FEATURE SELECTION AND FUZZY EXTREME LEARNING MACHINE (FELM) CLASSIFIER FOR HEART DISEASE DIAGNOSIS

G.Sunitha, Research Scholar, Department of Computer Science, Rayalaseema University, Kurnool. E-Mail: gundasunitha01@gmail.com
Dr.N.Geethanjali, Professor, Department of Computer Science & Technology, Sri Krishnadevaraya University, Anantapur.

ABSTRACT: The considerable growing of cardiovascular disease and its effects and complications as well as the high costs on society makes medical community seek for solutions to prevention, early identification and effective treatment with lower costs. Thus, valuable knowledge can be established by using artificial intelligence and data mining; the discovered knowledge makes improve the quality of service. Until now, different researches have been carried out in order to predict heart disease based on data mining methods such as classification and feature selection methods; however, what has been less noticed is the exact diagnosis of disease with the lowest cost and time. Early detection and treatment of heart disease will reduce the patient mortality rate. Accordingly, herein propose a Particle Swarm Optimization (PSO) and highly accurate hybrid Fuzzy Extreme Learning Machine (FELM) method for the diagnosis of coronary artery disease. The proposed FELM based prediction model is able to detect coronary artery disease based on clinical data without the need for invasive diagnostic methods. Making use of such methodology, we achieved good accuracy, sensitivity and specificity rates on Z-Alizadeh Sani dataset.

Keywords: Diagnosis Systems, Heart Disease, Feature Selection, Fuzzy Extreme Learning, Machine Learning, Coronary Artery Disease.

1. INTRODUCTION

Data mining is the process of examining the raw data set to generate useful information. It looks for pattern of data according to different perspectives [1]. It is capable of delivering the effective course of action by comparing and contrasting the data. The data mining is largely used in retail, financial communication and marketing organization to drill down their transactional data. Machine learning plays a huge role in medical field [7].

The medical domain is the great beneficiary of data mining where huge data are handled easily [1,5,7]. There still prevails a great challenging common disease called Coronary Artery Disease (CAD) which takes off more life in the world. The deadly disease is caused due to the constriction of the blood vessels in the heart. The Angiography is the
widely used diagnosis method which might place the life at risk by causing cancer and other health issues [1]. Some of works have been done in the recent work is described as follows:

Zeinab Arabasadi et al., [1] proposed a four ranking methods such as Gini index, weight by SVM, information gain and Principal component analysis (PCA). Thus the model is based on artificial neural network and genetic algorithm to diagnose the heart disease by clinical means there by eliminating the angiography.

Kale et al., [2] proposed a methodology for handling weighted classification problem and Feature Subset Selection (FSS) for F-ELM classifier. Finally these comparisons resulted in validation of effectiveness and accuracy. Kavipriya et al., [3] proposed the system that uses comprehensive feature selection which is applicable for selection of attributes that had chosen the dataset of PIMA Indian Diabetes under SVM classifier.

Chunhong Lue et al [4] had used trace-based separable criterion developed a selection algorithm on the basis of Genetic Algorithm (GA). Finally the comparatively obtained result brings out this proposed system has been providing better lung cancer diagnosis. Xiao Liu et al [5] proposed the system that uses the Rough Set (RFRS) and novel relief feature selection (ReliefF) based for diagnosis the heart disease. It brought out that superior performance is acquired by RFRS base on specificity and sensitivity. Thus proposed system provides the tool for heart disease diagnosis based on empirical analysis.

Kale et al [6] bring out the system called Improved Genetic – Particle Swarm Optimization (IG-PSO) algorithm based on optimal features returned by Extreme Learning Machine (ELM) as well as it provides an optimal features. Subbulakshmi et al [7] proposed a methodology on machine learning called hybrid methodology. This helps to integrate the PSO algorithm along with ELM classifier with UCI Machine Learning Repository. Thus results helps to achieve better performance.

Neol Perez Perez [8] aims to enhance the feature selection methods and the performance of AUC which is based on classifier especially for breast cancer. To increase the various, then enhancing the performance in various domains and made the uFilter method is made to extend and allow to use on various problems such as multiclass classification.

Nithin Kumari et al [9] introduced the system that deals beyond the analysis on health care. On comparing these all techniques authors conclude that Neuro-fuzzy approach is best among other two techniques especially for diagnosing the heart disease. Arun Kumar et al [10] proposed the system for detecting and classifying the brain tumor based on human by the help of brain images. This system uses Support Vector Machine, PCA and K-mean. Thus it
provides higher results for tumor detection. In addition K-mean helps for Image Segmentation.

Therefore a new hybrid model [1] based on artificial neural network and genetic algorithm has been proposed here and is used for detecting the heart disease without employing the angiography. The neural network helps to improve the performance and the genetic algorithm provides the best possible outcome as solution. The rest of the paper is organized as follows. Section 2 discusses the literature review. Section 3 overviews the proposed technique for feature selection and classification. Experimental results of the proposed scheme are presented in Section 4. Concluding remarks with future work are covered in Section 5.

2. PROPOSED SYSTEM

This study combines the Fuzzy Extreme Learning Machine (FELM) and PSO based on the classifier. In this process, the optimal solution from the dataset is obtained by utilising the PSO. For the classification of the data, FELM is employed. This system is further divided into two phases:

- Phase 1: Feature Selection using PSO
- Phase 2: Classification of Selected Attributes Using Fuzzy Extreme Learning Machine (FELM) based classifier

2.1 Feature selection using particle swarm optimization

The computation technique which is evolutionary, called Particle Swarm Optimisation is inspired by the social behaviour [11]. Swarm is used for monitoring the particle population by PSO and also in search space it encodes the candidate solution. The random position initialisation in the space and iteration of each particle’s position depending on its neighbours and experience is initialised by PSO. The vector \( x_i = (x_{i1}, \ldots, x_{in}) \) represents the position of the particle and \( n \) stands for search space dimension. The vector \( v_i = (v_{i1}, \ldots, v_{in}) \) denotes the velocity and the predefined range \([-v_{max}, v_{max}]\) limits the vector of each component. The personal best \( pbest_i = (p_{i1}, \ldots, p_{in}) \) in which the particle \( i \) is represented as the best previous position (based on the few fitness function) and \( gbest = (g_1, \ldots, g_n) \), the global best is the one in the population as a whole is found by best position and recorded. The position of each particle and velocity of the particle is updated by PSO at each iteration and the following equation defines that:
\[ x_{i,d}^{t+1} = x_{i,d}^t + v_{i,d}^{t+1} \]  
\[ v_{i,d}^{t+1} = w \cdot v_{i,d}^t + c_1 \cdot r_{1,i} \cdot (p_{i,d} - x_{i,d}^t) + c_2 \cdot r_{2,i} \cdot (g_d - x_{i,d}^t) \]  

The velocity vector or the component’s position is represented by \( 0 < d \leq n \), \( t \)-th iteration is denoted by \( t \) in the algorithm, the inertia weight’s predefined constant is represented by \( w \), constants for predefined acceleration is given by \( c_1 \) and \( c_2 \). The random values which are uniformly distributed over \([0,1]\) is represented by \( r_{1,i} \) and \( r_{2,i} \). The search spaces of the real value is applicable to the description of PSO. The modified algorithm is in demand as the issues occur along with the feature selection in the discrete search in which the Binary PSO has also been included. The restriction of the values to 0 or 1 by the values of all position vectors (\( x_i, p_{best,i}, \text{and } g_{best,i} \)) of the components prevails in binary PSO. The corresponding component with the probability in the position vector of each component is being 1 in the Equation (2) and is employed for updating velocity. For the transformation into a unit range, sigmoid function \( s(v_i,d) \) is used. Based on the following equation the position of the each particle is updated by binary PSO.

\[ x_{i,d} = \begin{cases} 1, & \text{rand}() < s(v_{i,d}) \\ 0, & \text{otherwise} \end{cases} \]

Where

\[ S(\lambda) = \frac{1}{1+e^{-\lambda}} \]

According to the fitness function PSO places its focus on searching of best classification performance in the feature selection. Besides the performance of classification, PSO demands the assistance for searching the feature subset which is off small size and classification has high accuracy since in most cases, search space is large. Thus the size-controlled PSO is proposed in which the target size guides for search of PSO. The updating equation’s velocity is produced by the influence of \( T \) and is given in the equation (4).

\[ v_{i,d}^{t+1} = w \cdot v_{i,d}^t + c_1 \cdot r_{1,i} \cdot (p_{i,d} - x_{i,d}^t) + c_2 \cdot r_{2,i} \cdot (g_d - x_{i,d}^t) + c_3 \cdot r_{3,i} \cdot S(T - |p|) \]

where the particles \( i \) for the selection of number of particles is represented by \( |p| \), third generation constant is given by \( c_3 \), uniform distribution of random values \( r_{3,i} \) over \([0,1]\).  

**PSO:**  
**Input:** Data Set \( D \)  
- Divide Dataset into a training set and a test set  
**Output:** Selected Featured Subset
Algorithm:
Random initialization of velocity and position of each particle

1. Do, while the criterion for stopping is not met
2. Fitness evaluation on training set of each particle
3. For each particle p do
4. Update the pbest and gbest of p
5. End for
6. For each particle p do
7. Update the velocity and position of p
8. End for
9. End while
10. Return the position of gbest (the selected feature subset)

2.2 Fuzzy Extreme Learning Machine (FELM) based classifier
This work uses the FELM. Thus combination of relative based metrics of Fuzzy sets and an Extreme based Machine Learning has been done by using classifier. Clinical datasets has been used in proposed System to carry out the above said processes. Mainly three types of subsystems are employs in this proposed system. These Subsystems are in the framework named FELM. The subsystems include fuzzification subsystem, classification subsystem and preprocessing subsystem. The fuzzification subsystem which is one among the three subsystems will map every feature to each fuzzy set. In addition, classification subsystem includes the process of using the various classification algorithms available in extreme machine based learning in order to perform classification.

Fuzzification subsystem
Transformation of features has been performed by using function namely Trapezoidal membership. This process of transformation is applied on clinical datasets which is specially selected in order to obtain fuzzy set additionally with the membership value. Thus fuzzification over clinical datasets is performed using functions of membership.

Classification subsystem
Classification subsystem involves mainly two approaches such as classifier construction and testing. This research aims on the classification which is carry out using Feed forward based neural network. This Feed Forward Neural Network performs classification with the help of
separate hidden layer using ELM. ELM stands for Extreme Learning Machine is used in order to identify the weights between output and hidden layer neurons.

Thus amount of neurons in input, hidden and output layers are symbolically represented with the help of notations such as p, q, and r. The Weight vector is measured between hidden and input layer neurons by using the representing as $W_{ih}$. Additionally $W_{ho}$ is used to represents the weight vector in between the neurons of hidden layer and output layer. The fuzzified based clinical datasets obtains the value that expected is represented as T. Hidden layered neurons uses the function called Sigmoid activation. For each value of q which means hidden layer neurons, the description of SLFNN’s training has been obtained as follows,

Step 1: FELM is allowed to take fuzzified features as input which is clinical dataset.

$$I_i = X_i^a = 1, 2, ..., p$$  \hspace{1cm} (5)

Where, the notation p used to refer the total amount of fuzzified features in X which means clinical dataset.

Step 2: Randomly initialization of weights between input and hidden layer neurons has been applied which ranges from 0 to 1 and can be represented in $W_{ih}$, Where the letters i and h specifies the neurons such as input layer and hidden layer. On the other hand i have the values that ranging between 1 and amount of fuzzified features which includes in the clinical based data set. The value of letter h, ranging between 1 and q ie number of neurons especially in hidden layer.

Step 3: By using Equation 6, Computation of hidden layer neurons input has been done and denoted as $H^i_j$,

$$H^i_j = \sum_{i=1}^{p} (I^i_i W_{ij}^{ih})j = 1, 2, ..., q$$  \hspace{1cm} (6)

$I_i$ denotes the output of input layer neurons, whereas $W_{ij}^{ih}$ denoted as weight between the hidden and input layer of neuron.

Step 4: Using 7th equation, $H^o_j$ means hidden layer’s output has been computed;

$$H^o_j = \frac{1}{1+e^{-H^i_j}}j = 1, 2, ..., q$$  \hspace{1cm} (7)

Step 5: Between the hidden layer and output layers of neurons, the weights such as $W^{ho}$ has been determined with the help of ELM method as in equation 8.

$$W = H^\dagger T$$  \hspace{1cm} (8)

Where $H^\dagger$ is Moore-Penrose’s generalized inverse of H.
Here, T is used to denote the target class and H represents the output value of hidden layer neuron. Weights which could make connection between hidden layer neurons and the output layer neurons has made and said as W.

\[ T = HW \]  \hspace{1cm} (9)

Step 6: Equation 10 helps to get the value for output layered neuron shortly \( O_k \).

\[ O_k = f\left(\sum_{j=1}^{q} (H_j W_{jk}^{ho})\right) \quad k=1,2,...n \]  \hspace{1cm} (10)

Let f denotes the activation function that we used; number of hidden layer neurons shortly as q and number of dataset used for training is represents as n.

2.3 PSO+FELM

In this section, we describe the proposed PSO-FELM classification method for the diagnosis of coronary artery disease cardiac auscultation. The aim of this system is to optimize a set of weighting factors for the feature set, such that the highest accuracy of the classifier can be achieved.

1. Initialization
   - Generate an initial swarm of size S.
   - Set the velocity vectors \( v_i \) (i=1,2,...,S) for each particle in the swarm with a value of zero.
   - Train an Fuzzy Extreme Learning Machine (FELM) based classifier and compute the corresponding fitness function \( f(i) \) (i.e. the accuracy) of each position \( p_i(t) \) for each \( p_i(t) \) in the swarm.
   - Select the best position from each particle with its initial position as in (11):

\[ P_{bi} = P_i, (i = 1,2,...,S) \]  \hspace{1cm} (11)

2. Optimization process
   - Determine the best global position \( p_g \) from particles in the swarm by the fitness function over all explored trajectories.
   - Update the velocity and position of each particle.
   - Train an Fuzzy Extreme Learning Machine (FELM) based classifier and compute the corresponding fitness function \( f(i) \) for each candidate particle \( p_i \) (i=1,2,...,S)
   - Update the best position \( p_{bi} \) of each particle, if the corresponding position has a smaller fitness function value.
Classify the given data with the trained Fuzzy Extreme Learning Machine (FELM) based classifier.

3. Stopping Criteria

- If not at maximum iteration, repeat the optimization process. Otherwise, continue step

3. EXPERIMENTAL DESIGN

Z-Alizadeh Sani dataset is considered from the UCI machine learning repository as benchmark problems to evaluate the performance of proposed method. Table 1 shows the dataset characteristics of the Z-Alizadeh Sani data set. In the experimental design, Accuracy, Sensitivity and Specificity parameters are used. It is evaluated with our proposed method and compared with existing methods.

Table 1. Data Set Characteristics

<table>
<thead>
<tr>
<th>Name of the Dataset</th>
<th>Z-Alizadeh Sani dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Instances</td>
<td>303</td>
</tr>
<tr>
<td>Number of Features</td>
<td>54</td>
</tr>
<tr>
<td>Number of Groups</td>
<td>4 (Demographic, Symptom and examination ECG, Echo feature)</td>
</tr>
<tr>
<td>Classification</td>
<td>Each patient could be in two possible categories Normal or Not based his/her diameter.</td>
</tr>
</tbody>
</table>

The selected features and their weights of the Z-Alizadeh Sani dataset samples are shown in the table 2.

Table 2. Z-Alizadeh Sani dataset selected feature weights

<table>
<thead>
<tr>
<th>Features</th>
<th>Sex</th>
<th>BMI</th>
<th>DM</th>
<th>HTN</th>
<th>Current Smoker</th>
<th>EX-Smoker</th>
<th>CVA</th>
<th>Airway disease</th>
<th>Thyroid Disease</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weights</td>
<td>0.25</td>
<td>0.01</td>
<td>0.15</td>
<td>0.18</td>
<td>0.23</td>
<td>0.15</td>
<td>0.12</td>
<td>0.08</td>
<td>0.05</td>
<td>0.16</td>
</tr>
<tr>
<td>Features</td>
<td>DLP</td>
<td>BP</td>
<td>Edema</td>
<td>Diastolic Murmur</td>
<td>Atypical</td>
<td>Nonanginal</td>
<td>Exertional CP</td>
<td>LowTHAng</td>
<td>St Depression</td>
<td>BBB</td>
</tr>
<tr>
<td>Weights</td>
<td>0.17</td>
<td>0.085</td>
<td>0.106</td>
<td>0.112</td>
<td>0.136</td>
<td>0.056</td>
<td>0.067</td>
<td>0.71</td>
<td>0.82</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Performance Metrics

The following are the performance metrics used up for the evaluation of the model.

- **A confusion matrix** is a classification table used to describe the performance of a classifier.
- **True positive (TP):** The proportion of actual positives that are correctly identified.
• **False positive (FP):** The proportion of actual positives that are in correctly identified.
• **True negative (TN):** The proportion of negatives that are correctly identified.
• **False negative (FN):** The proportion of negatives that are incorrectly identified.
• **Positive Predictive Value (PPV)** is the probability that the disease is present given a positive test result.
• **Positive Predictive Value (NPV)** is the probability that the disease is absent given a negative test result.
• **Sensitivity** defines how well the classifier can identify a diseased heart correctly, also called true positive rate.
  \[ Sensitivity = \frac{TP}{TP + FN} \]
• **Specificity** means the tests ability to exclude healthy heart from diseased heart correctly.
  \[ Specificity = \frac{TN}{FP + TN} \]
• The **accuracy** results show the proportion of correctly classified instances and incorrectly classified instances.

\[
\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Negative} + \text{False Positive} + \text{True Negative}}
\]

**Results and Discussion**

The proposed approach FELM is compared with three algorithms. Naïve Bayes (NB), Neural Network (NN) and Hybrid Neural Network - Genetic Algorithm (HNN-GA) algorithms are taken into consideration.

**Table 3. Performance Analysis**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>No Feature Reduction(%)</th>
<th>Feature Reduction(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>79.61</td>
<td>87.3</td>
</tr>
<tr>
<td>Neural Network</td>
<td>81.55</td>
<td>84.5</td>
</tr>
<tr>
<td>HNN-GA</td>
<td>83.49</td>
<td>84.9</td>
</tr>
<tr>
<td>FELM</td>
<td>85.44</td>
<td>83.5</td>
</tr>
</tbody>
</table>
Table 3 shows the performance analysis of four algorithms. All four algorithms applied based on 2 scenarios. In the first scenario, no features are reduced. In the second scenario Feature Reduction (FR) done using PSO.

![Figure 1. Comparison of Accuracy for Different Algorithms](image1)

![Figure 2. Comparison of Sensitivity for Different Algorithms](image2)
Comparison of accuracy, sensitivity and specificity for different algorithms are presented in Figure 2, figure 3 and Figure 4. Compared to the previous methods, the proposed method shows an improvement of classification in the case of correctly classifies patients. In all cases, the proposed system outperforms existing methods, except sensitivity in the no feature reduction scenario. However, the system cannot perfectly classify all the instances. It could be concluded that the method proposed by this study obtains promising results for Z-Alizadeh Sani dataset classification system.

4. CONCLUSION AND FUTURE WORK

Hybridization of Particle Swarm Optimization (PSO) and highly accurate hybrid Fuzzy Extreme Learning Machine (FELM) is proposed for the diagnosis of coronary artery disease in this paper. In the proposed algorithm, the performance of the hybrid Fuzzy Extreme Learning Machine is improved by applying PSO to FELM. The Z-Alizadeh Sani dataset obtained from the UCI machine learning repository is used for evaluating the performance of the proposed method. The performance of the proposed FELM algorithm is estimated based on classification accuracy, sensitivity and specificity. The proposed algorithm FELM with PSO provides better classification results. In future, EFS might combine the results of three feature selection methods such as Bat Algorithm (BA), and Firefly Algorithm (FA) which give a better approximation to the optimal subset or ranking of features. Parallel Deep
Learning Algorithm (PDLA) based prediction model can be proposed to detect coronary artery disease based on clinical data without the need for invasive diagnostic methods.

REFERENCES