International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 15, Number 2 (2019), pp. 97-113

http://www.ripublication.com

Adaptive Grasshopper Optimization Algorithm based Feature Selection with Enhanced Extreme Learning Machine Classifier for Coronary Artery Heart Disease Prediction

Rajkumar R.[1] and Anandakumar K.[2] and Bharathi A.[1,3]

[1.] *Department of Computer Applications, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India.*

[2.] *Department of Computer Applications, Bannari Amman Institute of Technology, Tamil Nadu, India.*

[3.] *Department of IT, Bannari Amman Institute of Technology, Tamil Nadu, India*

Abstract

Data mining is the broad field of research and is having wide scope of application areas which includes disease prediction. This research work aims to classify patients those who are prone to get CAHD and not prone to get CAHD. At first feature selection is performed by using adaptive grasshopper optimization algorithm and afterwards classification is performed using enhanced ELM. Two datasets are taken into account for evaluating the performance of the proposed classifier using the metrics such as true positive, true negative, false positive, false negative, accuracy, sensitivity, specificity and elapsed time. From the results it is evident that the AGOA – EELM classifier outperforms than that of existing methods taken for comparison.

Keywords: Data mining, heart disease prediction, machine learning, extreme learning machine, optimization, dataset, true positive, true negative, false positive, false negative, accuracy, sensitivity, specificity, elapsed time.

1. INTRODUCTION

Data mining is the expansion of hauling out hidden knowledge from existing data which is proficient to unveil the patterns and relationships among large amount of data in a single or several datasets. Data mining is proved to be implemented and working in many real world applications which include crime detection, risk evaluation and market analysis. A preference of industries like banking, insurance, and marketing are using data mining for cut down costs, and augment profits. Cardiovascular diseases are among the most common reasons of death all over the world. One major type of these diseases is coronary artery heart disease (CAHD). Twenty five percent of people, who have CAHD, die suddenly without any previous symptoms. CAHD is one of the most imperative types of diseases affecting the heart, and possibly lead to severe heart attacks in patients. Being aware of disease symptoms, can aid in time treatment, and reduce the severity of disease's side effects. The motivation of the research work starts from these preliminaries. The problem statement is quite obvious. Patients who have diabetes are more prone to CAHD. There are several machine learning algorithms and data mining techniques are employed in this research arena. Many of the algorithm deals only with proposing a classifier. Only very few literatures are found that focuses on feature selection task before performing the classification task. Hence this research work concentrates on employing the efficient feature selection strategy by employing adaptive grasshopper optimization algorithm. After that, enhanced extreme learning machine is used to perform the task of classification. The main objective of this research work is to reduce the elapsed time of the overall classification task during the testing phase. It is evident that without making use of feature selection strategy, the classifier will consume more time to perform the classification task. Hence there is a wide scope for employing an appropriate feature selection mechanism. This paper is organized as follows. This section gives a quick view of the significance, motivation, problem statement. Section 2 discusses on the related works. Section 3 portrays the proposed work. Section 4 narrates the dataset taken, performance metrics with results and discussions. Section 5 provides the concluding remarks of the paper.

2. RELATED WORKS

Weighted Fuzzy Rule based Clinical Decision Support System [1] was proposed to automatically receive the information from the patient data. At first, weighted fuzzy rule generation approach was carried out to select and receive the attribute data. At last, decision support system based on fuzzy logic was developed. The result of the comparison made with baseline schemes indicate that the system have low sensitivity and specificity.

Alphabet Entropy Method [2] was proposed to classify the cardiac arrhythmias which allow the inferences of prediction markers. With the use of nonlinear entropies, classification was done with the symbolic dynamics. Feature selection was given more priority and extraction was done with the help of random forest algorithm. In the result, F-Measure came with very low value.

Artificial Neural Network based Fuzzy System [3] was proposed to extract and make analysis to predict the heart failure. Variability of heart rate was used as base level signal, which was used as a input for the fuzzy system for the classification. This method was considered as feed-forward method, and result has low F-Measure.

Artificial Neural Networks based Decision System [4] was proposed with the assumption of heart rate attributes having common risk level. Deep analysis on the risk level provides the information that the artificial neural network won't work in a proper manner to detect and classify the status of heart rate, due to giving poor results on classification accuracy.

Enhanced Support Vector Machine Method [5] was proposed to classify the heart diseases among the patients. For the processing of input, Magneto Cardiograph signals were used. Actually it measures the magnetic fields produced from heart. Low Classification accuracy results indicate that the features were not fir for the prediction of heart disease.

Three Dimensional Cardiovascular System [6] was proposed with the base of images related to echo cardiographic. In order to construct the cardiovascular system, couple of algorithm was also proposed. To increase the interactivity, heart vessels thickness was considered as a feature. Low F-Measure shows that the system was not fit for the prediction of congenital heart disease.

Sophisticated Three Dimensional Classifier [7] was proposed to detect the diseases that are present in the heart vessels. This model was dependent on the patients precise pathological circumstance. Feature selection was performed to increase the accuracy, but the classification accuracy didn't got increased.

Data Quality Quantitative Assessment [8] was proposed to enhance the classification of high level frequency sounds from heart for increased classification towards prediction of coronary arteries. It is combined with the correlation analysis to provide an estimation towards the signal to noise ratio. Increased false positive rates shows that the prediction of coronary arteries is not possible only the heart beat sounds.

Sound Feature Selection [9] was proposed with the intention of classify the features based on heart sounds. The processing of digital signal was done with the help of data mining methods. Salient features which describes the low level frequency were identified. The low classification rate indicates that the method need further development to make more accurate results.

Dynamic Bayesian Network [10] was a temporal probability based graphic model, which focus on sequential events, its cause and dependency. It ensembles the concept of temporal abstractions for the risk level of coronary artery disease. A network structure was built to study the parameters. The increased false negative rate indicates that there exist need to improve the method even more.

3. PROPOSED WORK

The proposed work has two phases. In the initial phase, grasshopper optimization algorithm is employed for feature selection. In the second phase extreme learning machine classifier is used for performing the classification task. The reason for choosing extreme learning machine classifier is that in the earlier research works, we have employed neural networks [14] and support vector machine [15]. Also from the literatures it is inferred that ELM outperforms than that of SVM in reducing errors during training the classifier.

3.1. Feature Selection using Adaptive Grasshopper Optimization Algorithm (AGOA)

AGOA begins optimization by creating a set of random solutions. A random initial population matrix of size $\left(NV \times DNF\right)$ is created where NV is the number of initial vectors in population and RNF is the required number of features to be selected. Each vector in the population represents the indices of candidate features. Each individual element in the aforementioned vectors (as referred to the position of the grasshoppers) is an integer value between 1 and the total number of features $\left(NoF\right)$ .Thus, the lower boundary of the search space along each dimension, referred to by lb , is lb = 1, while the upper boundary, referred to by ub , is $ub = NoF$.

All candidate agents are evaluated with regard to a fitness value and the best search agent in the current population is considered as target. The error rate of classification is considered as the fitness value. After that, the search agents of the AGOA update their positions and also the decreasing factor is obtained. During every cycle (iteration), the position of the best target obtained so far keeps on updated. Updating the position is performed iteratively until the maximum number of iteration is reached. The position and fitness of the best target are finally returned as the best approximation for the global optimum.

In this AOGA, additionally the obtained values in the process of updating the position of the grasshoppers are set in the interval between $\left[1 - NoF\right]$. For that reason, the out of range values are identified and replaced with new random integers in the range of $\left[1 - NoF\right]$. Another contribution is AGOA elimination of repetitive values in the obtained vectors. A new feature distribution factor is added in AGOA to help the substitution of the duplicate features. By this added enhancement in AGOA, the probabilities of each feature while being used in forming promising subsets (i.e. subsets whose fitness value is less than the mean fitness of the whole population) or in less competitive subsets (i.e., subsets whose fitness is higher than the mean fitness of the whole population) within the current iteration are calculated. Accordingly, we define the feature goodness factor for the feature **j** in iteration **i** as FG_{ji} and calculate it by

$$FG_{ji} = \alpha PG_j + (1 - PB_j)(NoF - DNF)/NoF \quad \ldots (1)$$

where PG_j is the probability of using feature j in the promising subsets. PB_j is the probability of using feature j in the less competitive subsets. NoF is the total number of features and RNF is the required number of features to be selected. α is a positive constant that emphasizes the importance of features in good subsets. In this work, this parameter is set to $\alpha = 2$. To compute the distribution probabilities of features, take an example of the first feature which has been used twice in good subsets and four times in bad subsets. Thus, PG = 2/10, PB=4/10, and $FG = 2 \times 2/10 + (1 - 4/10)(6 - 3)/6 = 7/10$. Finally, the feature indices are sorted according to the highest FG value and its position in the FG vector, and then the next higher value and so on in a descending order.

Higher FG_j values indicate better features which can be used to replace the duplicate features in the trial vector. The justification behind this idea is from the fact that the first term in Eq. (1) indicates the degree to which feature j contributes in forming promising subsets while the second term indicates the weighted probability that feature j is not selected in less competitive subsets. The factor $(NoF - DNF)/NoF$ is used as a variable weight. Thus, when small number of features is required to be selected, the role of the second term is more prominent since this factor is close to 1. Hence the features that are best will be obtained using AGOA. Once AGOA selects the features, then ELM and Enhanced ELM are used for classification which is described in the next sub section.

3.2. Extreme Learning Machine Classifier

In general ELM is modeled as a single hidden layer feed forward neural network not including tuning of the hidden layer. One of the primary feature of ELM when comparing with other machine learning classifier is the it overcome limitations of back propagation algorithms which are commonly used in artificial neural networks by randomly generating input weights and analytically calculating output weights. Learning speed and performance of ELM and Enhanced ELM are significantly better than other conventional learning algorithms. There are three different layers in ELM: the input layer, the hidden layer and the output layer. It is understand that ELM has M neurons in the input layer, K neurons in the hidden layer and C neurons in the output layer, for N arbitrary distinct samples $(\vec{x}_i, \vec{t}_i)$, where $\vec{x}_i = [x_{i1}, x_{i2}, x_{i3}, ..., x_{im}]^T \in R^M$ and $\vec{t}_i = [t_{i1}, t_{i2}, t_{i3}, ..., t_{iC}] \in R^C$. The m, i and c represent the index of features, samples and neurons respectively. ELM with K hidden neurons is mathematically modeled as

$$\vec{o}_i = \sum_{k=1}^{K} \left[\vec{\beta}_k \cdot G\left(\vec{w}_k, \vec{b}_k, \vec{x}_k\right)\right] = \sum_{k=1}^{K} \left[\vec{\beta}_k \cdot G\left(\vec{w}_k, \vec{x}_i, b_k\right)\right]$$
$$\left(\vec{o}_i, \vec{\beta}_k \in R^C, \vec{w}_k \in R^M, b_k \in R, i = 1, 2, ..., N\right) \quad \ldots (2)$$

where $\vec{w}_i = [w_{k1}, w_{k2}, w_{k3}, ..., w_{km}]^T$ and b k are random input weights in the k - th hidden node, $\vec{\beta}_i = [\beta_{k1}, \beta_{k2}, \beta_{k3}, ..., \beta_{kc}]^T$ is the weight vector connecting the k - th hidden node and the output nodes, $\vec{o}_i = [o_{i1}, o_{i2}, o_{i3}, ..., o_{ic}]$ is the i - th output vector of ELM, and finally G(*) corresponds to an output of an activation function used in the neurons of the hidden layer. Particularly, the value of elements in $\vec{t}_i$ is 1 when the output of neuron belongs to the sample class and the rest are -1. ELM can evaluate these N samples with zero error, which is the basic principle of least squares algorithm. The evaluation is shown in

$$\sum_{n=1}^{N} \left\|\vec{o}_i - \vec{t}_i\right\| = 0 \quad \ldots (3)$$

and can be expressed as

$$H \cdot \beta = T \quad \ldots (4)$$

where

$$H = \begin{bmatrix} G(\vec{w}_1, b_1, \vec{x}_1) & . & . & . & G(\vec{w}_k, b_k, \vec{x}_1) \\ . & & . & & . \\ . & & & . & . \\ . & & & . & . \\ G(\vec{w}_1, b_1, \vec{x}_n) & . & . & . & G(\vec{w}_k, b_k, \vec{x}_n) \end{bmatrix}_{N \times K}$$
$$H = \begin{bmatrix} g_{11} & . & . & . & g_{1k} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ g_{n1} & . & . & . & g_{nk} \end{bmatrix}_{N \times K} \quad \ldots (5)$$

$$\beta = \begin{bmatrix} \vec{\beta}_1 \\ . \\ . \\ . \\ \vec{\beta}_k \end{bmatrix}_{K\times C} \quad and \quad T = \begin{bmatrix} \vec{t}_1 \\ . \\ . \\ . \\ \vec{t}_n \end{bmatrix}_{N\times C} \quad \ldots (6)$$

H is named as the hidden layer output matrix of ELM with a specific input dataset $X = [\vec{x}_1, \vec{x}_2, \vec{x}_3, ..., \vec{x}_n]$. The smallest norm least-squares solution of above linear system can be expressed as:

$$\vec{\beta} = H \cdot T = \left(H^T H\right)^{-1} H^T H \quad \ldots (7)$$

$$H^T H = \begin{bmatrix} u_{11} & . & . & . & u_{1k} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ u_{k1} & . & . & . & u_{kk} \end{bmatrix}_{K\times K} \quad \ldots (8)$$

$$H^T T = \begin{bmatrix} v_{11} & . & . & . & v_{1c} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ v_{k1} & . & . & . & v_{kc} \end{bmatrix}_{K\times C} \quad \ldots (9)$$

where H is the pseudo inverse which extrapolates the inverse of matrix H in Eq. (4) .

3.3. Enhanced ELM Classifier

In order to speed up the learning process in Enhanced ELM an incremental learning method is used in this research work.

Let the original training dataset be D = (X,T) where $D = \left\{\left(\vec{x}_j, \vec{t}_j\right) | \, \vec{x}_j \in R^M, \, j = 1, 2, ..., N\right\}$and N is the number of original samples. The newly arrived training dataset is represented by $\Delta D = \left(\Delta X, \Delta T\right)$,where $\Delta D = \left\{\left(\vec{x}_j, \vec{t}_j\right) | \, \vec{x}_j \in R^M, \, j = N+1, N+2, ..., N+\Delta N\right\}$ and ΔN is the number of new samples. When the newly arrived training dataset is merged with the original one, we have $D' = \left(\begin{bmatrix} X \\ \Delta X \end{bmatrix}, \begin{bmatrix} T \\ \Delta T \end{bmatrix}\right) = \left\{\left(\vec{x}_j, \vec{t}_j\right) | \, \vec{x}_j \in R^M, \, j = 1, 2, ..., N, N+1, N+2, ..., N+\Delta N\right\}$

The new hidden layer output matrix H' can be derived from H and ΔH. Now we have H' and T', where

$$H' = \begin{bmatrix} H \\ \Delta H \end{bmatrix} \; and \;\; T' = \begin{bmatrix} T \\ \Delta T \end{bmatrix} \; \ldots (10)$$

According to the matrix multiplication operator, Eqs. (8) and (9) is mathematically modeled as

$$H'^T H' = \begin{bmatrix} H \\ \Delta H \end{bmatrix}^T \begin{bmatrix} H \\ \Delta H \end{bmatrix} = \begin{bmatrix} H^T & \Delta H^T \end{bmatrix} \begin{bmatrix} H \\ \Delta H \end{bmatrix} = H^T H + \Delta H^T \Delta H \; \ldots (11)$$

$$H'^T T' = \begin{bmatrix} H \\ \Delta H \end{bmatrix}^T \begin{bmatrix} T \\ \Delta T \end{bmatrix} = \begin{bmatrix} H^T & \Delta H^T \end{bmatrix} \begin{bmatrix} T \\ \Delta T \end{bmatrix} = H^T T + \Delta H^T \Delta T \; \ldots (12)$$

As their approach defines $U = H^T H$ and $V = H^T T$, Eq. (11) and(12) are presented as

$$U' = U + \Delta U \;\; and \;\; V' = V + \Delta V \; \ldots (13)$$

And Eq. (7) is mathematically modeled as

$$\vec{\beta} = U^{-1} V \; \ldots (14)$$

Dual Class ELM

ELM usually deals with the data with more classes, while the proposed Enhanced ELM derived that E. By examining Eqs. (2) and (4) , when a target class T is dual class, namely all 1s and 0s in T, β becomes a linear approximation mapping from H to T , which geometrically is a hyper plane approximation. Then a distance of any point over the sample to the hyperplane constructed by the Enhanced ELM is defined as

$$d = |H \cdot \beta - T| \; \ldots (15).$$

By measuring the distance, it shows that Enhanced ELM is capable enough to classify patients those who are prone to CAHD.

4. DATASETS AND PERFORMANCE METRICS

Two datasets namely Statlog Heart dataset [16] and UCI heart disease PIMA dataset [17] are taken for performance evaluation. Statlog heart dataset contains 270 instances with 14 attributes including class label. Out of the 270 instances 120 instances are true positive and remaining 150 instances are true negative. PIMA dataset contains 768 instances with 9 attributes including class label. Out of the 768 instances 268 instances are true positive and remaining 500 instances are true negative. The details of the dataset are given in chart format in table 1. The performance metrics namely true positive, true negative, false positive, false negative, accuracy, sensitivity, specificity, elapsed time are taken for comparison with other classifiers.

Table 1. Dataset Details

Statlog dataset		**PIMA dataset**	
Total instances	270	**Total instances**	768
Number of attributes	14	**Number of attributes**	9
Attributes	1) age 2) sex 3) chest pain type 4) resting blood pressure 5) serum cholestoral 6) fasting blood sugar 7) resting electrocardiographic results 8) maximum heart rate achieved 9) exercise induced angina 10) oldpeak 11) the slope of the peak exercise 12) number of major vessels 13) thallium 14) class label	**Attributes**	1) Number of times pregnant 2) Plasma glucose concentration 3) Diastolic blood pressure 4) Triceps skin fold thickness 5) 2-Hour serum insulin 6) Body mass index 7) Diabetes pedigree function 8) Age 9) Class label
TP	120	**TP**	268
TN	150	**TN**	500

RESULTS AND DISCUSSIONS

Table 2. Performance Evaluation – Statlog Dataset

Method	TP	TN	FP	FN	Accuracy (in %)	Sensitivity (in %)	Specificity (in %)	Elapsed Time (in seconds)
Firefly Algorithm[11]	104	122	21	23	83.70	81.89	85.31	497.248
Neural Networks Classifier [12]	112	126	16	16	88.15	87.50	88.73	361.881
Modified Differential Evolution with Neural Networks [13]	113	127	15	15	88.89	88.28	89.44	318.765
Neuro Fuzzy Classifier [14]	116	133	11	10	92.22	92.06	92.36	223.324
Improved Bacterial Foraging Optimization based Twin Support Vector Machine [15]	117	136	9	8	93.70	93.60	93.79	38.475
Enhanced ELM with AGOA	119	148	2	1	98.89	99.17	98.67	12.218

Table 2. Performance Evaluation – PIMA Dataset

Method	TP	TN	FP	FN	Accuracy (in %)	Sensitivity (in %)	Specificity (in %)	Elapsed Time (in seconds)
Firefly Algorithm[11]	207	435	61	65	83.59	76.10	87.70	1527.781
Neural Networks Classifier [12]	231	439	48	50	87.24	82.21	90.14	1304.362
Modified Differential Evolution with Neural Networks [13]	235	441	44	48	88.02	83.04	90.93	1086.115
Neuro Fuzzy Classifier [14]	242	450	37	39	90.10	86.12	92.40	697.164
Improved Bacterial Foraging Optimization based Twin Support Vector Machine [15]	256	455	30	27	92.58	90.46	93.81	108.574
Enhanced ELM with AGOA	266	483	10	9	97.53	96.73	97.97	39.982

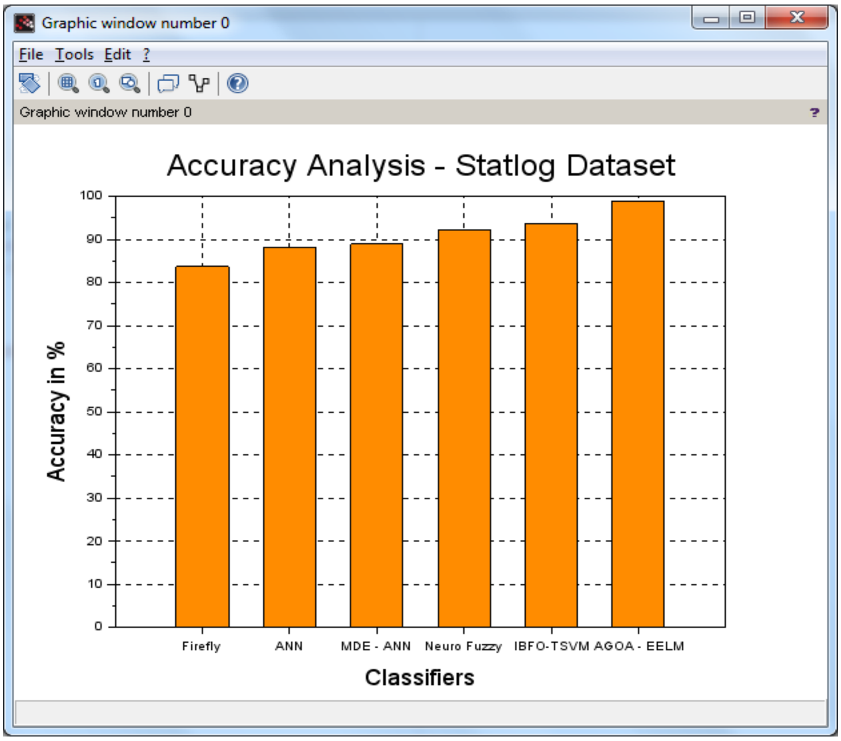

Fig.1. Accuracy Analysis

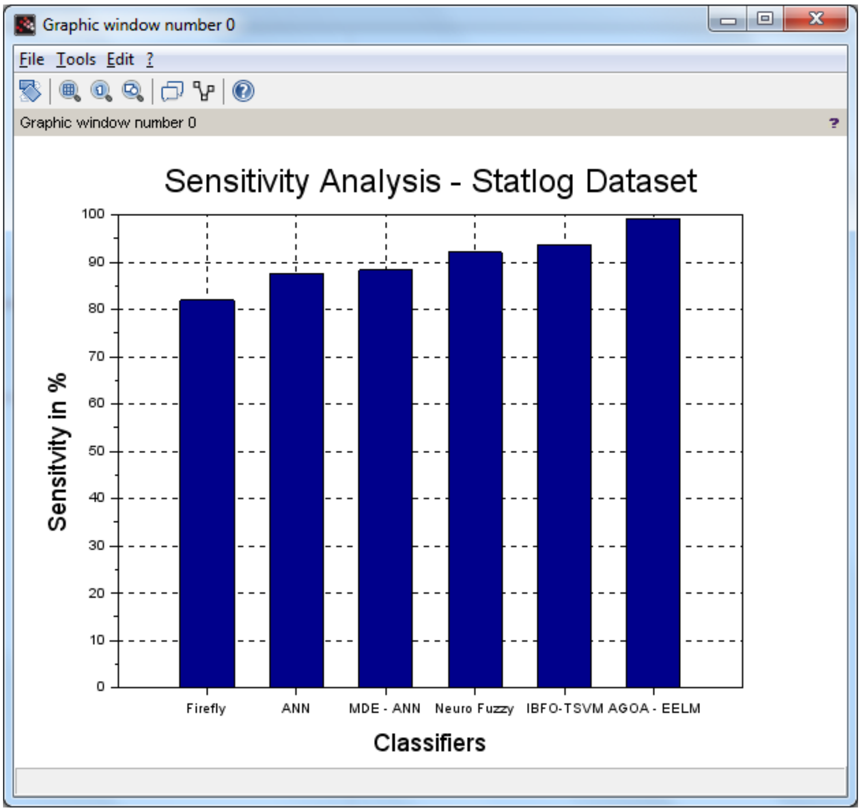

Fig.2. Sensitivity Analysis

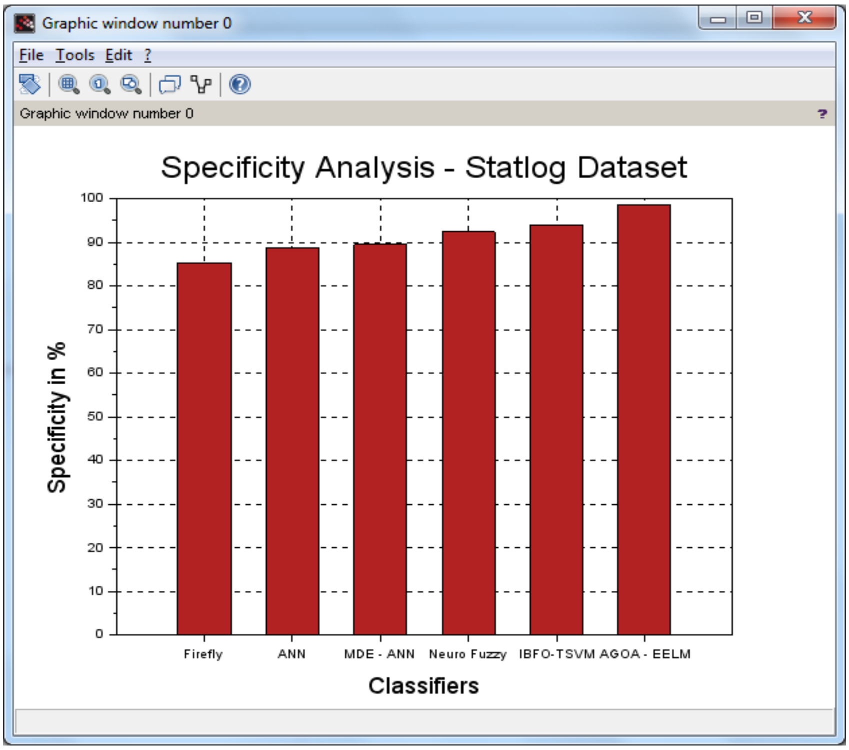

Fig.3. Specificity Analysis

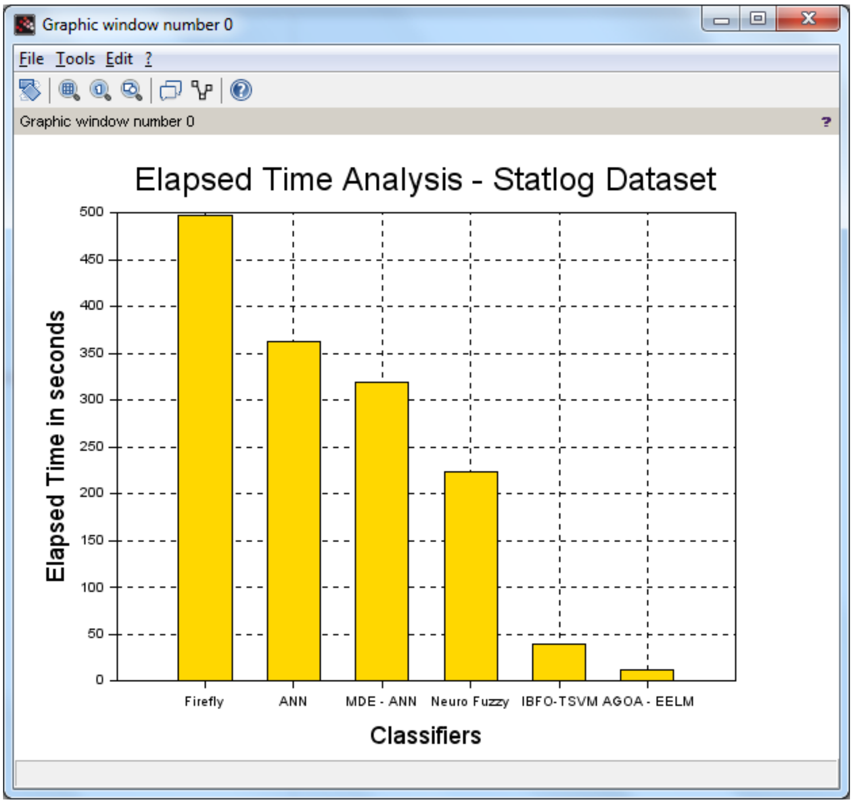

Fig.4. Elapsed Time Analysis

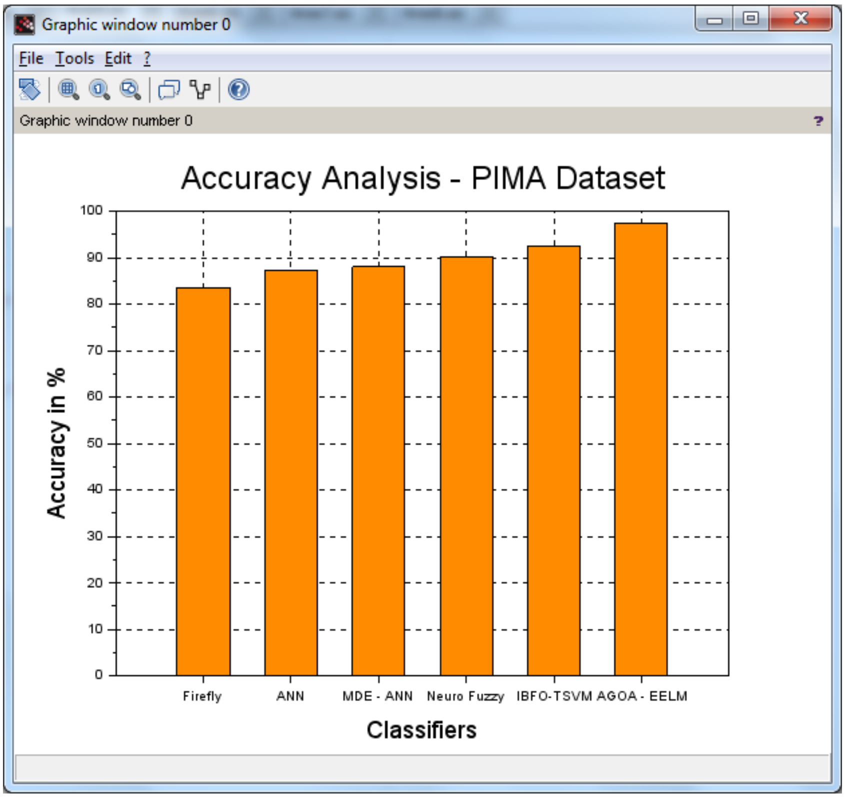

Fig.5. Accuracy Analysis

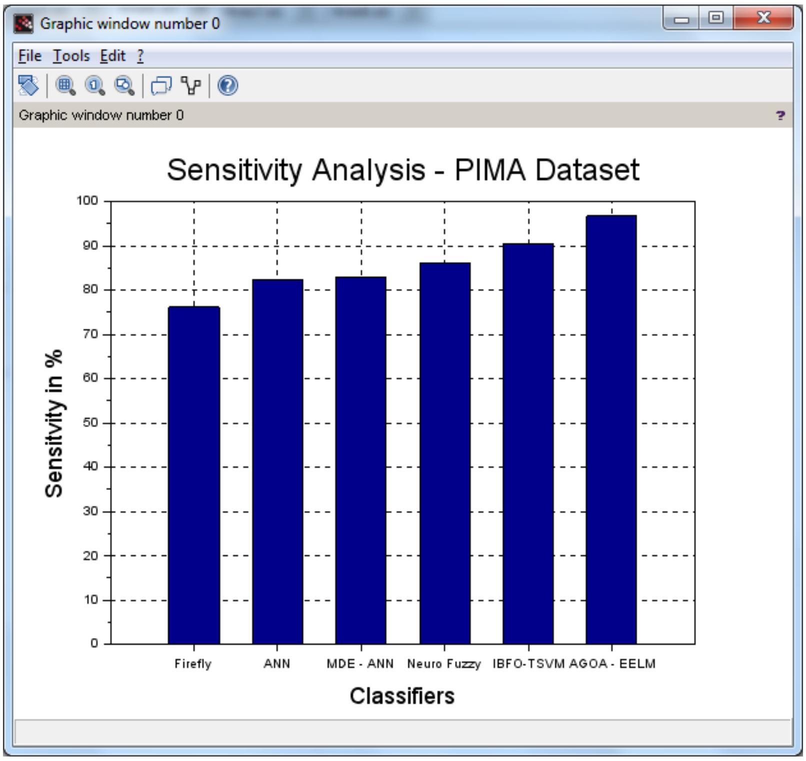

Fig.6. Sensitivity Analysis

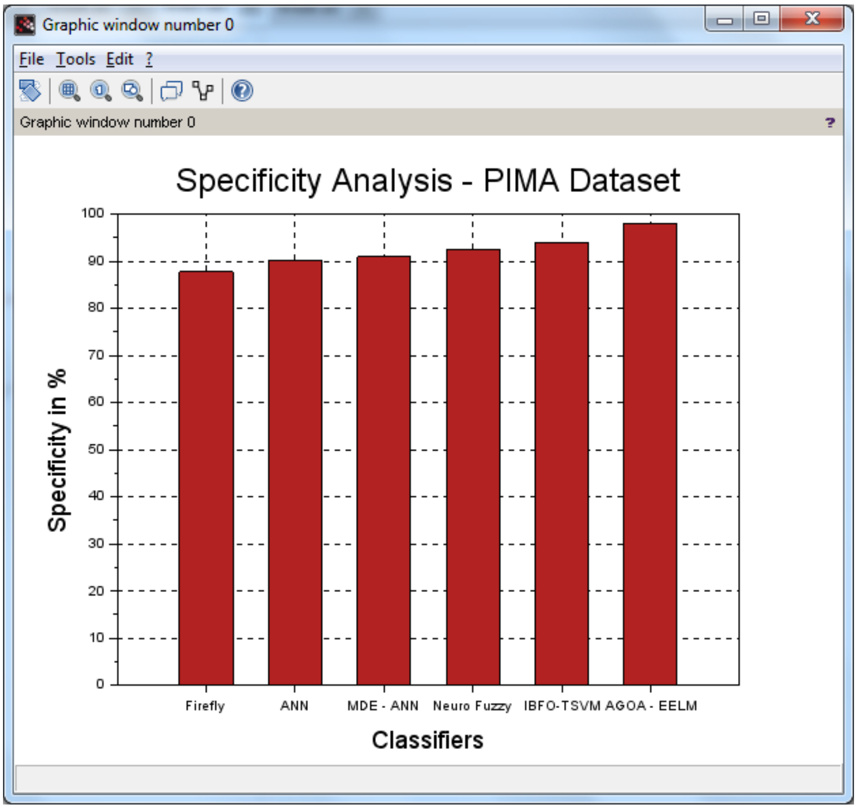

Fig.7. Specificity Analysis

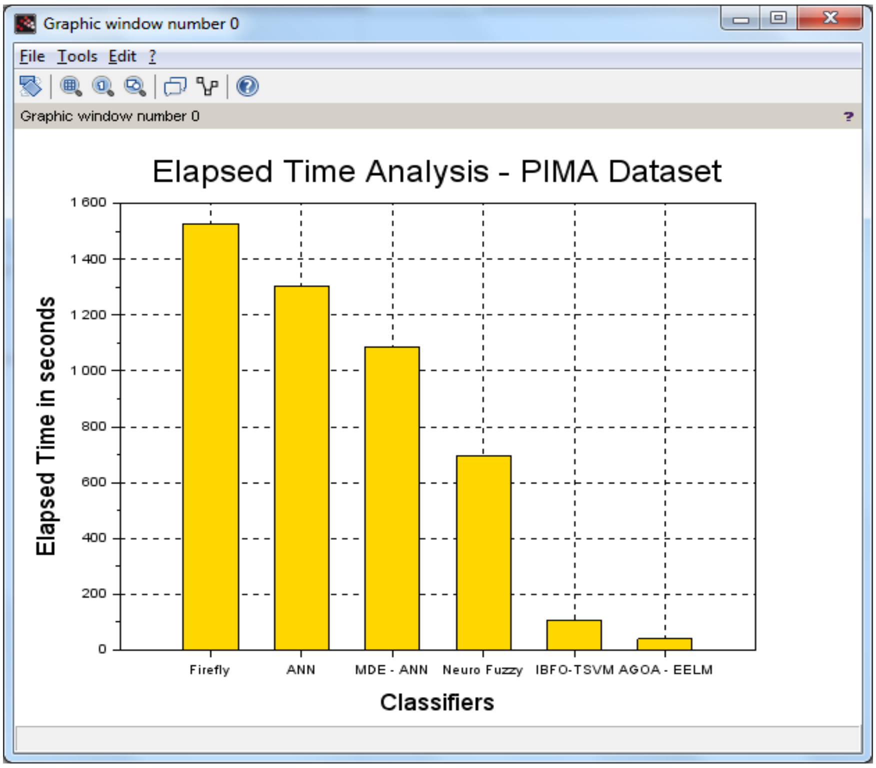

Fig.8. Elapsed Time Analysis

Fig.1. shows the performance of the classifiers in terms of accuracy for the statlog dataset. From the results it is inferred that the proposed Enhanced ELM with AGOA obtains 98.89% accuracy which is better than that of all classifiers. It is because of the improved true positive and true negative values as well as reduced false positive and false negative values. The same also impacts on sensitivity which is obtained 99.17 % (shown in Fig.2) and specificity obtained 98.67% (shown in Fig.3). The total elapsed time is eventually reduced to 12.218 seconds (shown in Fig.4) which is significantly less when compared to other classifiers.

Fig.5. shows the performance of the classifiers in terms of accuracy for the PIMA dataset. From the results it is inferred that the proposed Enhanced ELM with AGOA obtains 97.53% accuracy which is better than that of all classifiers. It is because of the improved true positive and true negative values as well as reduced false positive and false negative values. The same also impacts on sensitivity which is obtained 96.73 % (shown in Fig.6) and specificity obtained 97.97% (shown in Fig.7). The total elapsed time is eventually reduced to 39.982 seconds (shown in Fig.8) which is significantly less when compared to other classifiers.

5. CONCLUSION

Extreme learning machine classifier is enhanced in this research work and adaptive grasshopper optimization is employed for performing the feature selection task. The reduced feature set is given as the input for the Enhanced ELM classifier. This work is the extension of the previous work done namely Neuro Fuzzy classifier and Improved Bacterial Foraging Optimization based Twin Support Vector Machine Classifier. From the obtained results it is understood that Enhanced ELM with AGOA feature selection outperforms than that of existing chosen classifiers in terms of selected performance metrics.

REFERENCES

[1] P. K. Anooj, Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules, Journal of King Saud University - Computer and Information Sciences, Vol 24, pp 27-40, 2012.

[2] A. Jovic, F. Jovic, Classification of cardiac arrhythmias based on alphabet entropy of heart rate variability time series, Biomedical Signal Processing and Control, vol.31, pp 217-230, 2017.

[3] U. R. Acharya, P. S. Bhat, S.S. Iyengar, A. Rao, S. Dua, Classification of heart rate data using artificial neural network and fuzzy equivalence relation, Pattern Recognition, Vol 36, pp 61-68, 2003.

[4] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, G. Li, An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction, Expert Systems with Applications, Vol 68, pp 163-172, 2017.

[5] H. K. Lim, N. Chung, Y. Ko, Y. Lee, Y. K. Park, Magnetocardiogram Difference Between Healthy Subjects and Ischemic Heart Disease Patients, IEEE Transactions on Magnetics, Vol 45, No 6, pp 2890-2893, 2009.

[6] M. Nakao, Cardiovascular Modeling of Congenital Heart Disease Based on Neonatal Echocardiographic Images, IEEE Transactions on Information Technology in Biomedicine, Vol 16, No. 1, pp 70-79, 2012.

[7] A. Redaelli, A Model of Health: Mathematical modeling tools play an important role in optimizing new treatment options for heart disease., IEEE Pulse, Vol 6, No 4, pp 27-32, 2015.

[8] J. L. Semmlow, Improved Heart Sound Detection and Signal-to-Noise Estimation Using a Low-Mass Sensor, IEEE Transactions on Biomedical Engineering, Vol 63, No 3, pp 647-652, 2016.

[9] S. E. Schmidt, C. Holst-Hansen, J. Hansen, E. Toft, J. J. Struijk, Acoustic Features for the Identification of Coronary Artery Disease, IEEE Transactions on Biomedical Engineering, Vol 62, No 11, pp 2611-2619, 2015.

[10] K. Orphanou, A. Stassopoulou, E. Keravnou, DBN-Extended: A Dynamic Bayesian Network Model Extended With Temporal Abstractions for Coronary Heart Disease Prognosis, IEEE Journal of Biomedical and Health Informatics, Vol 20, No 3, pp 944-952, 2016.

[11] N. C. Long, P. Meesad, H. Unger, “A Highly Accurate Firefly Based Algorithm for Heart Disease Prediction”, Expert Systems with Applications, vol. 42, no. 21, pp. 8221 – 8231, 2015.

[12] C. H. Weng, T. C. K. Huang, R. P. Han, “Disease prediction with different types of neural network classifiers”, Telematics and Informatics, vol. 33, no. 2, pp. 277 – 292, 2016.

[13] T. Vivekanandan, N. Ch Sriman Narayana Iyengar, “Optimal Feature Selection using a Modified Differential Evolution Algorithm and its Effectiveness for Prediction of Heart Disease”, Computers in Biology and Medicine, vol. 90, pp. 125 – 136, 2017.

[14] R. Rajkumar, K. Anandakumar, A. Bharathi, “Risk Level Classification of Coronary Artery Heart Disease in Diabetic Patients using Neuro Fuzzy Classifier”, International Journal of Computational Intelligence Research, vol. 13, no. 4, pp. 575 – 582, 2017.

[15] R. Rajkumar, K. Anandakumar, A. Bharathi, "Improved Bacterial Foraging Optimization based Twin Support Vector Machine (IBFO-TSVM) Classifier for Risk Level Classification of Coronary Artery Heart Disease in Diabetic Patients", International Journal of Applied Engineering Research, vol. 13, no. 3, pp. 1716 – 1721, 2018.

[16] http://archive.ics.uci.edu/ml/datasets/statlog+(heart)

[17] https://archive.ics.uci.edu/ml/datasets/diabetes