{k}$ is the output count from the nodes of the neural network. The output layer is $\boldsymbol{M}$ and $\boldsymbol{\varphi}$is the activation function, and the learning rate is $\boldsymbol{\mu}$. They can increase the convergence rapidly, but they can lead to network oscillation at higher

levels and can affect the performance of the training. For achieving rapid conjunction with minimum fluctuation, momentum can be added with the standard formula for maintaining the weight. The efficiency and speed are held for the training process by making continuous adjustments, and its effects pass to the learning rate. Then the weight value maintained in the multilayer networks is frozen and passed to the testing phase.

Radial Basis Function Network

RBFN are good alternatives to traditional neural networks. Although it is one more traditional approximation function, it plays a vital role in the medical field. Typically function approximation calculates the Euclidean distance from the point to the center of each neuron. The RBFN uses the space to calculate and maintain the weight of every neuron.Steps involved in the radial base functions are:-

The random values are generated for the weights in the layer. The following equation calculates each unit j for the hidden layer.

$$\boldsymbol{y_i(x) = \sum_{j=1}^{k} w_{ij}\, \varphi_j(\|x - c_j\|), i = 1, \ldots . m' j = 1, .. k} \qquad \textbf{(18)}$$

where,

$\boldsymbol{x}$- no. of input dimensions vector

$\boldsymbol{\varphi(.)}$- Base function (neuron activation function or Gaussian function)

It is described as,

$$\boldsymbol{x - c_j} \qquad \textbf{(19)}$$

$\boldsymbol{c_j}$- hidden$\boldsymbol{j}$neuron's central vector. (with an equal number of dimensions like $\boldsymbol{x}$)

$\boldsymbol{w_{ij}}$- weight connecting the **j**th node of the output layer.

$\boldsymbol{m}$- no. of neuron node of the output layer.

$\boldsymbol{k}$-number of hidden layer nodes.

The sigmoid function is applied to every node in the output layer by the below equation.

$$\boldsymbol{\psi_k(x) = f\left[\sum_{j=1}^{L} W_{jk\varphi_j(x)}\right]} \qquad \textbf{(20)}$$

Where the no. of hidden nodes is $\boldsymbol{L}$. The following equations update the connection between the output and hidden layers by the weights:

$$\boldsymbol{w_{jk}new = w_{jk}old + \Delta w_{jk}} \qquad \textbf{(21)}$$

$$\boldsymbol{\Delta w_{jk} = \eta\big(t_k(x) - \psi_k(x)\big)\varphi_j} \qquad \textbf{(22)}$$

where,

$\boldsymbol{\eta}$- learning rate (value between 0 to 1)

$\boldsymbol{\psi_k(x)}$- network's actual output.

$\boldsymbol{t_k(x)}$- target vector's output for each pair.

The steps followed before step 3 are repeated with a small marginal error rate. But the desired output has not been achieved till now (Liu et al. 2011). The Processing steps are involved in the hybrid technique.

Step 1: Input data is obtained.

Step 2: PCA or MCA is calculated for each node.

Step 3: The reduced dimensions were obtained through the previous techniques and

inputted to the FCM analysis.

Step 4: The FCM analysis leads to clusters. These clusters are used as input for the multilayer network analysis and RBFN.

The dimensions of the data are obtained from PCA or MCA. These dimensions become the input layer of the Multilayer networks and RBFN. FCM analysis gives rise to clusters that form the output layer.

Result and Discussion

This paper has experimented with and analyzed all the proposed methods. For avoiding more complexity in understanding the functional performance of the methods, the experimental results are given for a few methods individually. The SPSS software is used in the experiment. Initially, the total variance is calculated for MCA and PCA methods, as shown in **Figure-4.** From **Figure-4**, it is noticed that MCA obtained a high total variance of 94.13%, which is better than 81.6% obtained by PCA. Then, the data dimensionality is reduced into six PCA components, and the eigenvalues with the variance are calculated. The eigenvalues and variance obtained for PCA are 80.99% and 78.22%, respectively, as shown in **Figure-5.**

The goodness of fit of a fuzzy clustering solution is evaluated by estimating multiple measures. The avg(Silhouette/cluster) should range from -1 to +1). The various measures are calculated as:

$$\begin{cases} \text{if avg(Silhouette)} \geq 0.71 & then\ structure = strong \\ \text{if } 0.51 \geq \text{avg(Silhouette)} \leq 0.70 & then\ structure = resonable \\ \text{if } 0.26 \geq \text{avg(Silhouette)} \leq 0.50 & then\ structure = weak \\ if -1 \geq avg(Shilhouette) \leq 0.25 & then\ structure = no\ substantial\ strucutre \end{cases} \qquad (23)$$

The following measure is getting normalized Dunn partition co-efficient **Fc(U)** within the range of

$$\mathbf{0} \geq \boldsymbol{Fc}(\boldsymbol{U}) \leq \mathbf{1} \qquad \mathbf{(24)}$$

a hard clustering with normalized Kaufman co-efficient **Dc(U)**within the range of

$$\mathbf{0} \geq \boldsymbol{Dc}(\boldsymbol{U}) \leq \mathbf{1} - (\mathbf{1}/\boldsymbol{K}) \qquad \mathbf{(25)}$$

Where,**K** indicates the number of clusters.

From the experimental result, it is noticed that Fc(U) is large than Dc(U) since the output of FCM is unpredictable. Hence Fc(U) is selected for normalized clustering, and it is integrated with MCA and PCM for improving the clustering performance by reducing the dimensionality. **Table-1** explains that FCM is combined with MCA and PCA, and running the analysis based on their dimensions exceeds the performance of the FCM clustering method. From the values given in Table-1, it is identified that incorporating MCA, FCM, and MLP can provide better performance in terms of the eigenvalues.

Classification by Neural Networks

In addition to the above-said methods, the initial layers of machine learning algorithms are used to analyze the output obtained from MCA and PCA. For evaluating the performance of the classification algorithms FCM such as FCM and other methods integrated with it. The performance comparison is given in **Table-2**.

Dataset

Population data is collected from the medical industry for cardiac, chest, and vascular diseases from Ain Shams University from 2010 to 2020. The main objective of this paper is to reduce the error rate and improve the classification accuracy of the deep learning algorithms. For analyzing and classifying the dataset, 17 attributes are taken.

Classification by neural networks

The performance evaluation is obtained from the experimental results by comparing the results such as error, accuracy in training, and testing. Table-3 shows that classification accuracy increased in all cases of incorporating FCM with other methods. In contrast, the results of MLP and RBF before merging have less performance which confirms the importance of integrating these methods. Still, it is noticed that the relative error percent in training and testing sample of MLP via FCM–MCA structure is least in comparison to MLP, RBF, MLP via FCM–PCA, RBF via FCM–MCA, and RBF via FCM–PCA. **Table-3** also revealsTable-2. Performance results comparison Method No. clusters Average silhouette The normalized Dunn partition coefficient Fc (U) The partition coefficient Dc (U) FCM 2 0.12 0.00 0.91 FCM via MCA 2 0.72 0.74 0.08 FCM thro' PCA 2 0.71 0.74 0.15 **Figure-6**. Scree plot of eigenvaluesfrom MCA+FCM+MLP, Hybrid techniques 231 that the classification accuracy ratio is higher in the case of MLP via FCM–MCA technique comparable with the techniques referred to earlier. The training and test results are also depicted for all structures in **Table-3**. It indicates that the training time for MLP via FCM–MCA is 00.016s.

On the other hand, MLP, RBF, MLP via FCM–PCA, RBF via FCM–MCA, and RBF via FCM–PCA structures have the training time 00.039, 00.603, 00.065, 00.065, 00.155, and 00.170 s, respectively, which is much more than that of MLP via FCM–MCA structure. It is noticeable that MLP via FCM–MCA performance was superior to other methods. It is noticeable that the performance of MLP via FCM–MCA was superior to other methods, so the normalized importance chart presented for it only. The results were compared based on each value of percent of relative error in training with value 2.5%, percent of relative error in testing with value 5.1%, the ratio of classification **Table-3**.From the experiment, the output obtained for all the structures in terms of time efficiency is given in **Table-2**. It includes both training and testing phases. The time complexity is calculated for evaluating the performance of FCM based methods.

Table-1. Results Obtianed for Training and Testing Process

Method	No. of Classes	Avg. Silhouette	Fc(U)	Dc(U)
FCM	5	0.12	0.01	0.93
FCM-MCA	5	0.73	0.75	0.09
FCM-PCA	5	0.72	0.75	0.16
MCA+FCM+MLP	5	0.73	0.77	0.08

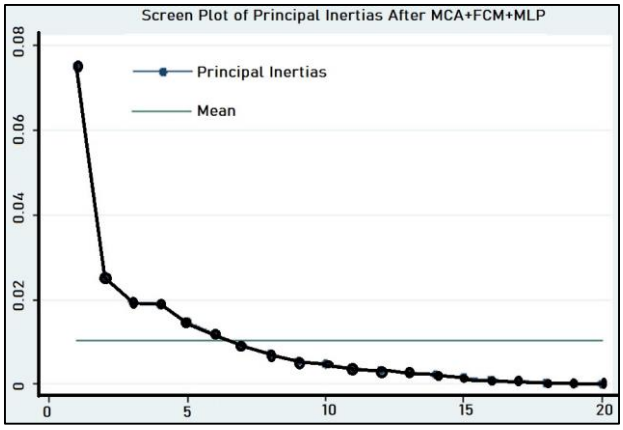


Figure-5. Eigenvalues Obtained Using MCA+FCM

Table-2. Time Complexity Comparison

Method	Time (hr)
MLP	0.39
RBF	0.60
FCM	0.016
MLP + FCM AND PCA	0.065
RBF + FCM AND MCA	0.155
RBF + FCM AND PCA	0.170
MCA+FCM+MLP	0.02

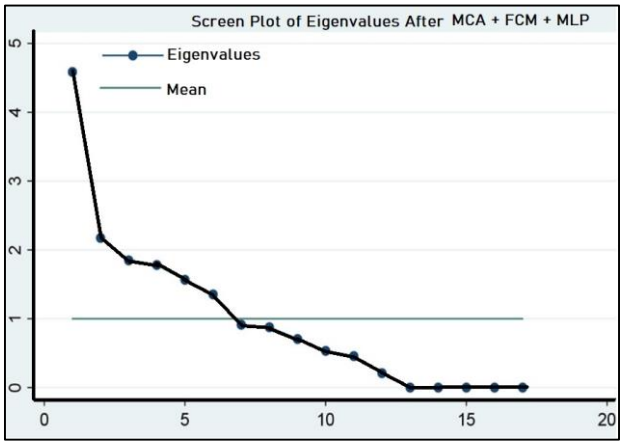


Figure-6. Eigenvalue Obtained Using MCA+FCM+MLP

Table-3. Accuracy Comparison

Method	Time (hrs)	Error-Testing	Training Accuracy	Testing Accuracy
MLP	0.39	23.0	78.7	77.0
RBF	0.60	36.1	76.1	63.9
FCM	0.016	21.1	98.1	96.6
MLP + FCM AND PCA	0.065	5.1	95.4	93.7
RBF + FCM AND MCA	0.155	6.3	97.5	94.9
RBF + FCM AND PCA	0.170	7.9	94.1	92.1
MCA+FCM+MLP	0.12	4.8	98.99	99.4

The comparison found that the incorporation of MLP, FCM, and PCA takes 0.16sec, which is very less than other methods (see **Table-2**).The MLP methods takes 0.39, RBF methods takes 0.60 FCM methods takes 0.016 MLP + FCM AND PCA methods takes 0.065 RBF + FCM AND MCA methodstakes 0.155 RBF + FCM AND PCA methods takes 0.170 (see**Figure-5**). Finally, the performance of all the FCM combined methods is evaluated by comparing their performance in terms of relative error, accuracy, and time for training and testing phases (see**Table-3**). From the comparison, it is understood that FCM + MLP + PCA outperforms other methods.

Conclusion

This paper's main objective is to provide a better hybrid approach for improving the accuracy of data mining. So it involves various state-of-art methods such as FCM, MCA, PCA, MLP, and RBFN. By integrating the special features of the methods mentioned above and investigating their performances. SPSS software is used and experimented with heart diseases data set. The purpose of the study is to enhance the prediction accuracy of the heart disease data. Based on the prediction, the treatment methodology can be selected. The experimental results conclude that the integration of MLP, FCM, and MCA is the best hybrid model for improving classification accuracy. The performance comparison estimates the error values, classification accuracy, and time complexity.

The selected best method will be experimented on various healthcare data sets and confirmed for betterness in future work.

REFERENCES

1. Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. *Encyclopedia of measurement and statistics*, *2*(4), 651-657.
2. Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.

3. Bhateja, V., Le Nguyen, B., Nguyen, N. G., Satapathy, S. C., & Le, D. N. (Eds.). (2018). *Information Systems Design and Intelligent Applications: Proceedings of Fourth International Conference INDIA 2017* (Vol. 672). Springer.
4. Bhatla, N., & Jyoti, K. (2012). An analysis of heart disease prediction using different data mining techniques. *International Journal of Engineering*, *1*(8), 1-4.
5. Chakraverty, S., Sahoo, D. M., &Mahato, N. R. (2019). *Concepts of soft computing: fuzzy and ANN with programming*. Springer Singapore.
6. Dalvi, R. D. F., Zago, G. T., &Andreão, R. V. (2017). Heartbeat classification system based on neural networks and dimensionality reduction. *Research on Biomedical Engineering*, *32*, 318-326.
7. Deng, S. (2020). Clustering with fuzzy C-means and common challenges. In *Journal of Physics: Conference Series* (Vol. 1453, No. 1, p. 012137). IOP Publishing.
8. El-Bialy, R., Salamay, M. A., Karam, O. H., & Khalifa, M. E. (2015). Feature analysis of coronary artery heart disease data sets. *Procedia Computer Science*, *65*, 459-468.
9. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, *2018*.
10. Hunt, V. D. (2012). *Artificial intelligence & expert systems sourcebook*. Springer Science & Business Media.
11. Hwang, H., Dillon, W. R., &Takane, Y. (2010). Fuzzy cluster multiple correspondence analysis. *Behaviormetrika*, *37*(2), 111-133.
12. Hwang, H., Dillon, W. R., &Takane, Y. (2006). An extension of multiple correspondence analysis for identifying heterogeneous subgroups of respondents. *Psychometrika*, *71*(1), 161-171.
13. Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013, March). Heart disease prediction using lazy associative classification. In *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)* (pp. 40-46). IEEE.
14. Kumari, M., &Godara, S. (2011). Comparative study of data mining classification methods in cardiovascular disease prediction 1.
15. Kumar, M., Shambhu, S., & Sharma, A. (2018). Classification of heart diseases patients using data mining techniques. *International journal of research in electronics and computer engineering*, *6*(3), 1495-1499.
16. Kurt, I., Ture, M., &Kurum, A. T. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert systems with applications*, *34*(1), 366-374.
17. Le, T. L. (2019). Fuzzy C-Means clustering interval type-2 cerebellar model articulation neural network for medical data classification. *IEEE Access*, *7*, 20967-20973.
18. Patra, D., & Pradhan, S. (2008, November). Integration of FCM, ICA and Neural network for ECG signal classification. In *Proc. of IEEE International Conference on Soft Computing (ICSC 2008), IET Alwar* (pp. 8-10).
19. Riahi-Madvar, H., Dehghani, M., Seifi, A., Salwana, E., Shamshirband, S., Mosavi, A., & Chau, K. W. (2019). Comparative analysis of soft computing techniques RBF, MLP, and ANFIS with MLR and MNLR for predicting grade-control scour hole geometry. *Engineering Applications of Computational Fluid Mechanics*, *13*(1), 529-550.
20. Wiharto, W., &Suryani, E. (2020). The comparison of clustering algorithms k-means and fuzzy c-means for segmentation retinal blood vessels. *Acta Informatica Medica*, *28*(1), 42.
21. Ziasabounchi, N., &Askerzade, I. N. (2014). A comparative study of heart disease prediction based on principal component analysis and clustering methods. *Turkish Journal of Mathematics and Computer Science (TJMCS)*, *16*, 18.