International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 32087-32101

http://www.ripublication.com

An Expert System for Diabetes Diagnosis using Extreme Learning Machine and Simulated Annealing

Anto S[1] and Chandramathi S[2]

[1] Department of Computer Science and Engineering,
Sri Krishna College of Technology, Coimbatore
[2] Dean, Electrical Sciences, Sri Krishna College of Technology, Coimbatore.
Email: [1]georgeantocse@gmail.com, [2]chandrasrajan@gmail.com

Abstract

As diagnosing diabetes is a challenging and tedious task for physicians, it has attracted researchers to focus on designing medical decision support systems with higher accuracy. In this paper, a clinical decision support system based on Extreme Learning Machine (ELM) and Simulated Annealing (SA) is proposed for the diagnosis of diabetes using the Pima Indian Diabetes (PID) dataset of the UCI machine learning repository. Classification is performed using ELM while optimization of ELM parameter is carried out by Simulated Annealing heuristics. The performance of the proposed system is analyzed based on the parameters such as classification accuracy, sensitivity and specificity using 10-fold cross-validation and confusion matrix. The accuracy of the proposed system is found to be superior to other existing systems in the literature. The system is validated for different datasets like Hepatitis, Breast Cancer and Cardiac Arrhythmia.

Index Terms – Medical Expert System, Extreme Learning Machine (ELM), Simulated Annealing (SA), Diabetes Diagnosis, Classification.

1. INTRODUCTION

Nowadays, number of diabetes patients has increased. Though it is incurable, it can be controlled to a certain extent by proper diet. Moreover, for most diabetes affected patients, the symptoms remain invisible leading to identification at a latter, critical stage.

Diabetes diagnosis is done during periodic blood checkup or during blood donation. It can be a challenging task for the less experienced physicians to predict its existence. For diagnosing a disease it is important to know the details about the

disease [1]. Diabetes is the irregular blood glucose utilization, commonly known as blood sugar. Glucose is a vital source of energy to the muscles, tissues and more significantly, the brain.

An automated system which helps in diagnosis will benefit the medical industry. Researchers are working on designing a better system for diagnosis of various diseases by paving way to new ideas and combinations of algorithms on the medical datasets.

Machine learning is a branch of Artificial Intelligence (AI) which concerns with the study, design and building of systems that can learn from data. The data representation and evaluation of function on these data are part of all machine learning systems. Knowledge representation and generalization is the core of machine learning.

The training data used in supervised machine learning consists of a pair - an input object (typically a vector) and a desired output value (class label). Supervised learning is the task of deducing a function from labeled training data. A supervised learning algorithm is used to analyze the training data and produce an inferred function for prediction. Later, this function can be used for mapping unknown or new examples(test data). An optimal scenario will allow the new function of the algorithm to correctly determine the class labels for hidden instances. This creates a need for the learning algorithm to generalize from the training data for new instances in a reasonable way.

A decision support system is computer software that attempts to act like a human on a particular subject area [3]. During the past few years, medical expert systems for the diagnosis of different diseases have received more attention. Such systems have higher optimization potential and reduced financial cost [12]. Pattern recognition and data mining are the techniques used in these expert systems that allow retrieval of meaningful information from large scale medical data.

Classification is one of the data mining techniques used to make decisions on real world data. Direct use of data from the database may affect the performance of the system. Thus to use correct, complete and provide the data in desirable form, preprocessing is done. Data preprocessing involves missing value handling, feature selection and scaling of dataset. The processed data is classified to provide a better decision on diagnosis. Classification techniques like Support Vector Machine (SVM) and Extreme Learning Machine (ELM) require proper selection of certain parameters to achieve the best performance. Simulated Annealing (SA) is a process involved in metallurgy, where the metals are slowly cooled to make them less brittle. It follows the same concept as a greedy approach of selecting the optimum temperature. This concept can be used to select the best parameters for classifiers.

2. RELATED WORKS

Cheng-Lung Huang et al [4] have proposed a general adaptive optimization search methodology based on grid algorithm combined with SVM classifier. GA based approach with the Grid algorithm is used for validating several real world medical data such as diabetes, heart disease and breast cancer.

Yuan Ren, Guang Chen Bai [5] have proposed two SVM parameter optimization approaches namely Genetic Algorithm-Support Vector Machine (GA-SVM) and Particle Swarm Optimization-Support Vector Machine (PSO-SVM). The objective function is based on the leave-one-out cross-validation technique and the parameters of SVM are optimized using GA and PSO respectively.

Fayssal et al [6] have designed a diagnosis system for diabetes using fuzzy classifier and modified Artificial Bee Colony (ABC) algorithm. The accuracy of the system has not improved and has paved way for further research.

Santi Wulan Purnami et al [7] have implemented a smooth SVM for the diabetes diagnosis. Multiple Knot Spline-Smooth SVM (MKS-SSVM) is used for classification, where the original spline function is modified to provide better accuracy.

Diabetes diagnosis is also dealt using Multi-Layer Neural Network by Hasan Temurtas et al [2]. The accuracy of the system is comparatively low.

The key problem in a Neural Network (NN) is determining the number of hidden nodes which affects accuracy. To overcome this problem, the proposed system uses ELM on Single Hidden Layer Feed Forward Network, in which the hidden nodes are randomly selected and the optimal number of hidden nodes is determined by SA.

The proposed expert system uses ELM for classification and SA for optimizing the ELM parameter. The performance of an ELM is mainly decided by the number nodes to be present in the hidden layer. This parameter (number of nodes) is optimized by the SA. The feature selection is done using FS method. The overall flow and working of the proposed system is depicted in Fig.(1).

Feature selection using FS is explained in Section 3. Section 4 is dedicated for the classification of PID dataset using ELM. SA is applied for the optimization of the number of neurons and weights as discussed in Section 5. The performance of the system is analyzed in Section 6.

3. METHODOLOGY

Feature selection

Feature selection is the process of selecting the most significant features in the dataset. As the dimensionality influences the accuracy of the system, it is necessary to reduce the complexity of the system [9]. Fisher Score method is used to select the best features of all the four medical dataset dataset.

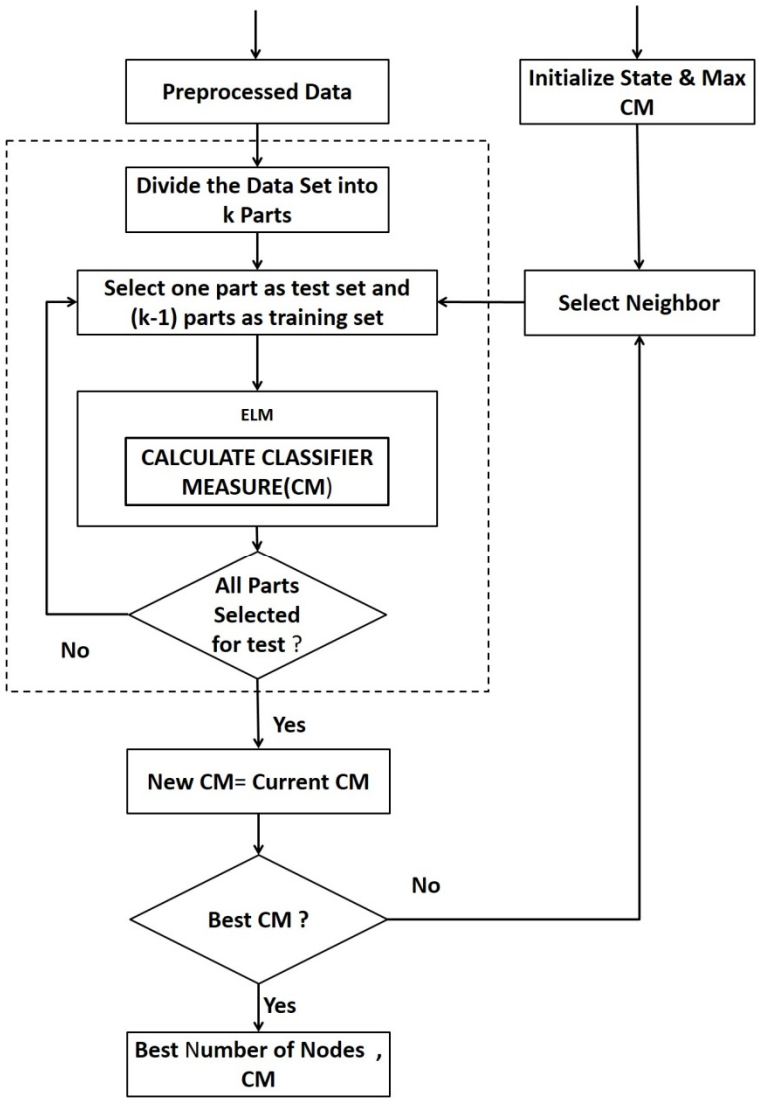

Figure 1. Flow Diagram of the proposed FS-ELM-SA

Fisher Score (FS)

The Fisher Score (FS) algorithm is widely used for many supervised learning systems for predicting the most discriminant features [10, 11]. The FS algorithm generates score for each attribute based on its importance in the whole dataset.

The essential features can be selected based on the scores generated. Datasets consist of (x_i, y_i) for 'N' instances. Let 'x_i' be the input vector with 'p' features and 'y_i' be the class label. FS selects the most relevant features 'm' from the given set of features. To find the most discriminating features, two basic steps are involved.

1. Feature score calculation for all features.
2. Selction of top 'm' features based on the score.

FS is computed using the following formula.

$$F(x_j) = \sum_{k=1}^{c} \frac{n_k(\mu_k^j - \mu_j)^2}{(\sigma_j)^2}, \; j=1, 2\ldots p \quad (1)$$

where,
'n_k' is the number of instances in each class 'k', 'μ^j' is the 'j^{th}' feature mean of whole dataset, 'σ^j' is the 'j^{th}' feature Standard Deviation (SD) of the whole dataset. It is given as

$$(\sigma_j)^2 = \sum_{k=1}^{c} n_k(\sigma_k^j)^2 \quad (2)$$

Scaling
To avoid the domination of attributes with greater numerical values over smaller values, linear transformation of numerical values within a range is computed in scaling. The values of the selected features from the dataset are normalized from 0 to 1.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} (\text{upperbound} - \text{lowerbound}) \quad (3)$$

where 'X' is the original data, 'X_{max}' is the maximum value of X, 'X_{min}' is the minimum value in X and 'X_{norm}' is the normalized value within the given upper and lower bound.

4. CLASSIFICATION

Feed forward neural networks (FFNN) are the most popular and most widely used models in classification problems. Fig. 2 illustrates a single hidden layer Feed Forward network with inputs 'x_i' and output 'O_j'. Each arrow in the figure symbolizes a parameter in the network. The network is divided into layers.

- The input layer consists of the inputs to the network.
- The hidden layer consists of neurons or hidden units placed in parallel. Each neuron performs a weighted summation of the inputs and then passes a nonlinear activation function, also called the neuron function.
- The output layer

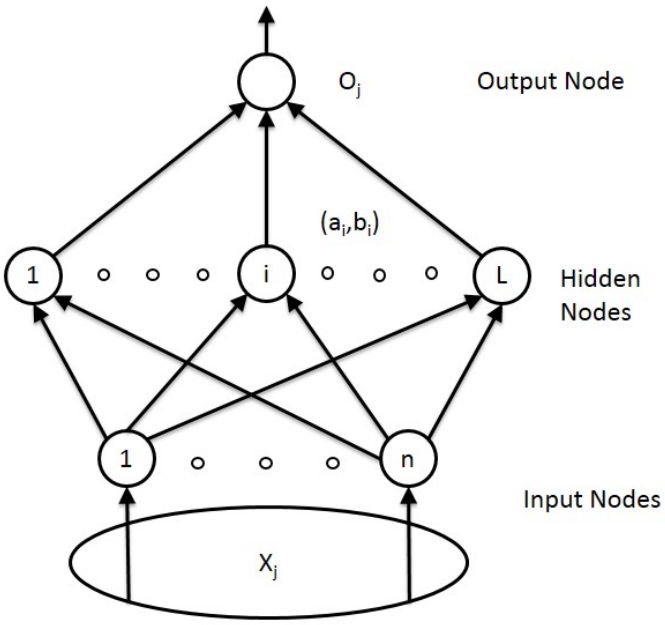

Figure 2. Single Hidden Layer Feed Forward Network

Extreme Learning Machine (ELM)

Huang et al., proposed a new learning algorithm called the Extreme Learning Machine (ELM). It is a Single-hidden Layered feed Forward neural networks (SLFNs).

According to Hanang et al., [18], ELM may randomly choose and fix all the hidden node parameters and then analytically determine the output weights. Once the weights of the SLFNs are randomly assigned, then SLFNs should be considered as a linear system. The output weights can be obtained analytically through a generalized inverse operation of the hidden layer output matrices.

Non-linear activation function used in neural network (sigmoid, hyperbolic function etc.,), Radial Basis Function (RBF) and complex activation function [19] are used as activation functions.

In the proposed system the SLFN has 'L' hidden nodes and it can be approximated by the given 'N' pairs of input /output values, namely, $(x_i, t_j) \in R^n \times R^m$ with zero error.

$$\sum_{i=1}^{p} \beta_i iG(a_i, X_j, b_i) = t_j, \text{ for } j= 1, 2, \ldots. L \quad (4)$$

where (a_i, b_i) is the parameter associated with 'i^{th}' hidden node and 'β_i' is the output weight linking the 'i^{th}' hidden node to the output node.

In this paper, a non-linear activation function called RBF (Equation 5) is used.

$$G(a_i, X_j, b_i) = g(b_i \| X_j - a_i \|) \quad (5)$$

Hence, Equation (5) can be rewritten as,

$$H\beta = T \quad (6)$$

where

$$H = \begin{bmatrix} G(a_1, X_1, b_1) G(a_2, X_1, b_2) \dots \dots \dots G(a_p, X_1, b_p) \\ G(a_1, X_2, b_1) G(a_2, X_2, b_2) \dots \dots \dots G(a_p, X_2, b_p) \\ \vdots \\ G(a_1, X_N, b_1) G(a_2, X_N, b_2) \dots \dots \dots G(a_p, X_N, b_p) \end{bmatrix} \quad (7)$$

$$\beta = \left[\beta_1^T, \beta_2^T, \beta_3^T \dots . \beta_p^T\right]^T \quad (8)$$

$$T = \left[t_1^T, t_2^T, t_3^T \dots t_N^T\right]^T \quad (9)$$

$$\tilde{\beta} = H^{\#}T \quad (10)$$

'$\tilde{\beta}$' is used as the estimated value of 'β', where '$H^{\#}$' is the Moore-Penrose generalized inverse of the hidden layer output matrix 'H' [20]. The formal ELM algorithm proposed by Huang et al., is given below [21]. Consider the ELM learning algorithm as 'L_θ' where 'θ' is the RBF hyper parameter.

ELM Algorithm

Given:

- A training set of input / output values,

$$(x_i, t_i) \in R^n R^m, \text{ for } i = 1, 2, \dots, N. \quad (11)$$

- An activation function,

$$G(a_i, X_j, b_i) = g(b_i \| X_j - a_i \|)$$

- The number of hidden nodes L.

Step 1: By using continuous sampling distribution, assign hidden nodes by randomly generating parameters (a_i, b_i), for i=1, 2...N
Step 2: Compute the hidden layer output matrix H.
Step 3: Compute the output weight '$\tilde{\beta}$', by using the relation $\tilde{\beta} = H^{\#}T$.

5. OPTIMIZATION

ELM based classification system performance is highly influenced by the number of hidden nodes (L). SA is employed in the arduous task of computing the optimal value for 'L'.

Simulated annealing is one of the most popular optimization techniques used for finding solutions for the optimization problems. It is a local heuristic search algorithm which uses the greedy method for finding optimal solutions [16]. It is a kind of Monte Carlo method used for examining the state and frozen state of the n-body system [17]. A maximum of 'k_{max}' iterations is performed until maximum accuracy is achieved. CallNeighbor( ) function finds the ensuing values of 'N'. CallRandom() function generates a random value from '0' and '1'. The basic SA process is shown in Fig. 3.

```
I←I0; A←A (I);
Ibest←I; Abest←A; //Initial Iteration, Accuracy.
k←0; kmax←Constant Value; //Initial "best" solution
MaxAccuracy←A constant Value //Evaluation count
while k < kmax and A < = MaxAccuracy
{
//while time left & not good enough
Inew←Neighbor (I) //Select number of Neurons
Anew←A (Inew) //Compute it's Accuracy
If exp (Anew −A) > Random () then
I←Inew; A←Anew // next Iteration
if Anew > Abest then //Is this a new best?
Ibest←Inew; Abest←Anew // Best Accuracy
k←k + 1 //One more evaluation done
}
return Ibest, Abest //Return the best Accuracy
```

Figure 3. Simulated Annealing Search Algorithm

In the hierarchical SA search, the Classification Measure (CM) is the basic parameter used for optimization technique. Classification accuracy of ELM is considered as the CM.

Initially, the SA parameters and ELM parameters are initialized. Then, the neighbours of the ELM parameters are selected and tried to tune these neighbours using the SA optimization search. The output of the first stage is given as the input to

the next stage. This decides whether the parameter is acceptable or not. If it is not acceptable, the parameter is tuned further.

The k-fold cross validation procedure is applied to the selected feature set to divide the set into training and testing sets. The cross validation technique returns a CM for 'k' classifiers built by the ELM algorithm. Each fold of the dataset is optimized using hyper parameter search strategy. The procedure for cross validation is as follows:

- **Division of datasets** - The medical datasets are divided into training and testing sets. The k-non overlapping equal sized subsets are formed on the given dataset 'D_i', where i=1, 2...k.
- **Classifier training -** 'k-1' folds are trained using classifier algorithm and the remaining one fold is tested on the trained classifier. Each classifier output generates accuracy for the predicted sets. The class performance is analyzed using performance parameters.
- **Classification Measure (CM)** - The CM is obtained by the Eq. (12).

$$CM=\frac{\text{Number of True Records Predicted}}{\text{Number of Total Records}} \quad (12)$$

CM is calculated for every sequential increase of the number of hidden nodes (L).

- **Optimization parameters -** The number of neurons in the hidden layer of the SLFN influences the performance of the system. To find the optimal number of neurons (L), SA optimization technique is used. The CM is calculated by increasing the number of neurons from 1 to 200. The value of parameter 'L' for which maximum CM is obtained, is chosen as the best value.

6. EXPERIMENTATION AND RESULTS

Diabetes Disease Dataset

The Pima Indian Diabetes (PID) dataset from University of California (UCI, Irvine) repository of machine learning is used [8]. The PID dataset consists of 768 instances with 268 (34.9%) cases in class '1' and 500 (65.1%) cases in class '0', where '1' means a positive test for diabetes and '0' is a negative test for diabetes.

This dataset consists of 8 features namely; number of times pregnant, Plasma Glucose Concentration (PGC), Diastolic Blood Pressure (DBP), Triceps Skin Folds Thickness (TSFT), Insulin, Body Mass Index (BMI), Diabetes Pedigree (DP) function and age. The most significant features among the 8 features are decided using Fisher Score (FS). Based on the scores of features, the best ones are selected. For diabetes, 4 features are selected. The top four features selected using both the methods are listed in Table 1 and the details of the number of selected features for all the four dataset is listed in Table 3.

Table 1. Selected Features

Features
PG Concentration
BMI
Number of times pregnant
Age

Performance metrics

The performance of the proposed system is analyzed using accuracy, sensitivity and specificity from confusion matrix. The confusion matrix holds both the actual and predicted instances classified by the classifier system. The confusion matrix of the system is shown in Table 2.

Table 2. Confusion Matrix

Predicted	**Actual**	
	Positive	**Negative**
Positive	TP (true positive)	FP (false positive)
Negative	FN (false negative)	TN (true negative)

- TN is the correct predictions of an instance as negative.
- FN is the incorrect predictions of an instance as positive.
- FP is the incorrect of predictions of an instance as negative.
- TP is the correct predictions of an instance as positive.

Classification accuracy is the rate of number of correct predictions, Sensitivity is the true positive rate, and specificity is the true negative rate. They are defined as shown in equations (13), (14) and (15).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (14)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (15)$$

Experimental Setup

The proposed model is implemented using MATLAB 8.0.1(R2013a). Datasets are divided into 10-folds using k-fold cross validation to make sure that all the instances in the datasets are trained and tested. The confusion matrix is formed from the predicted classifier output from which accuracy can be calculated (Eq.12). Accuracy

of the system varies with respect to the number of neurons in the hidden layer of ELM.

The parameters of SA algorithm (Fig. 3) are initialized and the CM is calculated in each iteration. The neighbors are chosen based on number of neurons. For every neighbor, the CM value is calculated.

In this work, diagnosis of diseases such as Diabetes, Hepatitis, Breast Cancer and Cardiac Arrhythmia using ELM-SA is done. All these four medical dataset are accessed from UCI machine learning repository [8, 13, 14, 15].

Table 3. The number of selected features by FS method

Disease	Number of features	
	Initial Set	**Selected Set**
Diabetes	8	4
Hepatitis	19	10
Breast Cancer	32	10
Cardiac Arrhythmia	279	15

The FS method is applied to all the four dataset to reduce the number of features. The data of the selected features for each medical datasets are normalized using min-max normalization.

After normalization, the data can be classified using ELM by dividing the datasets into k-folds and iterated for k times which will lead to k number of classification accuracy – one value for each fold. In each iteration, k-1 folds are taken as training sets and remaining one set is taken as the test set which ensures the testing of all the tuples on the dataset. Here 10-fold cross validation is done. ELM uses RBF as the kernel function to map data into the feature space. The critical parameter 'L' is optimized using SA to get the best combination of the optimal parameters and higher classification accuracy. Confusion matrix is obtained from the predicted classes. Based on the confusion matrix, parameters such as sensitivity, specificity and classification accuracy are calculated. The ELM classification of medical datasets using 10-fold cross validation is done and 96.45% classification accuracy (average of 10 folds) is obtained for diabetes.

Simulation and results

The classifier is iterated 200 times with incremental values of 'L' ranging from 1 to 200. It is inferred that the maximum performance (CM) of the system is achieved for the following 'L' values - 89, 107, 108, 131 and 182 (Fig. 4).

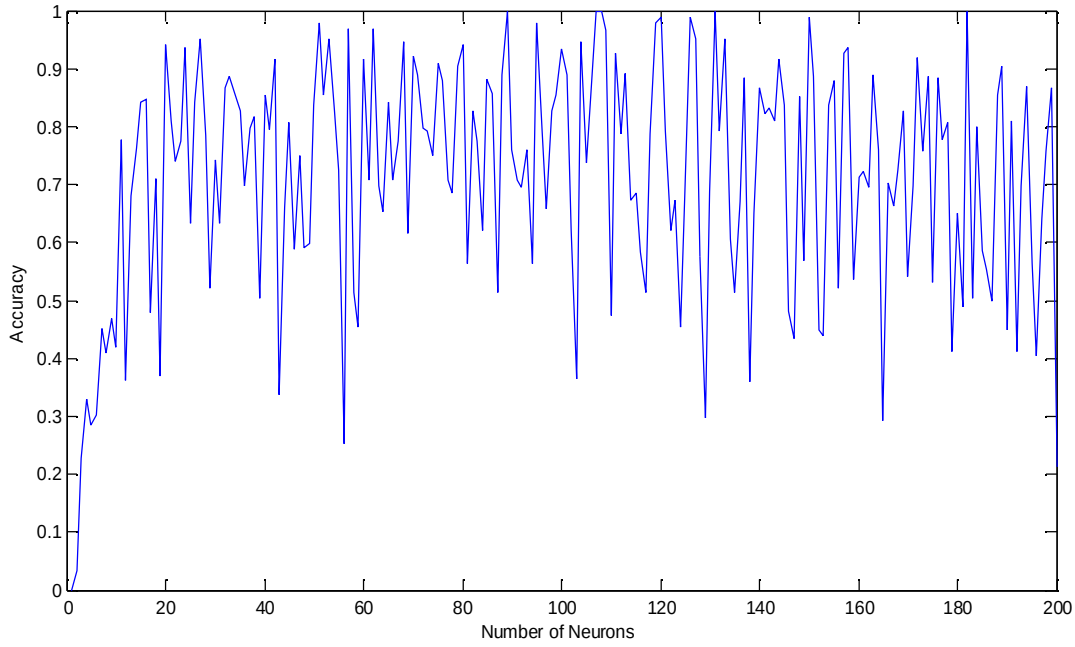

Figure 4. Variation of accuracy with iterations for FS_ELM_SA

Table 4. Accuracy, sensitivity and specificity of the proposed model

Medical Dataset	FS_ELM_SA			
	Best Accuracy	Average Accuracy	Specificity (Sp)	Sensitivity (Se)
Diabetes	1	0.9645	1	1

Table 5. Comparison of accuracy of the proposed method with existing methods for PID dataset

Diabetes	SVM	GA-SVM	GA – based approach	MKS-SSVM	MABC Fuzzy	Proposed FS-ELM-SA
Average Accuracy%	77.73	78.64	82.98	93.2	84.21	**96.45**

Table 6. Comparison of accuracy of the proposed method with existing methods for Hepatitis

Hepatitis	SVM	GA-SVM	Naive Bayes	KNN	Proposed FS_ELM_SA
Average Accuracy%	84.16	86.12	82.05	83.45	**88.08**

Table 7. Comparison of accuracy of the proposed method with other methods for Breast Cancer

Breast Cancer	SVM	GA-SVM	GA - based approach	GA	LLWNN-RLS	Proposed FS-ELM-SA
Average Accuracy %	75.87	76.57	94.23	90.78	97.2	**94.39**

Table 8. Comparison of accuracy of the proposed method with other methods for Cardiac Arrhythmia

Cardiac Arrhythmia	VFI5-GA	KNN-HITON	RF-CBFS	AIRS-FWP	Proposed GA-ELM-SA
Average Accuracy %	68	65.3	76.3	76.2	**76.09**

The sensitivity and specificity of the system is calculated using equations (13) and (14). Table 4 shows the highest accuracy on the best folds, the average accuracy over 10-folds, sensitivity and specificity of the proposed system for the PID dataset. The system achieves 96.45% accuracy which is better than the other existing methodologies.

The proposed system is validated using the medical datasets of Diabetes, Hepatitis, Breast Cancer and Cardiac Arrhythmia of the UCI machine learning repository. The classification accuracy of the proposed system is compared with the existing systems (Tables 5 - 8). It is observed that the classification accuracy of the proposed system is better for Breast Cancer and Hepatitis and falls in line with other methods for Cardiac Arrhythmia dataset.

7. CONCLUSION

In this paper, a decision support system based on FS-ELM-SA is proposed for the diagnosis of the diabetes. A Gaussian RBF is used as a kernel of ELM and the parameter of ELM is optimized using SA search. The robustness of the proposed system is analyzed in terms of metrics like classification accuracy, sensitivity and specificity using 10-fold cross-validation and confusion matrix. The accuracy of the system for the PID dataset is found to be 96.45%. This model is validated using three more medical datasets of UCI machine learning repository.

REFERENCES

[1] E. I. Mohamed, R., Linderm, G. Perriello, N. Di Daniele, S. J. Poppl, A. De Lorenzo., 2002, “Predicting type 2 Diabetes using an electronic nose-base

artificial neural network analysis", Diabetes Nutrition & Metabolism, Vol. 15, pp.215-221.

[2] Hasan Temurtas, Nejat Yumusak, Feyzullah Temurtas., 2009, "A comparative study on diabetes disease diagnosis using neural networks", Expert Systems with Applications, Elsevier, Vol. 36, pp.8610-8615

[3] Steve Copley., 2013, URL: https://igcseict.info /theory / 7_2/expert.

[4] L. H. Cheng, J. W. Chieh., 2006, "A Genetic Algorithm - Based Feature Selection And Parameters Optimization For Support Vector Machines", Expert Systems with Applications, Elsevier, Vol. 31, pp.231-240.

[5] R. Yuan, B. Guangchen., 2010, "Determination Of Optimal SVM Parameters By Using Genetic Algorithm/Particle Swarm Optimization", Journal of Computers, No. 5, pp.1160-1168.

[6] B. Fayssal, M. A. Chikh., 2013, "Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm", Computer methods and programs in biomedicine, No.1, pp.92-103.

[7] W. P. Santi, E. Abdullah, M. Z. Jasni, S. P. Rahayu., 2009, "A New Smooth Support Vector Machine and Its Applications in Diabetes Disease Diagnosis", Journal of Computer Science., 12(5), pp.1003-1008.

[8] M. Forina., 1991, "Pima Indian Diabetes Dataset", URL: http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes.

[9] R. Payam, T. Lei, L. Huan., 2007, "On Comparison of Feature Selection Algorithms Association for the Advancement of Artificial Intelligence", pp.35-39.

[10] Y.Ersen., 2013, "An Expert System Based on Fisher Score and LS-SVM for Cardiac Arrhythmia Diagnosis", Computational and mathematical Methods in Medicine, pp.1-6.

[11] K. C. Tan, E. J. Teoh, Q. Yua, K. C. Goh., 2009, "A hybrid evolutionary algorithm for attribute selection in data mining", Expert Systems with Applications, Elsevier.

[12] P. David, M. Alan., 2010 "Artificial Intelligence", URL: http://artint.info/html/ArtInt_89.html.

[13] Nick Street., 1996, "UCI machine learning repository", URL: http://archive.ics.uci.edu/ml/datasets/BreastCancerWisconsin.

[14] G. Gong., 1990, "UCI machine learning repository", https://archive.ics.uci.edu/ml/datasets/Hepatitis.

[15] G. H. Altay., 1998, "UCI machine learning repository", http://archive.ics.uci.edu/ml/datasets/Arrhythmia.

[16] N. Mehdi, S. Mehdi, N. T. Adel, Azra Masoumi., 2012, "Hepatitis Disease Diagnosis Using Hybrid Case Based Reasoning and article Swarm Optimization International Scholarly Research Network", ISRN Artificial Intelligence.

[17] C. F. Aliferis, I. Tsamardinos, A. Statnikov, A. Hiton., 2003 "A novel Markov blanket algorithm for optimal variable selection", Proceedings of the AMIA Annual Symposium.

[18] M. B. Li, G. B. Huang, P.Saratchandran and N.Sundarajan., 2005 "Fully Complex Extreme Learning Machine, " Neurocomputing, Vol. 68, pp.306-314.

[19] Biswarup das, Pradee Kumar Verma., 2001"Artificial neural network-based optimalcapacitor switching in a distribution system", Electric Power Systems Research, Vol. 60, pp.55-62.

[20] D.Serre., 2002, "TMatrices: Theory andApplications", Springer, Berlin.

[21] G. B. Huang, L. Chen, and C. K. Siew., 2006 "Universal Approximation using incremental constructive feedforward neural random hidden nodes", IEEE transactions on Neural Networks., 17(4), pp.879-892.