Informatics in Medicine Unlocked 2 (2016) 1–11

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets

Kindie Biredagn Nahato [a], Khanna H. Nehemiah [a,*], A. Kannan [b]

[a] Ramanujan Computing Centre, Anna University, Chennai 600025, India
[b] Department of Information Science and Technology, Anna University, Chennai 600025, India

ARTICLE INFO

Article history:
Received 31 October 2015
Received in revised form
18 December 2015
Accepted 9 January 2016
Available online 15 February 2016

Keywords:
Extreme learning machine
Fuzzification
Fuzzy set
Classification
Euclidean distance
Membership function

ABSTRACT

Data mining techniques play a major role in developing computer aided diagnosis systems and expert systems that will aid a physician in clinical decision making. In this work, a classifier that combines the relative merits of fuzzy sets and extreme learning machine (FELM) for clinical datasets is proposed. The three major subsystems in the FELM framework are preprocessing subsystem, fuzzification subsystem and classification subsystem. Missing value imputation and outlier elimination are handled by the preprocessing subsystem. The fuzzification subsystem maps each feature to a fuzzy set and the classification subsystem uses extreme learning machine for classification.

Cleveland heart disease (CHD), Statlog heart disease (SHD) and Pima Indian diabetes (PID) datasets from the University of California Irvine (UCI) machine learning repository have been used for experimentation. The CHD and SHD datasets have been experimented with two class labels one indicating the absence and the other indicating the presence of heart disease. The CHD dataset has also been experimented with five class labels, one class label indicating the absence of heart disease and the other four class labels indicating the severity of heart disease namely low risk, medium risk, high risk and serious. The PID data set has been experimented with two class labels one indicating the absence and the other indicating the presence of gestational diabetes.

The classifier has achieved an accuracy of 93.55% for CHD data set with two class labels; 73.77% for CHD data set with five class labels; 94.44% for SHD data set and 92.54% for PID dataset.

1. Introduction

The major silent killer diseases are heart disease and diabetes [1]. Cardiovascular diseases (CVD) refer a group of disorders of the heart and blood vessels. Diabetes is one of the risk factor for CVD [2]. Diabetes is a chronic disease that is a consequence when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin produced by the pancreas [3].

"Clinical decision support systems (CDSS) are computer systems designed to impact clinician decision making about individual patients at the point in time that these decisions are made" [4]. CDSS focuses on increasing the accuracy of decision making and decreases the processing time and cost. Data mining algorithms can be used for developing CDSS. Data mining encompasses statistical analysis, machine learning techniques to discover useful and previously unknown patterns from voluminous amount of data from databases [5,6]. The major data mining functionalities are association rule mining, classification and clustering [5,7]. Association rule mining discovers interesting relationship between items. The interestingness of the relationship is measured using two metrics, namely, support and confidence [7]. Classification is the process of developing a model that describes for the purpose of being able to use the developed model to distinguish or predict the class of objects whose class label is unknown [7]. Clustering is performed on a dataset to categorize into a group by maximizing the similarity and minimizing the difference in the group [7]. The learning technique used in classification is supervised whereas in clustering it is unsupervised. In this work, first each clinical dataset is preprocessed for handling missing values and outliers. Missing values are imputed and instances with outlier value(s) are eliminated from the clinical dataset. Second fuzzification is performed on the preprocessed data and third the classifier is modeled using extreme learning machine (ELM).

2. Background of fuzzy extreme learning machine

Fuzzy extreme learning machine (FELM) combines the advantage of ELM and fuzzy set theory. ELM, a learning algorithm

* Corresponding author. Tel.: +91 44 22358013.
E-mail address: nehemiah@annauniv.edu (K.H. Nehemiah).

http://dx.doi.org/10.1016/j.imu.2016.01.001

developed by Huang et al. [8], is applied in a single layer feed-forward neural network (SLFNN) where the weights between the input layer neurons and hidden layer neurons are randomly generated and the weights between hidden layer neurons and output layer neurons are analytically determined through simple generalized inverse operations [8,9]. ELM overcomes the limitation of the popularly used backpropagation learning algorithm in SLFNN. Backpropagation neural network (BPNN) learning algorithm is either very slow due to improper learning rate or easily converges to local minima. Besides this, BPNN needs many iterative learning steps to accomplish the learning task. ELM has better generalization performance compared to BPNN and it tends to reach the solutions without using parameters like, learning rate, momentum rate as that of backpropagation learning algorithms. In ELM once the weights between the input layer neurons and the hidden layer neurons are randomly generated, the weights will not be iteratively tuned or adjusted as in BPNN [9]. This significantly reduces the time taken to train the network.

In SLFNN the output value of the output layer neurons (O_k) can be computed using the value of the hidden layer neurons (H_j) and the connecting weights (W_{jk}^{ho}) as follows

$$O_k = f\left(\sum_{j=1}^{q}\left(H_j W_{jk}^{ho}\right)\right) \quad k = 1, 2, \ldots n \tag{1}$$

Where, f is the activation function, $q$is the number of hidden layer neurons, n is the total number of training dataset, Using the identity function, O_k becomes the summation of the product of H_j and W_{jk}^{h0} as follows

$$O_k = \sum_{j=1}^{q}\left(H_j W_{jk}^{ho}\right) \quad k = 1, 2, \ldots n \tag{2}$$

The objective of ELM neural network is to minimize the error between output value (O_k) and the target class (T_k). Using the approximate zero error mean given by $\sum_{k=1}^{n} \|O_k - T_k\| \cong 0$, hence, Eq. (2) can be written compactly as

$$T = HW \tag{3}$$

Where, T is the target class, H is the output value of the hidden layer neurons, W is the weights that connects the hidden layer neurons and the output layer neurons. Then the unknown weights (W) can be computed as

$$W = H^{\dagger}T \tag{4}$$

where $H^{\dagger}$ is Moore-Penrose's generalized inverse of H.

Fuzzy set theory was introduced by Zadeh [10] for handling uncertainty. Fuzzification refers to the process of mapping each feature in the clinical dataset to a fuzzy set with a degree of membership ranging from 0 to 1 [11,12]. Each feature is represented by two or more linguistic variables. For example the feature Diastolic blood pressure can be represented as a fuzzy set with three members namely Hypotension, Normal and Hypertensive. Clinical datasets have uncertainty; hence fuzzy set theory is used to resolve the uncertainty problem. In this study, each feature value of the instance is represented by the membership value of the corresponding linguistic variables.

FELM was used by Zhang et al. [13] for weighted classification problem. In their proposed method, fuzzy set theory has been used for weighting the instances of dataset based on the number of distinct class labels. For example the dataset with three class labels has the weight values of 0.5, 0.3 and 0.2. The summation of the given weights becomes 1. Their FELM was tested by using only a fixed number of hidden layer neurons.

FELM takes the advantage of fuzzification and ELM. Fuzzification of features of the clinical dataset helps to get higher performance accuracy and the learning process of ELM helps to obtain not only higher accuracy, but also reduces the training time.

In this research work, the FELM classifier was developed and tested with a varying number of hidden layer neurons. The first classifier has used 10 hidden layer neurons, and increase by one for the second classifier. The increment is terminated when the hidden layer neurons becomes 200. The classifier with highest performance is selected.

The rest of the paper is organized as follows. The description of related work carried out by other researchers is presented in Section 2. In Section 3, the system framework of the proposed work is discussed. Experimental results and comparison of the proposed work with works carried out by other researchers is discussed in Section 4. Conclusion and scope for future work is discussed in Section 5.

3. Related work

Related works carried out by other researchers using clinical data sets taken from the UCI machine learning repository is discussed in this section.

Aslam et al. [14] in their work have used Pima Indian Diabetes (PID) dataset for diagnosing the presence or absence of diabetes. The researchers have carried out their work in three stages. In stage one, the diabetes features have been normalized to zero mean and unit standard deviation. Student's t-test, Kolmogrov–Smirnov test, f-score selection, Kullback–Leibler divergence and genetic programming (GP) have been employed to assess the effectiveness of the normalized features. The features were arranged in decreasing order of importance based on the above tests and different subsets of features were prepared using sequential forward selection (SFS) process. In stage two they have used GP with comparative partner selection to generate new features for each subset of features prepared by SFS. In stage three they have tested the performance of the features generated in stage two using KNN and SVM classifiers. They have achieved a classification accuracy of 80.5% using GP-KNN with ten-fold cross validation and 87% using GP-SVM. The researchers have not dealt with the uncertainty and vagueness of the feature value in the dataset. Furthermore their classification approach has been tested over only one dataset which may not generalize over the other clinical dataset.

Patil et al. [15] proposed a hybrid approach by combining K-means clustering algorithm and C4.5 for classifying of Pima Indian diabetes (PID) dataset. Their proposed system has three steps. First, the data has been preprocessed by removing inappropriate and inconsistent data. Due to the 0 values associated in the PID dataset, the researchers have removed two features namely serum-insulin and triceps skin fold, and 143 instances from the dataset. After preprocessing, PID dataset is reduced from 768 to 625 instances and from 8 to 6 features. Z-score method was applied to normalize the reduced PID. Second, patterns have been extracted using K-means clustering algorithm. The incorrectly clustered patterns were removed; thereby the dataset was reduced to 433 instances. Third, a decision tree model has been constructed using the extracted patterns. They have achieved a classification accuracy of 92.38% using ten-fold cross validation. This work suffers from overfitting problem as their proposed clustering technique eliminates 192 instances (about 30% of the preprocessed dataset) that are incorrectly clustered.

Alneamy et al. [16] in their work have used teaching learning based optimization (TLBO) and fuzzy wavelet neural network (FWNN) for diagnosis of heart disease. They have used Cleveland heart disease (CHD) dataset. Gaussian membership function has been used for fuzzification. TLBO is applied to update the weight of

FWNN and the wavelet parameter. They have used five-fold cross validation and scored accuracy of 90.29%. Their work is limited on classifying heart disease as present or absent. It does not predict the severity of heart disease.

Varma et al. [17] have used decision tree to classify the presence or absence of diabetes in Pima Indian diabetes dataset. They have remove all instances associated with 0 values and reduced the number of instances from 768 to 336. The split point has been computed by using Gini index with a fuzzy decision boundary. Gaussian membership function has been used for fuzzification. Three-fold cross validation has been used and an accuracy of 75.8% has been achieved. They have stated that the performance of their work can be enhanced by carrying out experiments using other fuzzy membership functions.

Nahato et al. [18] have proposed a disease diagnosis system using rough set theory and backpropagation neural network (RS-BPNN). The proposed method was tested using Statlog Heart Disease (SHD), Wisconsin breast cancer disease (WBCD) and hepatitis dataset. After handling missing value, rough set theory indiscernibility relation method has been applied for feature selection. The BPNN has been trained using the selected feature set. A classification accuracy of 97.3%, 98.6%, and 90.4% for hepatitis, WBCD and SHD has been achieved. The number of neurons in the hidden layer is fixed, however better result may be obtained by varying the hidden layer neurons in the neural network.

Seera et al. [19] in their work have proposed a hybrid intelligent system for classifying clinical data using fuzzy min–max neural network, classification and regression tree (CART), and random forest model. Their proposed method was tested using Wisconsin breast cancer disease (WBCD), Pima Indian diabetes (PID) and liver disorder datasets. They have used 90% of the dataset for training and the remaining 10% of the dataset was used for testing. They have achieved an accuracy of 98.84%, 78.39%, and 95.01% for WBCD, PID and liver disorder dataset respectively. The researchers have not discussed how missing value is handled in the datasets.

Dennis et al. [20] in their work have used Adaptive Genetic Fuzzy Systems (AGFS) for classification. First, the dataset has been discretized using minimum and maximum values of each feature for every class label. They have used genetic algorithm for optimized rule generation. Triangular membership function has been used to transform the features to a fuzzy set. Mamdani fuzzy inference system has been used for classification. They have tested their work with seven datasets among which four datasets are clinical datasets. For training 90% of the dataset has been used and for testing 10% of the dataset has been used. The accuracy achieved for the clinical datasets Cleveland heart disease, Pima Indian diabetes, Indian liver data and Mammogram dataset is 76.67%, 89.80%, 75.86% and 57.29% respectively. Their work neither highlighted the effect of preprocessing nor analyzed the clinical relevance of the generated rules.

Kalpana et al. [21] have proposed a fuzzy expert system for diagnosis of diabetes. They have used Pima Indian diabetes (PID) dataset. In their work, they have selected five features namely glucose, insulin, body mass index, diabetes pedigree function, and age of the patient for fuzzification using triangular membership function. Each feature has been transformed to three fuzzy sets namely low, medium, and high. The target class has been also transformed to fuzzy sets. Centroid method is used for defuzzification process. Their proposed method has achieved an accuracy of 90.38%. The researchers have not specified the data preprocessing technique used for handling missing values and how the five features were selected.

Christopher et al. [22] have used a rule based classifier for classifying clinical datasets. Their work has proposed swarm optimization approach for obtaining an optimal subset of rules from the set of rules generated using C4.5 algorithm. They have used six clinical datasets from the UCI repository. Their proposed approach has achieved accuracy of 77.89%, 83.99%, 94.88%, 92.88%, 64.20% and 82.05% for Cleveland Heart Disease, Pima Indians Diabetes, Wisconsin Breast Cancer, Hepatitis, Liver Disorders dataset, and Lymphoma dataset respectively. The researches have not analyzed the clinical relevance of the optimized rules.

Subbulakshmi and Deepa [23] in their work have integrated self regulated learning particle swarm optimization (SRLPSO) with extreme learning machine (ELM) classifier for disease diagnosis. Their proposed method has been experimented with five clinical datasets of the UCI machine learning repository. The PSO was designed to update the weights of input neurons and the bias value to increase the performance of the ELM classifier. The ELM learning algorithm has been applied in single hidden layer feed forward neural network for determining the weights that link the hidden layer neurons and output layer neurons. Their proposed method has achieved accuracy of 99.78%, 93.09%, 89.96%, 98.71% and 91.33% for Wisconsin Breast Cancer, Pima Indian diabetes, Statlog heart disease, hepatitis and Cleveland heart disease datasets respectively.

Kaya and Uyar [24] in their work have developed hybrid a decision support system using rough set theory and extreme learning machine (ELM) for classifying hepatitis. Rough set theory has been used for feature subset selection and ELM has been used to determine the weights that link the hidden layer neurons to output layer neurons of the single layer feed forward neural network. They have used hepatitis dataset from UCI machine learning repository. Each of the selected feature subset has been partitioned to training–testing as 50–50%, 70–30% and 80–20%. The ELM has been modeled using the corresponding number of features of the reducts; tangent sigmoid activation function has been used for obtaining the value of hidden layer neurons. Their ELM has been tested using various number of hidden layer neurons ranging from 10 to 100. The researchers have achieved 100% accuracy with the reduct of four features namely fatigue, malaise, protime, and histology with the data division of 80–20% for training–testing.

4. System framework

FELM framework proposed in this work has three major subsystems namely, preprocessing subsystem, fuzzification subsystem and classification subsystem. The FELM framework is illustrated in Fig. 1.

4.1. Clinical dataset

Cleveland heart disease (CHD), Statlog heart disease (SHD) and Pima Indian diabetes (PID) datasets from UCI machine learning repository have been used for this study [25].

The CHD dataset has 303 instances with a class label associated with each instance. There are 164 instances in class 0 (absence of heart disease), 55 instances in class 1 (low risk), 36 instances in class 2 (medium risk), 35 instances in class 3 (high risk), and 13 instances in class 4 (serious). In this study, the CHD dataset has been experimented with five class labels (CHD5) and with two class labels (CHD2). The CHD5 dataset has used the CHD dataset that is taken from the UCI machine learning repository. The CHD2 dataset is prepared by merging the risk level of the heart disease class labels to presence of heart disease class label.

The SHD dataset has 270 instances with a class label stating the presence or absence of heart disease. There are 150 instances in class 0 (absence of heart disease) and 120 instances in class 1 (presence of heart disease). Description of CHD and SHD dataset presented in Table 1.

PID dataset has eight features with 768 instances. All feature values of the PID are numerical data type. Each instance has a class

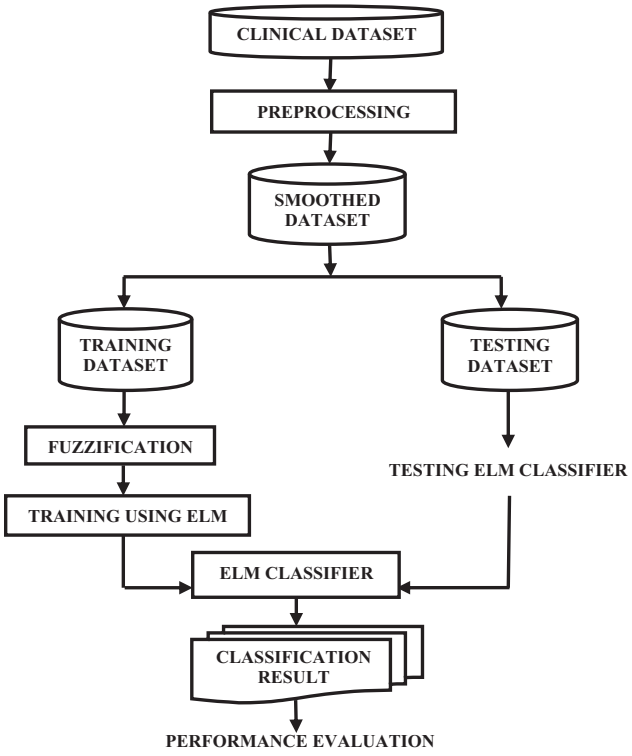

Fig. 1. System framework.

Table 1
Description of CHD and SHD dataset.

No	Feature	Description	Data type	Domain
1	Age	Patient age in year	Numerical	[29–77]
2	Sex	Gender	Binary	[0, 1]
3	Chp	Chest pain type	Nominal	[1,2, 3, 4]
4	Bp	Resting Blood pressure	Numerical	[94–200]
5	SCh	Serum Cholesterol	Numerical	[126–564]
6	Fbs	Fasting Blood sugar > 120 mg/dl	Binary	[0, 1]
7	ECG	Resting Electrocardiographic result	Nominal	[0, 1, 2]
8	Mhrt	Maximum Heart rate	Numerical	[71–200]
9	Exian	Exercise induce angina	Binary	[0–1]
10	Opk	Oldpeak	Numerical	[0–6.2]
11	Slope	Slope of peak exercise ST segment	Nominal	[1, 2, 3]
12	Vessel	Number of major vessel	Nominal	[0, 1, 2, 3]
13	Thal	Defect type	Nominal	[3, 6, 7]
14	Class label	Heart disease		[0, 1, 2, 3, 4] for CHD5[a] [0, 1] for SHD and CHD2[b]

[a] CHD with five class labels.
[b] CHD with two class labels.

label associated with it. There are 500 instances in class 0 (indicates absence of diabetes) and 268 instances in class 1 (indicates the presence of diabetes). Description of PID dataset is presented in Table 2.

4.2. Preprocessing subsystem

Data preprocessing aids in obtaining a smooth dataset and improves the quality of the patterns mined [7]. Data preprocessing includes removal of noisy data, handling of missing values, normalization, and data reduction. In this study, handling of missing values and eliminating instances with outliers from the clinical dataset are done by the data preprocessing subsystem. The CHD dataset has missing values and the PID dataset has the value zero associated with certain features where the value zero is considered to be an outlier. The SHD dataset has neither missing values nor outliers.

Table 2
Description of PID dataset.

No	Feature	Description	Domain	Zero entry
1	Preg	Number of times pregnant	[0–5]	111
2	Glu	Plasma glucose concentration a 2 h in an oral glucose tolerance test	[0–199]	5
3	Bp	Diastolic blood pressure (mm Hg)	[0–122]	35
4	Skin	Triceps skin fold thickness (mm)	[0–99]	227
5	Insulin	2-h serum insulin (mu U/ml)	[0–846]	374
6	BMI	Body mass index (kg/m^2)	[0–67]	11
7	DPF	Diabetes pedigree function	[0.078–2.42]	–
8	Age	Age (years)	[21–81]	–
9	Class	Class label	[0, 1]	

4.2.1. Preprocessing in CHD

In CHD dataset, the features vessel and Thal has 4 and 2 missing values. Since the missing values are less in number they are handled by imputing with the most frequent values of the first five nearest neighbors of the concept class. The nearest neighbor is computed by using the Euclidean distance measure presented in Eq. (3) below [7].

$$d(x_i, x_j) = \left[\sum_{s=1}^{p} (x_{is} - x_{js})^2\right]^{1/2} \quad j = 1, 2, \ldots, g \tag{5}$$

where, x_i is the instance with missing value; x_j is the instance without missing value belonging to the same class label of x_i; s is the corresponding feature values of x_i and x_j; p is the total number of features in the clinical dataset; g is the number of instances without missing values.

4.2.2. Preprocessing in PID

In PID dataset the features number of times pregnant, Plasma glucose concentration, diastolic blood pressure, triceps skin fold thickness, 2-h serum insulin and body mass index has 111, 5, 35, 227, 374 and 11 instances that have the value 0 associated with it. A total of 432 instances have one or more features mentioned above with value 0 associated with it. All the 432 instances have been rejected and the PID dataset has been reduced from 768 instances to 336 instances since the value 0 for the above mentioned features are considered as outliers and eliminated as in [17]. Among the 336 instances, 225 instances belong to class 0 (indicates absence of diabetes) and 111 instances in class 1 (indicates the presence of diabetes).

4.3. Fuzzification subsystem

Trapezoidal membership function is used to transform the features of the selected clinical datasets to fuzzy set with membership value. This membership function is also used by other researchers [26,27] for fuzzification of clinical dataset. The trapezoidal membership function is presented in Eq. (6).

$$f(X, a, b, c, d) = \begin{cases} 0, X < a, X > d \\ \frac{(X-a)}{(b-a)}, a \le X \le b \\ 1, b \le X \le c \\ \frac{(d-X)}{(d-c)}, c \le X \le d \end{cases} \tag{6}$$

The parameters $\{a, b, c, d\}$ are used to determine the membership values of the feature value X. These parameters in each feature of the clinical datasets are assigned by referring clinician's suggestion from [28–33].

4.3.1. Fuzzification of heart disease datasets

CHD and SHD datasets have thirteen features, among which five features are numerical, three features are binary and five features are nominal data type. Among the five features with numerical data type, four features namely, patient age in years (Age), resting blood pressure (Bp), maximum heart rate (Mhrt), and oldpeak (Opk) is represented by three fuzzy sets (fuzzy variables). The feature serum cholesterol (SCh) having numeric data type is represented by four fuzzy variables. Table 3 illustrates the fuzzy set and points corresponding to the features with numerical data type for CHD and SHD datasets. Fig. 2 represents fuzzification of the numerical data type features for CHD and SHD datasets.

The feature gender (sex) is transformed to male and female. The feature Fasting blood sugar (Fbs) stores the level of fasting blood glucose. This test is performed after the patient has been fasting for eight hours continuously. The feature Fbs is transformed to excess-glucose and normal-glucose. Excess-glucose indicates that the blood glucose level is >120 mg/dl and normal-glucose indicates that the blood glucose level is ≤ 120 mg/dl. The feature exercise induced angina (Exian) indicates whether chest pain occurred or not during exercise. This happens because sufficient blood through the arteries is not supplied to the walls of the heart. Exian is transformed to exian-positive and exian-negative. Exian-positive indicates that exercise has induced angina and Exian-negative indicates that exercise has not induced angina. The features chest pain type (Chp), resting electrocardiographic result (ECG), slope of peak exercise ST segment (Slope), number of major vessel colored by fluoroscopy (Vessel) and Thal with nominal data type are transformed to three or more linguistic variables based on the value associated with each feature. The feature Chp is transformed to typical angina, atypical angina, Non-anginal and Asymptotic. The feature ECG is transformed to value-0, value-1 and value-2 to indicate ECG result has normal wave, abnormal wave and hypertrophy of left ventricle respectively. The feature Slope is transformed to up-sloping, flat and down-sloping. The feature Vessel is transformed to colored-0, colored-1, colored-2 and colored-3. The feature Thal is transformed to normal, fixed and reversible.

Totally CHD and SHD datasets are transformed to 39 variables. Table 4 illustrates the linguistic variables of binary and nominal features for CHD and SHD datasets.

Table 3
Fuzzy set corresponding to features with numerical data type for CHD and SHD datasets.

Feature	Fuzzy Set	Points			
		a	*b*	*c*	*d*
Age	Young	20	20	30	35
	Middle-Aged	30	40	50	60
	Old	50	60	80	80
Bp	Normal	80	90	120	130
	Hypertension	120	130	160	170
	Hypertensive	160	170	200	200
SCh	Low	120	120	160	180
	Desirable	160	180	200	210
	Border-line	200	220	240	250
	Risk	240	260	600	600
Mhrt	Below	50	50	100	110
	Normal	100	110	180	190
	Above	180	190	220	220
Opk	Low	0	0	1.5	2
	High	1	2	3.5	4.5
	Terrible	3	4	7	7

4.3.2. Fuzzification of PID dataset

PID dataset has eight features with numerical data type. During fuzzification the seven features namely number of times pregnant (Preg), Plasma glucose concentration a 2-h in an oral glucose tolerance test (Glu), diastolic blood pressure (Bp), triceps skin fold thickness (skin), 2-h serum insulin (Insulin), diabetes pedigree function (DPF), and age is represented by three fuzzy variables. The remaining one feature, Body mass index (BMI) is represented by four fuzzy variables. Totally PID dataset features are transformed to 25 linguistic variables. Table 5 shows the fuzzy set to the corresponding features of PID dataset and Fig. 3 represents the fuzzification of PID dataset.

4.4. Classification subsystem

Classification subsystem consists of two processes namely, classifier construction and classifier testing. In this research work, classification is done using feed forward neural network with a single hidden layer using extreme Learning Machine (ELM) for determining the weights between the hidden layer neurons and output layer neurons [8].

Let $p, q$and$r$ represent the number of neurons in input layer, hidden layer and output layer respectively. Let W^{ih} represents the weight vector between the input layer neurons and the hidden layer neurons, W^{ho} represents the weight vector between the hidden layer neurons and output layer neurons and T represents the expected value of the fuzzified clinical dataset (X_i). Sigmoid activation function is used for hidden layer neurons. Table 6 shows the parameter values of the FELM.

For each value of hidden layer neuron (q), the training of SLFNN is described as follows.

Step 1: Use the fuzzified features of the clinical dataset as input to the FELM as shown in Eq. (7).

$$I_i = X_i i = 1, 2, \ldots, p \tag{7}$$

where, p refers to the total number of fuzzified features of the clinical dataset(X).

Step 2: Initialize the weights (W^{ih}) between the input layer neurons and hidden layer neurons randomly ranging from 0 to 1;

Where i denotes the input layer neuron and h denotes the hidden layer neuron. In this research work the value of i ranges from 1 to the number of fuzzified features in the clinical data set and the value of h ranges from 1 to q; where q represents the number of neurons in the hidden layer.

Step 3: Compute the input of hidden layer neurons (H_j^i) using Eq. (8);

$$H_j^i = \sum_{i=1}^{p} \left(I_i^o W_{ij}^{ih} \right) \quad j = 1, 2, \ldots, q \tag{8}$$

where, I_i^o is the output of the input layer neurons, W_{ij}^{ih} is the weight between the input and the hidden layer neuron.

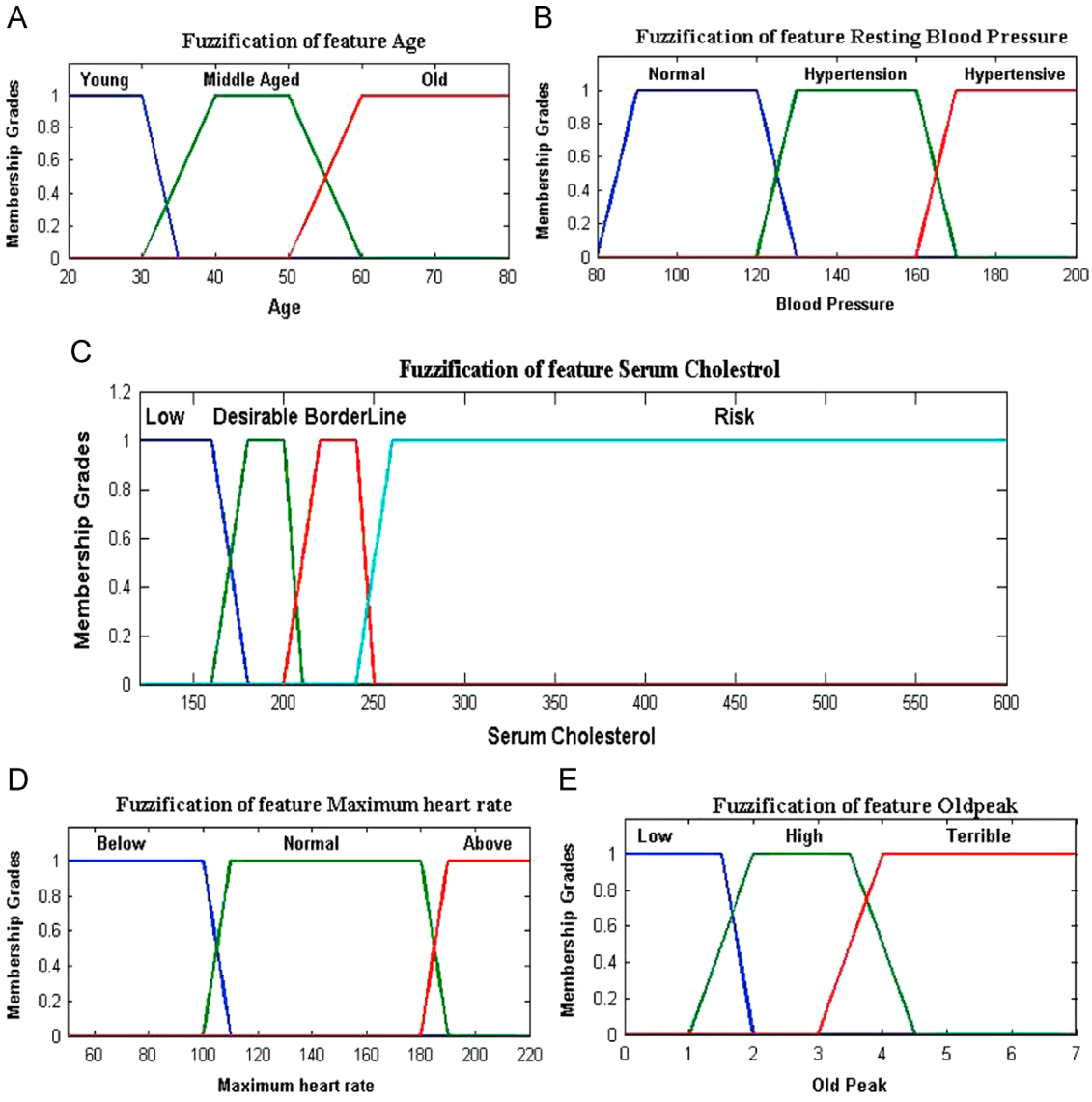

Fig. 2. Fuzzification of numerical features of heart disease datasets.

Table 4
Linguistic variable for corresponding binary and nominal data type features of CHD and SHD datasets.

Data Type	Features	Linguistic Variables
Binary	Sex	Male Female
	Fbs	Normal glucose Excess glucose
	Exian	Exian-positive Exian-negative
Nominal	Chp	Typical angina Atypical angina Non-anginal
	ECG	Value-0 Value-1 Value-2
	Slope	Colored-0 Colored-1 Colored-2 Colored-3
	Thal	Normal Fixed Reversible

Step 4: Compute the output of a hidden layer neurons (H^o) using Eq. (9);

$$H_j^o = \frac{1}{1+e^{-H_j^i}} \quad j = 1, 2, \ldots, q \tag{9}$$

Step 5: Determine the weights (W^{h0}) between hidden layer neuron and output layer neuron using ELM method as shown in Eq. (4)
Step 6: Obtain the value of output layer neuron (O_k) using Eq. (1).

5. Experimental results

Experiments were conducted on the selected clinical datasets using MATLAB tool version 7.10, release R2010a. The performance metrics namely, accuracy, sensitivity, specificity, True Positive Rate (TPR), False Positive Rate (FPR) and precision were used for evaluating the Fuzzy Extreme Learning Machine (FELM). The metrics are computed by considering True Positives (TP), False Negatives (FN), True Negatives (TN) and False Positives (FP). True positives (TP) refer to those instances that are truly identified as a diseased

Table 5
Fuzzification of PID dataset.

Feature	Fuzzy Set	Points			
		a	b	c	d
Preg	Low-Preg	1	1	2	4
	Normal-Preg	3	4	5	7
	High-Preg	5	7	20	20
Glu	Low-Glu	45	45	70	90
	Normal-Glu	70	90	140	150
	High-Glu	140	150	200	200
Bp	Hypotension	20	20	55	65
	Normal	60	70	85	90
	Hypertension	85	100	120	120
Skin	Thin-Skin	5	5	8	12
	Medium-Skin	10	15	20	30
	Thick-Skin	20	30	55	55
Insulin	Low-Insulin	15	15	25	50
	Normal-Insulin	25	60	150	180
	High-Insulin	160	200	850	850
BMI	Underweight	15	15	18.5	20
	Ideal	18.5	20	25	28
	Overweight	25	27	30	32
	Obese	30	35	60	60
DPF	Low-DPF	0.085	0.085	0.35	0.5
	Medium-DPF	0.35	0.5	0.85	1.0
	High-DPF	0.85	1.0	2.4	2.4
Age	Young	20	20	25	30
	Medium-aged	25	30	45	50
	Old	45	50	82	82

patient by the classifier. If the patient is not correctly classified, it becomes False Negatives (FN). Healthy instances correctly identified by the classifier becomes True Negatives (TN), if not it becomes False Positives (FP). The metrics used to measure the performance of the classifier using Eqs. (10)–(16) [7].

$$Sensitivity = \frac{TP}{TP+FN} \tag{10}$$

$$Specificity = \frac{TN}{TN+FP} \tag{11}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{12}$$

$$TPR = \frac{TP}{TP+FN} \tag{13}$$

$$FPR = \frac{FP}{TN+FP} \tag{14}$$

$$Precision = \frac{TP}{TP+FP} \tag{15}$$

$$F\_Measure = \frac{2TP}{2*TP+FP+FN} \tag{16}$$

The experiment was carried using the parameters as described in Table 6. The number of neurons in the hidden layer is initially set to 10. The network is trained using the training set and tested by a testing set. Additional neurons are incrementally added one at a time. A termination condition is reached when the number of neurons in a hidden layer is 200. The experiment is performed using training–testing rate of (80–20), (70–30), (60–40), and (50–50). Table 7 shows the highest result of each training–testing division with the number of hidden layer neuron (q).

As shown in Table 7, the highest accuracy is obtained for selected clinical dataset using 80–20 training–testing rate with accuracy of 73.77%, 93.55%, 94.55% and 92.54% for CHD with five class labels (CHD5), CHD with two class labels (CHD2), SHD and PID respectively.

The training and testing accuracy of the FELM using 80–20 training–testing rate with the hidden layer 10–200 is illustrated in Fig. 4–7.

As shown in the Figs. 4–7, highest accuracy is obtained with 33, 47, 25 and 39 hidden layer neurons for CHD5, CHD2, SHD and PID clinical dataset respectively. If number of hidden layer neuron become more than the corresponding values, the accuracy of testing dataset decreased gradually while the training set accuracy increased to 100%. From this result, it is concluded that as the

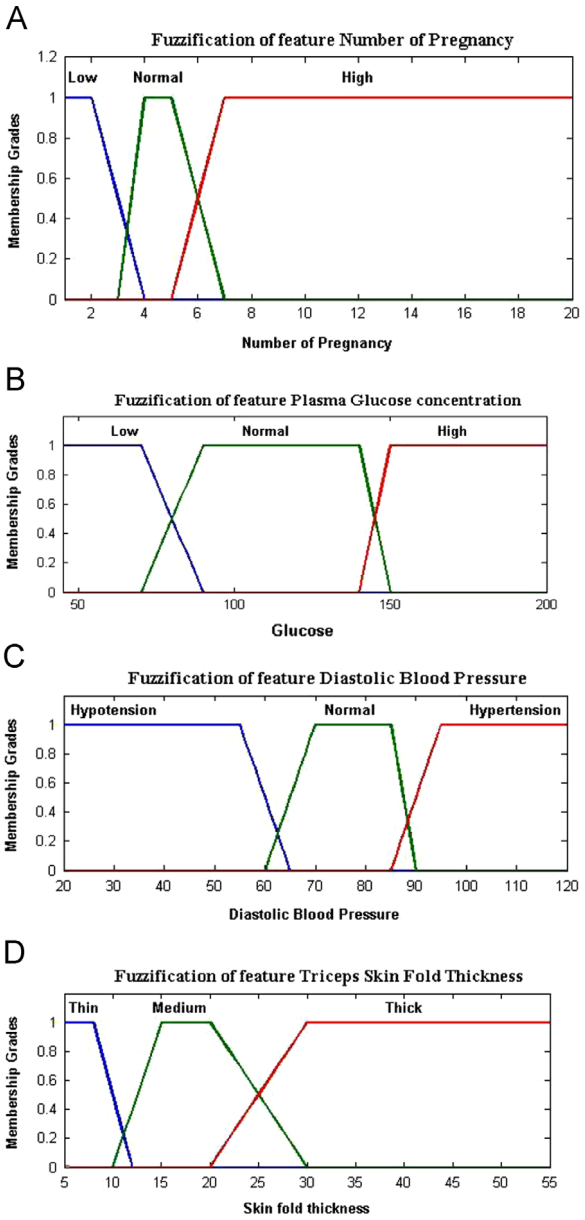

Fig. 3. Fuzzification of Pima Indian diabetes dataset.

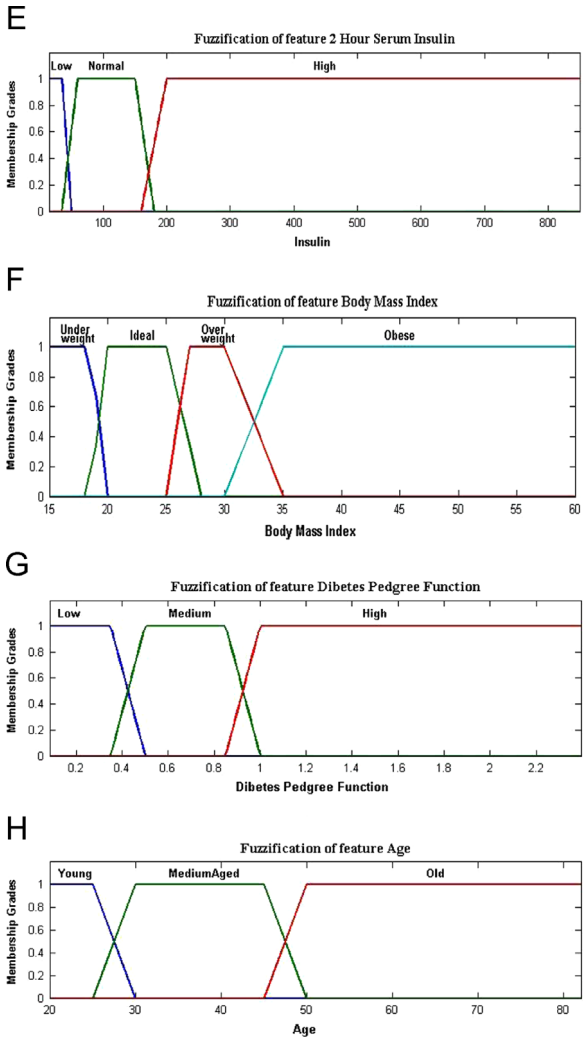

Fig. 3. (*continued*)

Table 6
Parameters for FELM.

	CHD5	CHD2	SHD	PID
Input layer Neurons (p)	39	39	39	25
Output Layer Neurons (r)	5	2	2	2
Hidden layer Neurons (q)	10, 11, 12,…, 200			
Learning algorithm	ELM			
Activation Function	Hidden Layer: Sigmoid			
Dataset division	Training–testing set % (80–20), (70–30), (60–40), (50–50)			

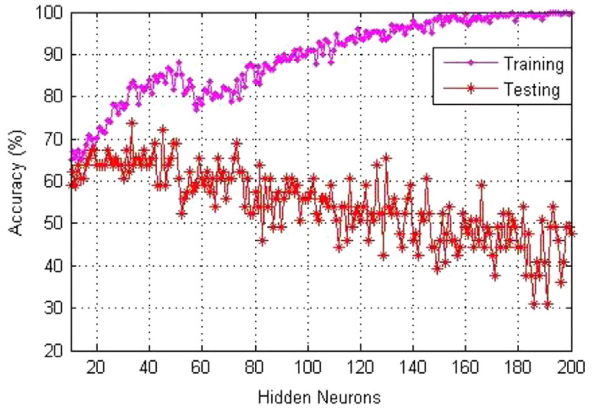

Fig. 4. Training and testing accuracy for CHD with five class labels dataset.

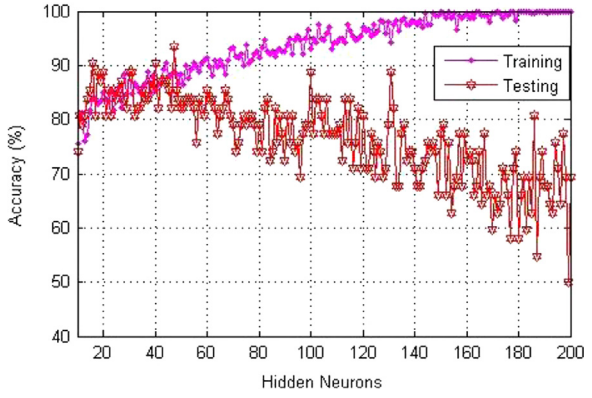

Fig. 5. Training and testing accuracy for CHD with two class labels dataset.

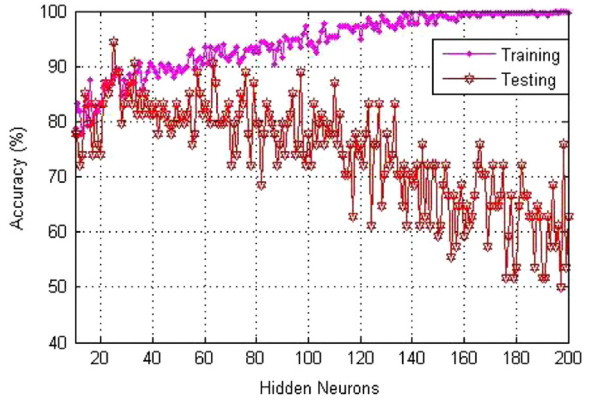

Fig. 6. Training and testing accuracy for SHD dataset.

Table 7
Comparison of fuzzy extreme learning machine accuracy.

Dataset	(Training–Testing)							
	(80–20)		(70–30)		(60–40)		(50–50)	
	Accuracy (%)	q	Accuracy (%)	q	Accuracy (%)	q	Accuracy (%)	q
CHD5[a]	73.77	33	68.48	37	67.77	29	64.71	39
CHD2[b]	93.55	47	91.21	32	88.52	31	86.16	28
SHD	94.44	25	92.59	24	91.67	50	91.11	29
PID	92.54	39	89.11	29	85.82	32	82.25	23

[a] CHD with five class labels.
[b] CHD with two class labels.

number of hidden layer neurons increased, the classifier leads to overfitting problem

Table 8 Summarizes the highest testing accuracy with the corresponding training phase accuracy and hidden layer neurons.

Table 9 presented contingency table for CHD with five class labels and Table 10 presented contingency table for CHD with two class labels, SHD and PID testing dataset.

The Receiver Operating Character (ROC) with the Area under ROC Curve (AUC) of the selected clinical dataset is illustrated in Fig. 8.

As shown in Fig. 8, the broken line drawn from the bottom left to upper right corner is the imaginary line to show that the AUC is greater than 0.5; The bold line shows the border region of the ROC curve. AUC of the three datasets; CHD2, SHD and PID becomes 0.935, 0.942 and 0.909 respectively. As the CHD5 has five class labels, it is not possible to draw a single ROC curve clinical dataset with binary class labels.

The overall performance of the clinical dataset is using the selected number of hidden layer is shown Table 11.

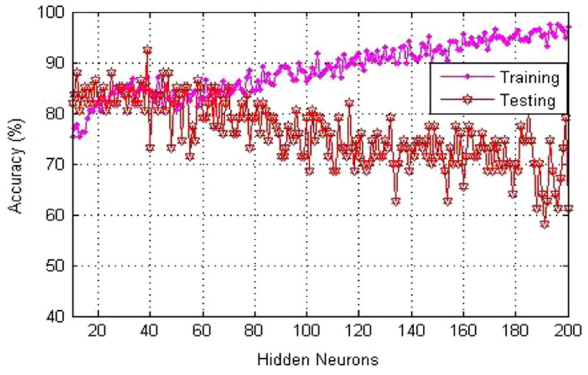

Fig. 7. Training and testing accuracy for PID dataset.

Table 8
Number of hidden layer neurons for the highest accuracy.

Dataset	*q*	Training accuracy (%)	Testing accuracy (%)
CHD5[a]	33	83.42	73.77
CHD2[b]	47	88.80	93.55
SHD	25	87.96	94.44
PID	39	86.99	92.54

[a] CHD with five class labels.
[b] CHD with two class labels.

Table 9
Contingency table for CHD5 testing set.

		Expected				
		Absence	Low	Medium	High	Serious
Predicted	Absence	33	3	0	0	1
	Low	1	4	2	0	0
	Medium	0	2	3	3	0
	High	0	2	2	4	0
	Serious	0	0	0	0	1

5.1. Performance comparison

The performance of FELM is compared with ELM, BPNN and Fuzzy Backpropagation Neural Network (FBNN) in terms of training time and accuracy. The training–testing rate for all learning method is designed to be 80–20. The ELM classifier developed and tested with a varying number of hidden layer neurons ranging from 10 to 200. The BPNN and FBPNN have designed with a single layer neural network with 25 and 50 neurons respectively. The parameter used for both BPNN and FBPNN are hyperbolic sigmoid activation function for hidden layer neurons and linear activation function for output layer neurons and with maximum number of iteration is 1000. Table 12 shows comparison of FELM with ELM, BPNN and FBPNN

As illustrated in Table 12, FELM achieves highest accuracy as compared to BPNN, ELM and FBNN. Because of no issues like minima and improper learning rate, training time using ELM and FELM takes less than 0.1 s. Even though both ELM and FELM has less training time, FELM has performs well in terms of accuracy.

Comparison of the proposed FELM method, with the existing method implemented by the other researchers using datasets from UCI machine learning repository is shown in Table 13.

As illustrated in Table 13, FELM classification system has obtained highest accuracy compared to other authors work for classifying, CHD2 and SHD. FELM achieves highest result for PID dataset except [23] work. The proposed classifier on CHD5 dataset has obtained competent accuracy to the work done by [20,36].

6. Conclusion and future work

Classification is a vital tool for diagnosis of disease. In this study a fuzzy extreme learning machine (FELM) is proposed for diagnosis of the two major silent killer diseases; heart disease and diabetes with Cleveland heart disease (CHD), Statlog heart disease (SHD), and Pima Indian diabetes (PID) datasets. CHD dataset has been tested in two ways; severity of heart disease (CHD with five class labels) and whether heart disease has occurred or not (CHD with two class labels). After handling missing value and removing outliers, fuzzification has been applied for mapping the features to fuzzy set with a degree of membership ranging from 0 to 1. The fuzzy input dataset has been fed to the ELM with the training–testing rate of 80–20%. Maximum accuracy for testing dataset is obtained with 33, 47, 25, and 39 hidden layer neurons for CHD with five class labels, CHD with two class labels, SHD and PID dataset respectively with the accuracy of 73.77%, 93.55%, 94.44%, and 92.54%. The performance with respect to sensitivity,

Table 10
Contingency table for CHD2, SHD and PID testing set.

		Expected					
		CHD2		SHD		PID	
		Presence	Absence	Presence	Absence	Presence	Absence
Predicted	Presence	TP (29)	FN (2)	TP (22)	FN (2)	TP (19)	FN (3)
	Absence	FP (2)	TN (29)	FP (1)	TN (29)	FP (2)	TN (43)

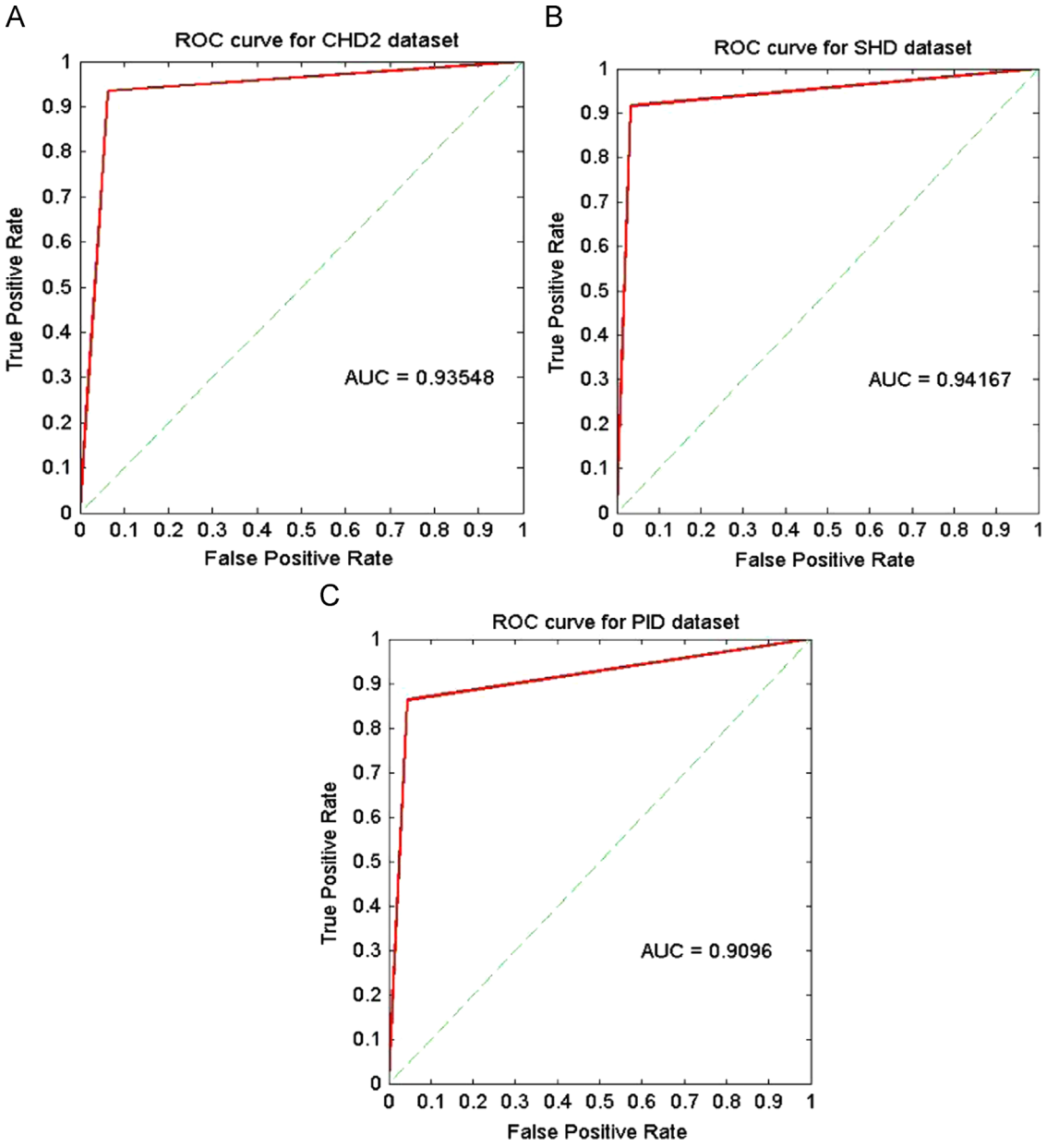

Fig. 8. ROC curve for (A) CHD2, (B) SHD and (C) PID dataset.

Table 11
Overall performance of dataset in the selected hidden layer neurons.

Metrics	CHD2	Dataset SHD	PID
Accuracy (%)	93.55	94.44	92.54
Sensitivity (%)	93.54	95.65	90.47
Specificity (%)	93.54	93.55	93.48
Precision	0.935	0.912	0.864
*F*_measure	0.935	0.936	0.884
TPR	0.935	0.657	0.905
FPR	0.138	0.936	0.209
AUC	0.935	0.942	0.909

specificity, TPR, FPR, and *F*-measure were evaluated. The proposed method when compared to ELM, BPNN and FBNN performs higher in terms of training time and accuracy. Performance of FELM other researchers' work exhibits better performance results for CHD

Table 12
Performance Comparison of FELM with ELM and BPNN.

Learning method	Performance measure	Clinical dataset			
		CHD5	CHD2	SHD	PID
BPNN	Accuracy (%)	63.04	85.2	85.2	79.1
	Time (s)	1.615	1.411	1.49	1.37
FBPNN	Accuracy (%)	64.69	87.46	85.2	80.6
	Time (s)	1.848	1.548	1.69	1.71
ELM	Accuracy (%)	65.57	88.52	85.19	82.09
	Time (s)	0.013	0.013	0.013	0.013
FELM	Accuracy (%)	**73.77**	**93.55**	**94.44**	**92.54**
	Time (s)	**0.013**	**0.013**	**0.013**	**0.013**

Table 13
Performance comparison of FELM with other authors work.

Author(s)	Proposed method	Accuracy (%)			
		PID	CHD5	CHD2	SHD
Aslam et al. [14]	GP-KNN	80.50	–	–	–
Patil et al. [15]	K-means with C4.5	92.38	–	–	–
Alneamy et al. [16]	TLBO-FWNN	90.29	–	–	–
Varma et al. [17]	Gini-Fuzzy	75.80	–	–	–
Seera et al. [19]	FMNN-CART-RF	78.39	–	–	–
Kalpana et al. [21]	FIS	90.38	–	–	–
Christopher et al. [22]	PSO-C4.5	83.99	–	77.89	–
Lee et al. [26]	Fuzzy Expert system	91.20	–	–	–
Dennis et al. [20]	AGFS	89.80	76.67	–	–
Subbulakshmi and Deepa [23]	SRLPSO-ELM	93.09	–	91.33	89.96
Bhatia [38]	SVM-GA	–	72.55	90.57	–
Mattila et al. [34]	Generic Statistical approach	78.04	–	85.01	–
Anooj [35]	Weighted fuzzy rule	–	–	62.35	–
Setiawan et al. [36]	Fuzzy based roughest	–	–	83.00	–
Kahramanli et al. [37]	Hybrid of AIS and ANN	–	–	87.40	–
Nahato et al. [18]	RS-BPNN	–	–	–	90.40
Proposed Method	**FELM**	**92.54**	**73.77**	**93.55**	**94.44**

with two class labels, SHD, PID datasets. The proposed work's performance is competent in respect of CHD5 when compared to other works. Hybrid FELM with bio-inspired optimization techniques will be considered as a future work.

Conflict of interest

We wish to confirm that there are no known conflicts of interest associated with this publication entitled "**Hybrid Approach using Fuzzy Sets and Extreme Learning Machine for Classifying Clinical Datasets**", and there has been no significant financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from the editorial office of the Journal "**Informatics in Medicine Unlocked"**.

References

[1] ⟨www.medindia.net/health_statistics/health_facts/silent-killer-diseases-facts.htm⟩ [accessed 15.04.15].
[2] ⟨www.who.int/mediacentre/factsheets/fs317/en/⟩ [accessed June, 2015].
[3] ⟨www.who.int/mediacentre/factsheets/fs312/en/⟩ (accessed June, 2015].
[4] Berner ES. Clinical decision support systems, theory and practice: health Informatics serious. 2nd ed. . USA: Springer Science & Business Media; 2007.
[5] Thuraisingham B. A primer for understanding and applying data mining. IT Prof 2000;2(1):28–31.
[6] Prather JC, Lobach DF, Goodwin LK, Hales JW, Hage ML, Hammond WE. Medical data mining: knowledge discovery in a clinical data warehouse. In: Proceedings of the AMIA annual fall symposium. American Medical Informatics Association; 1997. p. 101–5.
[7] Han J, Kamber M, Pei J. Data mining concepts and techniques. 3rd ed.. San Francisco: Morgan Kaufmann; 2012.
[8] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1):489–501.
[9] Huang GB, Wang DH, Lan Y. Extreme learning machines: a survey. Int J Mach Learn Cybern 2011;2(2):107–22.
[10] Rajasekaran S, Pai GV. Neural networks, fuzzy logic and genetic algorithms. New Delhi: PHI Learning Private Limited; 2011.
[11] Zadeh LA, Fuzzy-Set-Theoretic A. Interpretation of linguistic hedges. J Cybern 1972;2(3):4–34.
[12] Jang JSR, Sun CT, Mizutani E. Neuro-fuzzy and soft computing a computational approach to learning and machine intelligence. New Delhi: PHI Learning; 2011.
[13] Zhang WB, Ji HB. Fuzzy extreme learning machine for classification. Electron Lett 2013;49(7):448–50.
[14] Aslam MW, Zhu Z, Nandi AK. Feature generation using genetic programming with comparative partner selection for diabetes classification. Expert Syst Appl 2013;40(13):5402–12.
[15] Patil BM, Joshi RC, Toshniwal D. Hybrid prediction model for type-2 diabetic patients. Expert Syst Appl 2010;37(12):8102–8.
[16] Alneamy JSM, Rahma AHA. Heart disease diagnosis utilizing hybrid fuzzy wavelet neural network and teaching learning based optimization algorithm. Adv Artif Neural Syst 2014;6 (Article ID 796323).
[17] Varma KV, Rao AA, Lakshmi TSM, Rao PN. A computational intelligence approach for a better diagnosis of diabetic patients. Comput Electr Eng 2014;40(5):1758–65.
[18] Nahato KB, Harichandran KN, Arputharaj K. Knowledge mining from clinical datasets using rough sets and backpropagation neural network. Comput Math Methods Med 2015:460189.
[19] Seera M, Lim CP. A hybrid intelligent system for medical data classification. Expert Syst Appl 2014;41(5):2239–49.
[20] Dennis B, Muthukrishnan S. AGFS: Adaptive Genetic Fuzzy System For Medical Data Classification. Appl Soft Comput 2014;25:242–52.
[21] Kalpana M, Kumar AS. Design and Implementation of Fuzzy Expert System using Fuzzy Assessment Methodology. Int J 2012;1(1):39–45.
[22] Christopher JJ, Nehemiah HK, Kannan A. A swarm optimization approach for clinical knowledge mining. Comput Methods Progr Biomed 2015;121(3):137–48.
[23] Subbulakshmi CV, Deepa SN. Medical dataset classification: a machine learning paradigm integrating particle swarm optimization with extreme learning machine classifier. Sci World J 2015:418060.
[24] Kaya Y, Uyar M. A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease. Appl Soft Comput 2013;13(8):3429–38.
[25] Lichman M. UCI machine learning repository. Irvine, CA: University of California, School of Information and Computer Science; 2013. ⟨⟨http://archive.ics.uci.edu/ml⟩⟩.
[26] Vijaya K, Khanna NH, Kannan A, Bhuvaneswari NG. Fuzzy neuro genetic approach for predicting the risk of cardiovascular diseases. Int J Data Min Manag 2010;2(4):388–402.
[27] Lee CS, Wang MH. A fuzzy Expert System for Diabetes Decision Support Application. IEEE Trans Syst, Man,Cybern, Part B: Cybern 2011;41(1):139–53.
[28] ⟨http://www.webmd.com⟩ [accessed June 2015].
[29] ⟨http://www.nlm.nih.gov⟩ [accessed June 2015].
[30] ⟨http://www.healthline.com⟩ [accessed June 2015].
[31] ⟨http://www.fitday.com⟩ [accessed June 2015].
[32] ⟨http://www.cdc.gov⟩ [accessed June 2015].
[33] ⟨http://www.heart.org⟩ [accessed June 2015].
[34] Mattila J, Koikkalainen J, Virkki A, Van Gils M, Lotjonen J. Design and application of a generic clinical decision support system for multiscale data. Biomed Eng, IEEE Trans 2012;59(1):234–40.
[35] Anooj PK. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. J King Saud Univ—Comput Inf Sci 2012;24(1):27–40.
[36] Setiawan NA, Venkatachalam PA, Hani AFM. Diagnosis of coronary artery disease using artificial intelligence based decision support system. In: Proceedings of the international conference on man–machine systems. Batu-Ferringhi, Penang; 2009. p. 11–3.
[37] Kahramanli H, Allahverdi N. Design of a hybrid system for the diabetes and heart diseases. Expert Syst Appl 2008;35(1):82–9.
[38] Bhatia S, Prakash P, Pillai GN. SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features. In: Proceedings of the world congress on engineering and computer science, WCECS; 2008. p. 22–4.