



An Effective Fuzzy Neural Network with Reinforcement Learning Approach for Medical Data Classification

Panem Penchala Prasad^{1*}

Fidele Sagayaraj Francis²

Soudagar Zahoor Ul Huq³

¹*Gunampalli Pulla Reddy Engineering College, India*

²*Pondicherry Engineering College, India*

³*Gunampalli Pulla Reddy Engineering College, India*

* Corresponding author's Email: prasad.cse@gprec.ac.in

Abstract: Nowadays, several researchers are facing challenges on the prediction of diseases from the huge volume of medical databases. So, researchers are using data mining techniques like association rules, classification and clustering to address challenges. The physicians make the right decisions for successful diagnosis of various diseases by using the prediction. The existing work classifies the data to predict the certain diseases, but still it faces the difficulties due to overfitting in the training data. The main aim of this research work is to classify the Medical Data (MD) by developing the feature selection based approach in MD Classification as MDC. The irrelevant features are eliminated from the MD by using Recursive Feature Elimination (RFE) method, then ranked the features to reduce the computation cost of the proposed method. The ranked features from the RFE are given as input to the Fuzzy Neural Network (FNN) with Reinforcement Learning (RL), which is used for classification. The proposed RFE-FNN method has the accuracy of the 98.57%, 98.15% sensitivity, 98.64% specificity and 95.47% F-Measure in Heart disease dataset.

Keywords: Classification, Disease prediction, Fuzzy neural network, Medical dataset, Recursive feature elimination, Reinforcement learning.

1. Introduction

Nowadays, the relational databases are developed and enriched with useful informative MD due to the importance of healthcare information systems [1, 2]. In general, the huge volume databases analyzed and predicted by using various machine learning and data mining techniques. In several medical domains like diabetes, heart diseases, cancers and other complex diseases, various predictive models have been developed [3]. The manual practice is rapidly overcome by using the digitization of clinical procedure in the medical domain. Nowadays, medical specialists consider the promising application like Computer Aided Diagnosis (CAD) due to early processing and their effective handling of identification of diseases, however it is facing several challenges [4, 5]. But, the important issue is to handle the appropriate medical records, which will

help the doctors and technologist to predict the diseases rapidly and further proceed for a speedy recovery. To manage and organize the data records, the hospitals use effective and efficient systems, namely CAD systems, decision support systems and management systems [6].

The presence of specific diseases is identified in the patients by extracting and learning the useful information from the vast amount of data with attributes using data mining techniques [7]. In healthcare informatics, various data mining techniques such as supervised and unsupervised learning techniques, feature selection techniques are used to predict the different diseases [8]. There are various algorithms in use for supervised learning techniques and unsupervised learning technique. The algorithms are being applied in healthcare analytics [9] where feature processing algorithms, namely Principal Component Analysis (PCA), variations of PCA, i.e. Kernel Principal Component Analysis

(KPCA) and Linear Discriminant Analysis (LDA) are used to process and transform the collected patient data records into acceptable form [10]. The mathematical or optimized statistical techniques such as Genetic algorithms, RFE, Fisher score and so on are used to further extract or score the newly transformed features. The associated parameters are used to predict the patient's diseases as normal or abnormal by using classification techniques [11]. The techniques are used for achieving less false alarms with the maximum detection rate, where the common techniques are Neural Network (NN), Support Vector Machine (SVM), Decision Trees (DTs) and Naive Bayes [12]. In this research, the important features are selected by using RFE method which is used to increase the performance of the MDC. The important features of the medical data set are given to FNN with RL method for classifying the data. The existing work classifies the data to predict certain diseases, but still it faces the difficulties due to overfitting in the training data. In this research work, the main aim is to classify the Medical Data (MD) by developing the feature selection based approach in MD Classification as MDC. The irrelevant features are eliminated from the MD by using Recursive Feature Elimination (RFE) method, then rank the features to reduce the computation cost of the proposed method.

The rest of the paper is organized as: review of several recent techniques which is used for classifying the MD are explained in section 2. The detail description of proposed RFE-FNN is presented in section 3, whereas the validation of the proposed method with existing techniques is described in section 4. Finally, the conclusion of this research work is discussed with future work.

2. Literature survey

According to non-linear techniques, several researchers developed various techniques for MDC. A review of recent techniques [13-18] discussed in this section, where the advantage and the drawback of these techniques are presented below.

L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang, H. Li, and D. Liu [13] tried to solve the MDC problems by implementing SVM with Fruit-fly Optimization Algorithm (FOA). The generalization capability of SVM classifier was maximized by using optimal parameters of FOA technique for MDC. Compared to existing optimization techniques with SVM, the SVM-FOA algorithms provided better performance in four UCI dataset in terms of accuracy, specificity and Area Under Curve (AUC). But, an improvement on the performance of the SVM is crucial by

parameter setting and feature selection since they influence one another.

N. Kausar, A. Abdullah, B. B. Samir, S. Palaniappan, B. S. AlGhamdi, and N. Dey [14] differentiated the normal and abnormal patients by selecting the relevant clinical features using the PCA algorithm. The related parameters and their measures were adjusted and additionally SVM and k-means clustering algorithms were used for classification. The idea of the clustered approach was to minimize the classification time and increased detection rate with unseen patterns of similar behaviors within the selected cluster. The experiments were conducted on the heart disease dataset, which proved that SVM-PCA achieve maximum accuracy in less processing time for desired cluster sizes. But, this method cannot use for multi-classification of different medical datasets.

V. Chaurasia, P. Saurabh, and B. B. Tiwari [15] developed the J48, Naive Bayes and Radial Basis Function (RBF) to identify the breast cancer i.e. benign and malignant for patients. The prediction models were developed by applying data selection, pre-processing and transformation of data. In this work, the unbiased prediction accuracy was measured by using 10-fold cross validation procedure with stratified sampling technique. This method achieved better results of accuracy for Naive Bayes on Breast cancer dataset, but leads to very poor performance on J48 due to over fitting of data.

A. Khatami, A. Khosravi, T. Nguyen, C.P. Lim, and S. Nahavandi [16] classified the multiclass radiography images by using three step framework such as Wavelet Transform (WT), Kolmogorov Smirnov (KS) and Deep Belief Networks (DBNs). The WT and KS were used as the de-noising techniques to remove the insignificant features and noises of the image. An unsupervised DBN was used to learn the un-labelled features in the second step. The inputs were obtained in the first two steps from the discriminative feature subsets and evaluations were carried out in the third steps by using classifiers. The method provided better performance in multiclass radiography image classification, reduced the computational costs and achieve the robustness for handling complex medical images. This process is time consuming, because there are no optimal number of features selected for KS test.

M.Z. Alam, M. Saifur Rahman, and M. Sohel Rahman [17] developed a feature ranking, selection strategy and suitable classifier algorithm for MDC. The disease was predicted by using the various feature ranking strategies and random forest as final classifiers. The experiments were conducted on ten UCI dataset that obtained highly accurate predictors.

This method also useful in any other domain that exhibit similar characteristics of features. But, this method provided poor performance due to the presence of noise and outlier data in the training set.

J. Zhang, M. Liu, and D. Shen [18] identified huge scale anatomical landmarks by implementing a Two-Stage Task-Oriented Deep Learning (T2DL) method with limited training data. The inherent associations between target anatomical landmarks and local image patches were learned by using two deep Convolutional Neural Network (CNN). Among the image's patches, the correlations were captured by directly coordinating the predicted landmarks with CNN model in an end-to-end manner. In a real time, multiple landmarks were detected by effectively training the T2DL model in a global task-oriented manner. In real applications, various landmarks have several importance, but in this work, landmarks were treated equally in both prostate and brain datasets.

The existing techniques provide poor performance on MDC due to overfitting on the training data. To overcome this limitation, FNN with RL is proposed in this paper to avoid overfitting of training data which additionally improves the performance of the MDC.

3. Proposed methodology

MDC is a challenging task to classify the particular disease, as the data are available in huge amount. i^{th} variables are normalized machine learning techniques used effectively for classifying the data and to predict the disease. The features are selected by using RFE technique which is used to improve the effectiveness of classification in MD. The selected features are given as an input to the fuzzification techniques, which is further trained by ANN using Reinforcement Q-Learning for classification [19]. The performance of proposed techniques can be validated by using extensive experiments on medical UCI respiratory data [20]. The function of proposed method RFE-FNN can be explained in this section and Fig. 1 shows the block diagram of the proposed method.

The raw data can be collected by using standard UCI datasets for classifying the medical dataset. The data may contain noises, outliers, missing values which lead to poor classification accuracy. Hence, pre-processing should be done for this data to find the accurate risk factors for patients.

3.1 Pre-processing

After data collection, the data is pre-processed because there are four records having missing values,

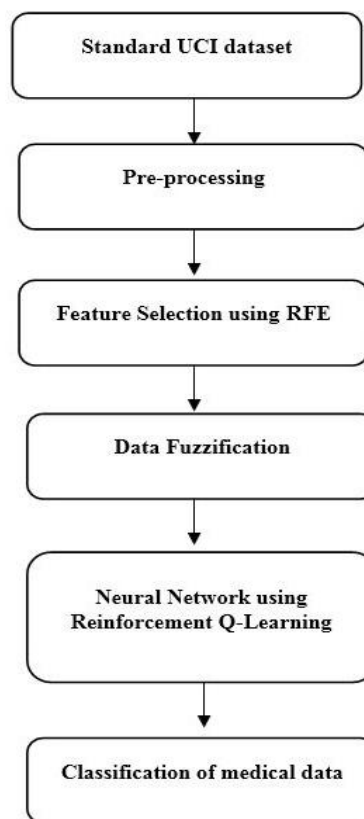


Figure. 1 Block diagram of RFE-FNN technique

which should be removed from the dataset. In this stage, the noises and missing values are reduced by several data pre-processing techniques like data cleaning, normalization, transformation, discretization and integration of data. In this research work, normalization method is used for pre-processing to reduce the dimensionality of data. After extracting the data, it is classified into three various types, namely training, testing and validation. In deep learning architecture, the effects of scales are avoided by normalizing the data into [0;1]. Eq. (1) presents the normalization method, which can be applied in the extracted dataset.

$$V_i = \frac{a_i - \min_{a_i}}{\max_{a_i} - \min_{a_i}} \quad (1)$$

Where, i^{th} variables are normalized by using a value as a_i , minimum valor in the training set is represented as \min_{a_i} , which is registered for this variable and maximum valor are described as \max_{a_i} , which is also registered for this variable. After the preprocessing stage, the data can be split into training and testing to obtain a high risk factor. For experimental analysis, the pre-processed data can be split into 70% for training and 30% for testing.

3.2 Selection of features

In this process, the original high dimensional feature vector is reduced by studying the importance of selected relevant features. Feature selection is considered as one of the most important process for building an efficient and powerful prediction engine. The intuitive insights are provided by the importance of various feature components and this insight is useful to make the prediction process more transparent to the end users. The selected features are used to reduce the computational complexity of data. In feature selection, still there are some unrelated data, which leads to poor accuracy. Hence, the irrelevant features should be removed from the selected data that can be carried out by using RFE technique. After the number of iterations, the remaining features are considered as the most useful in the process of discrimination. Therefore, RFE is used to eliminate the irrelevant features iteratively, instead of removing the related features. In the next section, the importance of features is discussed with RFE techniques to further predict the best features of data classification.

3.3 Recursive feature elimination (RFE) technique

According to important measures, the features are ranked by using RFE selection method, which is basically a recursive process. The less relevant features are removed and important features are measured in every iteration. To speed up the process every time, a group of features is removed by using RFE method. During the stepwise elimination process, each feature's relative importance is changed subsequently when validating the various subsets of features. This process is suitable for highly correlated features. A final ranking is constructed by eliminating the inverse order features, where the selection process consists only in taking the first n features from this ranking. The pseudo-code of RFE is shown below which is used to remove the irrelevant features and improves the classification of MD.

Algorithm 1: Recursive Feature Elimination

Inputs:

- Training set T
- Set of p features $F = \{f_1, \dots, f_p\}$
- Ranking method $M(T, F)$

Outputs:

- Final ranking R

Code:

```

Repeat for  $i$  in  $\{1:p\}$ 
  Rank set  $F$  using  $M(T, F)$ 
   $f_1 \leftarrow \text{lastrankedfeaturein}F$ 
   $R(p-i + 1) \leftarrow f_1$ 
   $F \leftarrow F - f_1$ 
    
```

The selected are given as an input for data fuzzification, which is explained in the below section.

3.4 Fuzzy model

The specification of a set of linguistic if-then rules is defined in the form of IF X Then Y, and fuzzy controller is introduced according to the fuzzy set theory. Here, the fuzzy sets' labels are described as X and Y. The combined output is produced by activating the membership functions independently and an appropriate defuzzification function is used to compute the combined output.

The linguistic notations are defined by sets in fuzzy and the truth-value of the linguistic expressions are defined by the membership function [21]. The universe of discourse is the domain and the range interval is $[0,1]$, which defines the functions (i.e. membership function) by the membership degree to a fuzzy set of an object. A fuzzy rule $R(i)$ with C rules in a D -inputs-single-output system are expressed as in Eq. (2).

$$R(i): \text{IF } s \text{ is } m(i) \text{ THEN } o(i), \text{ with } i \in \{1, \dots, C\} \tag{2}$$

Where, the input vector is described as $s \in R^D$, membership of fuzzy set for the input vector is represented as $m(i)$ in the premise part and finally the real number describes as $o(i)$ in consequent part. The number of pairs of input and output data is undergoing problems. Fig. 2 represents the process of pre- and post-process as fuzzification and defuzzification, which occur before and after the fuzzy inference system.

The rule based on the fuzzy control systems is static, even though the fuzzy systems achieved success in various applications. An expert operator manually tuned the parameters until it achieves better performance. A fuzzy system designer faces challenges includes computational time and efforts for developing the parameters like membership functions and rules. To overcome the above issues, the outputs are given to the neural network, where the overfitting of training data can be avoided by reinforcement Q-Learning.

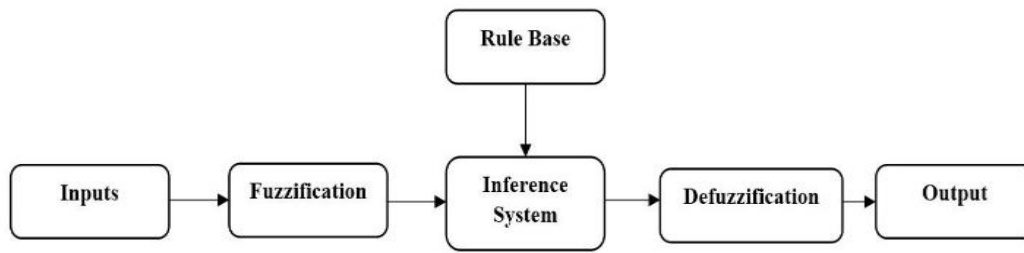


Figure. 2 Basic fuzzy inference system

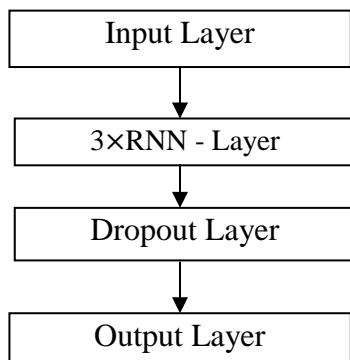


Figure. 3 Block diagram of the FNN model

3.5 Neural network with reinforcement Q-learning

NN model is used to detect structures in streams of data. The current state's output is predicted according to the output of previous states, but the feed forward NN computes unidirectional from the input to the output. The block diagram of the FNN with Reinforcement Q-Learning is shown in the Fig. 3. The input layer is the first layer where, all the data given to the network are collected, followed by RNN layer, Dropout layer and at last output layer. The RNN layer performs Sigmoid actuation function that generates the output values and those are processed for generating predicted values. The binary Cross Entropy Loss is calculated in order to find the loss in the prediction and the final results are obtained in the output layer.

The interactions with the environment occur at discrete time steps, where the learning in NN is a continuous process. The environment states (i.e., state, s_t) are received by the agent and selects the action (a_t) in a typical RL cycle (at time t) for interaction. Then, a new state (s_{t+1}) is progressed by responding to the action, which is carried out by the environment state. For the selected action, the agent may either get the reward (r_{t+1}) or not, and also the agent may identify that with the transition (s_t, a_t, s_{t+1}). According to certain policy, action-value function $Q(s, a)$ or value function $V(s)$ should be updated by the agent after every cycle. The

policy (π) is a function that maps states $s \in S$ to actions $a \in A$, i.e., $\pi : S \rightarrow A \Rightarrow a = \pi(s)$.

The most possible solution to address the RL problem by describing the policy, reward function, state value function pairs and policy. The pair of value function is separated into either action-value function as Q or state-value function as V .

In the state-value function, the expected outcomes for the prediction are carried out by following the policy π and s as state. At future time steps, the rewards are added to identify the policy using the given discount factor ($\gamma \in [0, 1]$). The Eq. (3) defines it as,

$$V^\pi(s) = E_\pi(\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s) \quad (3)$$

The expected outcomes are predicted in the action-value function, where policy π is calculated by summing the rewards for every state action pair as s , which is explained in Eq. (4),

$$Q^\pi(s, a) = E_\pi(\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a) \quad (4)$$

There is an achievement in the optimal policy and solve the Markov Decision Processes (MDP) using dynamic programming to identify the optimal policy and its value in the following way. At first, either improve the initial policy iteratively as policy iteration or compute the improved action-value or state-value function by recursively refining the arbitrary value function. For a given policy, the Bellman expectation equation is used to estimate the state-value function in some simple cases, which is mathematically explained in Eq. (5).

$$V^\pi(s) = E_\pi(r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s) \quad (5)$$

The policy evaluation process is considered to maximize the state-action values, to achieve the optimal improved policy as π' . But, in the unknown environment scenarios, the model-free methods are used without using MDP. In some cases, instead of state-value, the action-value is maximized. The

Table 1. Dataset description

| Dataset with ID | Total Features | No. of sample for Training | No. of Samples for testing |
|------------------|----------------|----------------------------|----------------------------|
| Heart Disease-HD | 13 | 180 | 90 |
| Breast Cancer-BC | 9 | 499 | 200 |
| Diabetes-PID | 8 | 576 | 192 |
| Parkinson's-Pks | 22 | 130 | 65 |

similar policy and improvement process evaluation are used to identify the optimal policy, i.e. π' . The Eq. (6) explains the process as,

$$Q^\pi(s, a) = E_\pi(r_{t+1} + \gamma Q^\pi(s_{t+1}, a_{t+1}) | s_t = s, a_t = a) \quad (6)$$

Hence, by using RL with Q, the overfitting is avoided in training data which is used to improve the performance of RFE-FNN. The experiments and their validated results are discussed in the next sections.

4. Experimental results

In this section, the validation of the proposed method and their experimental results are discussed with various existing techniques. Four biomedical datasets such as Pima Indians diabetes, Wisconsin breast cancer, heart disease and Parkinson's datasets are collected from the UCI machine learning repository for identifying the performance of the proposed method. Table 1 shows the dataset description with its ID, number of features, sample of training and testing.

The missing values are present only in Wisconsin dataset, the missing categorical attributes are replaced by using the mode of the attributes and the missing continuous data are replaced by mean of the attributes. During calculation, the numerical difficulties are addressed by scaling the data in the range of [-1,1] before constructing the FNN models. Hence, the feature values in the smaller numerical ranges are not dominated by those values present in the greater numerical ranges. In the below subsection, the evaluation of parameter settings and the experimental validated results of proposed RFE-FNN against various existing techniques are explained.

4.1 Parameters setting

The computer with 2.2 GHz of Intel Core i5, RAM of 8GB, where the RFE-FNN method is developed using the programming language of Python 3.7.3 version. The performance of proposed RFE-FNN method is validated by conducting several experiments on UCI dataset using various metrics namely AUC, accuracy, F-measure, specificity (precision) and sensitivity (recall).

The proportion of positive samples that are correctly classified as positive by using sensitivity rate i.e. true positive rate. In contrast with this, the negative samples are correctly classified as negative by using specificity measure i.e. true negative rate. Accuracy can be calculated by using the Eq. (7), and the Eq. (8) is used to evaluate the single combined metric, which is defined as F-measure. Among the number of labelled positive class samples, precision is used to identify the number of accurately labelled samples, which is shown in Eq. (9). On the contrary, according to the positive class, recall is used to predict the number of accurate positive class labelled samples, which can be divided by the total number of samples. The mathematical expression for recall is given in Eq. (10).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (7)$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

$$Specificity = \frac{TN}{TN+FP} \quad (9)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (10)$$

Where, TP is true positive, TN is true negative, FP is false positive and FN is false negative.

4.2 Performance evaluation of proposed RFE-FNN

The performance evaluation of existing techniques with proposed RFE-FNN are compared in terms of accuracy, F-measure, sensitivity, specificity and AUC for four datasets which are explained in this section. Compared to the RFE-FNN with SVM+FOA [13], PCA with SVM+K-means [14], hybrid approach such as RBF+Navie Bayes+ J48 [15], WT-DBN [16] and FR-RFA [17], the best supervised approach is determined by using extensive experiments. The validation of RFE-FNN are discussed with each parameter and their results are discussed in the sub-section.

4.2.1. Performance of RFE-FNN in terms of accuracy and AUC

In this section, accuracy and AUC of proposed RFE-FNN is compared with SVM+FOA, PCA with SVM+K-means, hybrid approach such as RBF + Naive Bayes+J48 and FR-RFA for four different datasets such as Heart disease, Breast cancer, Parkinson's and diabetic dataset. Table 2 describes the values of proposed RFE-FNN in terms of accuracy and the graphical representation is given in Fig. 4.

Compared to other existing techniques, the RFE-FNN method provides better performance for accuracy in all datasets, which is explained by Table 2 and Fig. 4. The proposed method obtained high accuracy (97.82%) in BC dataset, whereas 78.29% achieved in PID dataset. The proposed FNN achieves higher accuracy compared to the proposed SVM, as the overfitting problem is overcome in all dataset. The SVM achieved nearly 96% of accuracy in three

Table 2. Comparison of RFE-FNN over accuracy

| Methods | Accuracy(%) | | | |
|-------------|-------------|-------|-------|-------|
| | HD | BC | PID | Pks |
| SVM+FOA | 96.38 | 96.90 | 77.46 | 96.90 |
| SVM+K-means | 90.57 | 96.27 | 76.50 | 96.27 |
| Naïve Bayes | 97.36 | 95.64 | 76.48 | 96.24 |
| RBF | 96.77 | 95.57 | 76.47 | 95.45 |
| J48 | 93.41 | 96.24 | 76.26 | 95.57 |
| FR-RFA | 83.50 | 97.82 | 78.29 | 95.12 |
| RFE-FNN | 98.54 | 98.71 | 80.47 | 97.15 |

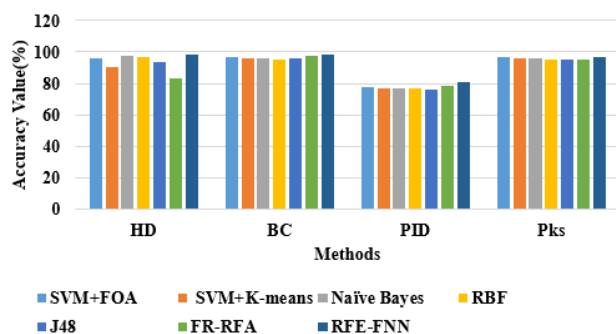


Figure. 4 Accuracy of proposed RFE-FNN

Table 3. AUC value for proposed RFE-FNN

| Methods | AUC (%) | | | |
|---------|---------|-------|-------|-------|
| | HD | BC | PID | Pks |
| SVM+FOA | - | 96.87 | 72.34 | 96.87 |
| FR-RFA | 87.05 | 98.45 | 85.40 | 96.13 |
| RFE-FNN | 89.58 | 99.12 | 87.46 | 97.67 |

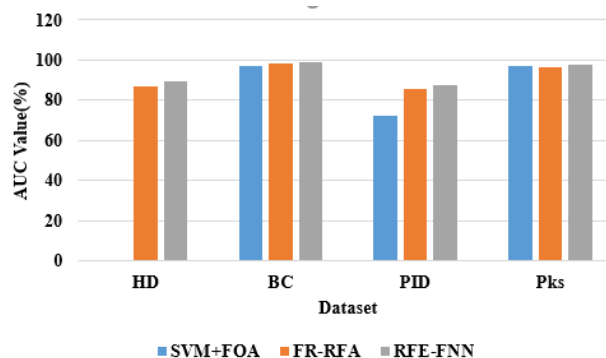


Figure. 5 Comparison of proposed method over AUC

datasets, but it provides very low accuracy (77.46%) in PID dataset. Table 3 describes the performance of RFE-FNN with SVM+FOA and FR-RFA in terms of AUC for all dataset.

The Fig. 5 shows that the RFE-FNN method achieved higher AUC value (99.12%) in BC dataset, whereas SVM+FOA achieved 96.87%. The RFE-FNN achieved 89.58%, 87.46% and 97.67% for HD, PID and Pks dataset, whereas FR-RFA technique achieved 87.05%, 85.40% and 96.87% AUC values for above three datasets. The existing SVM technique achieved a very low AUC in PID dataset as 72.34%.

4.2.2. Performance of RFE-FNN in terms of sensitivity and specificity

In this section, the parameters like specificity and sensitivity of RFE-FNN are compared to existing techniques such as SVM+FOA, RBF, Naive Bayes, J48, WT-DBN and FR-RFA. The experimental results are tabulated in Table 4 and Fig. 6, 7 where, the graphical representation of sensitivity and specificity of RFE-FNN with several existing techniques are shown.

Table 4. Performance of RFE-FNN in terms of sensitivity and specificity

| Methods | Sensitivity (%) | | | | Specificity (%) | | | |
|-------------|-----------------|-------|-------|-------|-----------------|-------|-------|-------|
| | HD | BC | PID | Pks | HD | BC | PID | Pks |
| SVM+FOA | 78.61 | 96.86 | 55.07 | 96.86 | 82.58 | 96.89 | 89.62 | 96.86 |
| WT-DBN | 93.25 | 92.94 | 92.64 | 95.89 | 97.05 | 96.70 | 95.71 | 97.85 |
| Naive Bayes | 97.40 | 96.24 | 53.59 | 96.42 | 97.90 | 96.59 | 88.80 | 96.59 |
| RBF | 97.07 | 96.27 | 54.13 | 94.57 | 96.23 | 94.32 | 88.16 | 94.32 |
| J48 | 93.43 | 96.62 | 53.84 | 96.62 | 90.37 | 95.45 | 88.61 | 95.45 |
| FR-RFA | 83.51 | 98.50 | 79.49 | 95.14 | 84.16 | 98.25 | 78.46 | 95.47 |
| RFE-FNN | 98.15 | 99.42 | 94.56 | 98.47 | 98.64 | 99.74 | 96.39 | 98.50 |

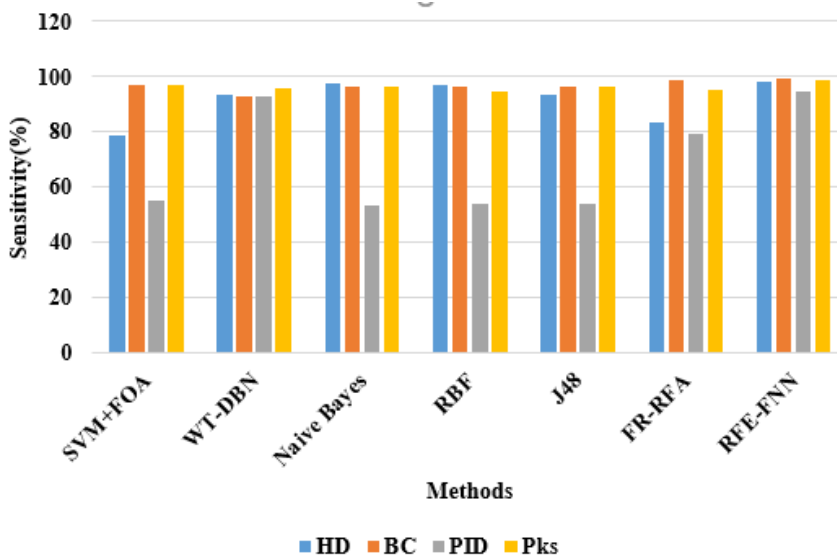


Figure. 6 Analysis of RFE-FNN over sensitivity

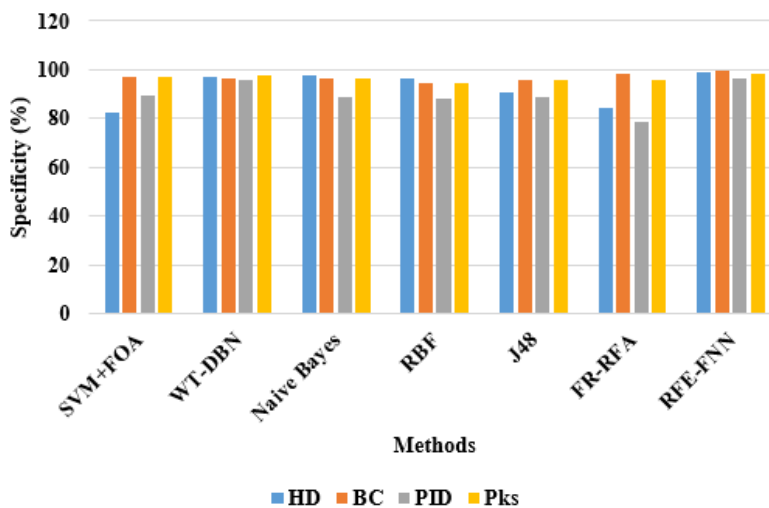


Figure. 7 Analysis of RFE-FNN over specificity

The above Fig. 7 represents the experimental results of RFE-FNN for sensitivity on all four datasets. This shows that RFE-FNN achieved 98.15%, 99.42%, 94.56% and 98.47% recall values for HD, BC, PID and Pks dataset, whereas the existing RBF achieved 97.07%, 96.27% and 94.57% recall for HD, BC and Pks. The sensitivity values for PID dataset are very low for existing methods as well as for proposed RFE-FNN method. This is due to overfitting of data in the training phase, which leads to poor performance in sensitivity.

Compared to sensitivity values for PID dataset, the specificity values are very high for all methods, which is shown in above Fig. 7. In BC dataset, the proposed RFE-FNN method achieved 99.74% precision, whereas Naive Bayes, WT+DBN and FR-RFA techniques achieved 96.59%, 96.70% and 98.25% precision values. When compared with SVM+FOA, the RFE-FNN method nearly improves

6% precision values on PID dataset. Finally, the F-Measure for proposed method is explained in below section.

4.2.3. Performance of RFE-FNN in terms of f-measure

The experiments are conducted on all dataset to validate the performance of RFE-FNN in terms of F-measure, which are shown in Table 5. The graphical representation for the F-Measure of RFE-FNN are compared with WT-DBN and FR-RFA is described in Fig. 8.

Table 5. Performance of f-measure

| Methods | F-Measure (%) | | | |
|---------|---------------|-------|-------|-------|
| | HD | BC | PID | Pks |
| WT-DBN | 93.05 | 92.94 | 91.83 | 95.84 |
| FR-RFA | 83.52 | 98.00 | 78.15 | 96.18 |
| RFE-FNN | 95.47 | 99.07 | 93.62 | 97.21 |

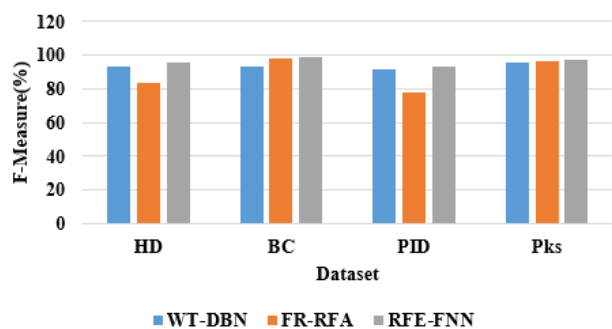


Figure. 8 RFE-FNN performance on f-measure

The experimental analysis on F-Measure proved that the RFE-FNN method achieved a higher F-measure than other existing methods for all the four datasets. The RFE-FNN method obtained 93.62% for PID dataset, whereas the FR-RFA and WT-DBN achieved 78.15% and 91.83% F-Measure. In BC dataset, the proposed method achieved 99.07% F-measure, where FR-RFA achieved 98%. The analysis of the proposed RFE-FNN method of classification of four datasets shows that RFE-FNN method is highly efficient compared to other existing methods such as SVM, Naive Bayes, RBF and RFA. This shows that the proposed RFE-FNN method avoids the overfitting of training data which can be applicable for effective classification performance.

5. Conclusion

To solve the MDC problems, a computationally effective and efficient technique i.e. the RFE-FNN method is implemented in this research work. The RFE method helps to decrease the irrelevant features from the datasets. The important features are selected in RL methods and FNN method is used to classify the MD. The issues of overfitting in the training data are addressed by developing the FNN with RL method, which is the main aim of the RFE-FNN method. The RFE-FNN method is evaluated with the four 'UCI respiratory medical dataset' such as Heart disease dataset, Breast dataset, Diabetic, and Parkinson's dataset. The experimental result shows that the RFE-FNN has the highest performance compared to the other existing methods. The proposed RFE-FNN method achieved 89.58% and 97.67% of AUC for Heart and Parkinson's datasets, whereas FR-RFA achieved 87.05% and 96.13% of AUC for the same dataset. In future, the RFE-FNN method is further improved by using optimization techniques for achieving higher accuracy in the standard UCI dataset.

References

- [1] T. Nguyen, A. Khosravi, D. Creighton, and S. Nahavandi, "Medical data classification using interval type-2 fuzzy logic system and wavelets", *Applied Soft Computing*, Vol.30, pp. 812-822, 2015.
- [2] G. Gui, H. Pan, Z. Lin, Y. Li, and Z. Yuan, "Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection", *KSCCE Journal of Civil Engineering*, Vol.21, No.2, pp. 523-534, 2017.
- [3] S. Jadooki, D. Mohamad, T. Saba, S. Almazayad, and A. Rehman, "Fused features mining for depth-based hand gesture recognition to classify blind human communication", *Neural Computing and Applications*, Vol.28, No.11, pp.3285-3294, 2017.
- [4] C. Fernandez-Lozano, J. A. Seoane, M. Gestal, T.R. Gaunt, J. Dorado, and C. Campbell, "Texture classification using feature selection and kernel-based techniques", *Soft Computing*, Vol.19, No.9, pp. 2469-2480, 2015.
- [5] S.K. Satapathy, D. Satchidananda, and J. Alok Kumar, "EEG signal classification using PSO trained RBF neural network for epilepsy identification", *Informatics in Medicine Unlocked*, Vol. 6, pp.1-11, 2017.
- [6] A.M. Abdel-Zaher and A.M. Eldeib, "Breast cancer classification using deep belief networks", *Expert Systems with Applications*, Vol.46, pp.139-144, 2016.
- [7] V. Chaurasia and P. Saurabh, "Data mining techniques: to predict and resolve breast cancer survivability", *International Journal of Computer Science and Mobile Computing IJCSMC*, Vol.3, No.1, pp.10-22, 2014.
- [8] A. Husham, M. Hazim Alkawaz, T. Saba, A. Rehman, and A.J. Saleh, "Automated nuclei segmentation of malignant using level sets", *Microscopy research and technique*, Vol.79, No.10, pp.993-997, 2016.
- [9] T. Saba, S. Al-Zahrani, and A. Rehman, "Expert system for offline clinical guidelines and treatment", *Life Science Journal*, Vol.9, No.4, pp.2639-2658, 2012.
- [10] R. Varatharajan, M. Gunasekaran, and M.K. Priyan, "A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing", *Multimedia Tools and Applications*, Vol.77, No.8, pp.10195-10215, 2018.
- [11] S. Vijayarani and S. Dhayanand, "Liver disease prediction using SVM and Naïve Bayes algorithms", *International Journal of Science*,

- Engineering and Technology Research (IJSETR)*, Vol.4, No.4, pp. 816-820, 2015.
- [12] C. Soguero-Ruiz, K. Hindberg, I. Mora-Jiménez, J.L. Rojo-Álvarez, S.O. Skrøvseth, F. Godtliebsen, and R. Jenssen, “Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods”, *Journal of Biomedical Informatics*, Vol.61, pp.87-96, 2016.
- [13] L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang, H. Li, and D. Liu, “Evolving support vector machines using fruit fly optimization for medical data classification”, *Knowledge-Based Systems*, Vol.96, pp.61-75, 2016.
- [14] N. Kausar, A. Abdullah, B.B. Samir, S. Palaniappan, B.S. AlGhamdi, and N. Dey, “Ensemble clustering algorithm with supervised classification of clinical data for early diagnosis of coronary artery disease”, *Journal of Medical Imaging and Health Informatics*, Vol.6, No.1, pp.78-87, 2016.
- [15] V. Chaurasia, P. Saurabh, and B. B. Tiwari, “Prediction of benign and malignant breast cancer using data mining techniques”, *Journal of Algorithms & Computational Technology*, Vol. 12, No.2, pp.119-126, 2018.
- [16] A. Khatami, A. Khosravi, T. Nguyen, C.P. Lim, and S. Nahavandi, “Medical image analysis using wavelet transform and deep belief networks”, *Expert Systems with Applications*, Vol.86, pp.190-198, 2017.
- [17] M.Z. Alam, M. Saifur Rahman, and M. Sohel Rahman, “A Random Forest based predictor for medical data classification using feature ranking”, *Informatics in Medicine Unlocked*, Vol.15, pp.100180, 2019.
- [18] J. Zhang, M. Liu, and D. Shen, “Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks”, *IEEE Transactions on Image Processing*, Vol.26, No.10, pp.4753-4764 2017.
- [19] M. Mahmud, M.S. Kaiser, A. Hussain, and S. Vassanelli, “Applications of deep learning and reinforcement learning to biological data”, *IEEE Transactions on Neural Networks and Learning Systems*, Vol.29, No.6, pp.2063-2079, 2018.
- [20] C.L. Blake, and C.J. Merz, “UJI Repository of Machine Learning Databases”, <http://www.ics.uci.edu/~mllearn/MLRepository.html>, 1996
- [21] Y. Al-Dunainawi, F. Maysam, and A. Jizany, “A new MIMO ANFIS-PSO based NARMA-L2 controller for nonlinear dynamic systems”, *Engineering Applications of Artificial Intelligence*, Vol. 62, pp. 265-275, 2017.