Hybrid Firefly-Bat Optimized Fuzzy Artificial Neural Network Based Classifier for Diabetes Diagnosis

Gadekallu Thippa Reddy[1*]   Neelu Khare[2]

Vellore Institute of Technology University, Vellore 632014, India
* Corresponding author's Email: krish.chaitanya143@gmail.com

Abstract: Huge amount of medical data is available today. In order to predict the disease we need a reliable method to diagnose the disease. In this paper we introduce a technique known as FFBAT-ANN prediction algorithm which is categorized as Feature reduction and Diabetes disease classification and such a process is carried out using LPP algorithm and FFBAT- artificial neural network classifier respectively. Initially, LPP algorithm is employed to produce fuzzy rules by recognizing the attributes corresponding to the diabetes disease. Later on, the classification is carried out by blending FFBAT optimization technique with Artificial Neural Network classifier. Finally, the experiment is performed on diabetes dataset. The experimental result proves that the performance of our proposed method outperforms the other conventional methods.

Keywords: Diabetes classification, Firefly-BAT (FFBAT) optimization algorithm, Artificial Neural Networks (ANN) classifier, Feature reduction, Locality Preserving Projection (LPP) algorithm, Membership functions.

1. Introduction

Data mining is a computational process of extracting patterns and knowledge from the large database which revolves around following processes, namely (1) Exploring the given data, (2) Building models or identifying patterns in the given data set and (3) Model identification; where the best fit model is selected from candidate list of models [1]. There are many applications we can visualize, where data mining can be applied. Among them, medical application is the extensive one [2]. However in medical applications, the extraction of original medical data is a difficult task due to missing and irrelevant data in the data set. In order to analyse and identify the relevant data from large data set, the medical decision support system is widely used and hence it helps to make clinical decisions more accurately.

Nowadays, Computer based patient record technology plays a great emphasis on generating huge health care data and provides manageable access to them and hence soft computing techniques facilitate the evaluation and analysis on them [3]. Hence, soft computing techniques have been extensively used in medicine, which has become highly multidisciplinary area connecting medicine, computer science, statistics and many others in the last two decades. Though a comprehensive comparative study of soft computing techniques in the context of biomedical data, it offered a great challenge in high-dimensional feature space [4-6].

In this paper, we propose a novel FFBAT-ANN classification algorithm by hybridizing the FFBAT and the Artificial Neural Networks classifier and it is categorised as i) feature reduction and ii) Diabetes disease classification using LPP algorithm and FFBAT-ANN classifier respectively.

The paper is structured as follows: Section 2 addresses the literature review related to the knowledge discovery of disease using different algorithms. Section 3 presents our proposed FFBAT-ANN classification algorithm and provides a detailed description of the disease classification. Section 4 presents the result and discussion of database collected from Sree Diabetic Care Center. Finally, the conclusion part is presented in section 5.

2. Review on Related Works

Nowadays, over 246 million people with the majority of women have been affected by diabetes disease. As per the WHO report, the number of affected people count is expected to rise to over 380 million in the upcoming year of 2025 and it is named as the fifth deadliest disease in the United States with no imminent cure in sight [1]. Iyer *et al.* [7] suggested the classification analysis of the pattern found in the raw data to diagnose the disease using Decision Tree and Naive Bayes algorithms. Esfandiari *et al.* [8] studied the relative work of utility of data mining in knowledge discovery in medicine. Wang and Kang [9] addresses the type 2 diabetes mellitus of the real data comprising 382 samples collected from the Hospital of BUPT. The reliability of Artificial Neural networks in data mining has been investigated by Lu *et al.* [10]. In the advancement of medical data diagnosis and classification, Seera *et al.* [11] suggested the hybridization of intelligent systems such as Fuzzy Min–Max neural network, the Classification and Regression Tree, and the Random Forest model. Eventhough the artificial neural networks posses several advantages over medical diagnosis, Amato *et al.* have demonstrated the philosophy, capabilities, and limitations through selected examples of medical data [12]. Since the coronary heart disease (CHD) caused from traditional genetic factors, an artificial neural networks-based (ANNs) diagnostic model has been proposed by Atkov *et al.*[13]. Barkat, Bardley [14] demonstrated the support vector machine and eclectic rule extraction for artificial neural network. C5 decision tree learner is discussed in [14]. Moreover, to disclose the hidden knowledge and patterns in the large database, various data mining techniques have been constructed throughout the various fields by many researchers [15]. More precisely, the main factor that affects the survival rate of breast cancer one of the application of data mining could be comprehensively predicted and classified by Dursan *et al*. and Neha *et al*. [16]. Mangai and Kumar, [17] stated that in medical diagnosis and making healthcare decisions, how the various classification techniques have been supported. Furthermore many researchers in their literature were focussed and investigated on obtaining the accurate result for the medical data. As for producing the accurate result, another significant characteristics of feature selection in data mining has been introduced that performs pre-processing on any given database [18,19]. This helps to perceive the effectiveness of data mining tool with the requirement of pre-processing the given data by healthcare professionals and medical researchers [20]. Nedic *et al.* [21] used genetic algorithms, Hooke and Jeeves algorithm, simulated annealing and particle swarm optimization algorithms together with data mining to model traffic noise.

For classifying the diabetes disease, a novel algorithm called FFBAT- Optimized Rule Based Fuzzy Logic Classifier has introduced by ThippaReddy and Neelukhare [22]. Here, initially the pre-processing of data has been carried out using Locality Preserving Projection (LPP) algorithm from which the relevant attributes are selected. Eventually RBFL classifier is used to classify the diabetes by generating fuzzy rules and it gets optimized by means of FFBAT algorithm. As per the optimized fuzzy rules and membership functions, fuzzy system has been successfully designed and classified the diabetes data. Particularly, FFBAT is the hybridized optimization algorithm which involves the aspects of both FF and BAT optimization algorithm.

3. Diabetes Disease Classification Using FFBAT-ANN Classifier

We have applied FFBAT algorithm to produce the optimal fuzzy rules and the optimized rules are fed to the ANN classifier to classify the diabetes disease. The Fig. 1 illustrates the process of FFBAT-Artificial Neural Network classifier (FFBAT-ANN).

3.1 Feature Reduction Using LPP

Initially we have to reduce the dimensionality of the dataset without any loss of data in the dataset. In addition, we reduce the number of features and remove the unrelated, unnecessary or noisy information. We use the Locality Preserving Projection algorithm (LPP) to decrease the dimensionality of the dataset. Locality Preserving Projections in the pre-processing step of data mining represents a linear approximation of the nonlinear Laplacian Eigen maps which has the feature of optimally protect the neighbourhood structure of the dataset.

3.2 FFBAT-ANN Classifier

After reducing the irrelevant attributes from the input dataset and optimal fuzzy rule generation, the classification of diabetes disease is performed by rule based Artificial Neural Networks (ANN). In this paper, optimal rules are obtained from dimension reduced dataset by means of FFBAT algorithm. In many literatures, Firefly [23-32] and

BAT [33-40] algorithms have been used effectively for optimization. Although, Firefly algorithm (FA) performs local search as well by a fixed parameter, it has severe limitation on experiencing the confinement of various local optimal solution such as slower process and less convergence rate. In addition, the implementation of FA makes to move the fireflies regardless of its earlier better positions for an individual firefly involved in the swarm and thereby end up its search by missing their situations. Besides, the BAT algorithm implementation is more complicated than other meta-heuristic algorithms since the parameters are not fixed i.e. for an individual bat has the varying set of parameters namely its position, velocity, pulse rate, loudness and frequency. Also this algorithm gets trapped in local minima. We have assumed that classification performance will be increased by combining FF and BAT algorithms. In our work, FF and BAT are combined and employed for optimal rule generation process in neural network-fuzzy classifier. The result section illustrates that the suggested classifier of rule based-fuzzy neural networks classifier attained improved performance than the individual process. The proposed FFBAT-ANN classifier based diabetes disease classification, includes two stages, of (1) optimal Rule generation and (2) classification using FFBAT algorithm and fuzzy neural network classifier respectively.
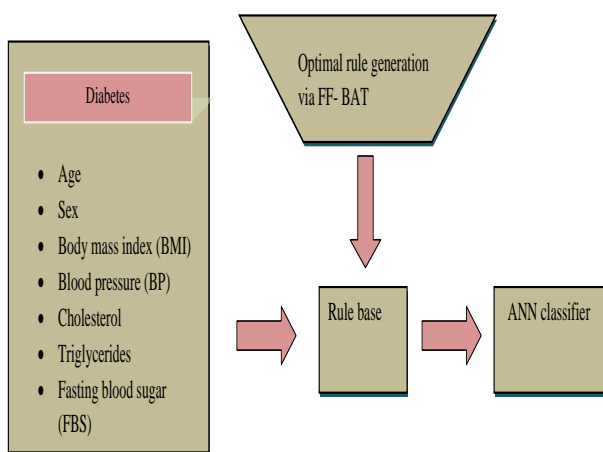

Figure .1   Architecture of the proposed algorithm

Rule Index	Attribute			Class Label	Operator
1	$A_1, A_2, A_3, \ldots, A_n$			$C_1, C_2, C_3, \ldots, C_l$	AND
2	L	M	H	L H	OR

Figure .2   Solution Encoding

Stage 1: Optimal rule generation using FFBAT algorithm

For various applications, BAT [41] and Firefly (FF), [42] optimization algorithms are used to obtain the optimal best solution. Here, we combine these two algorithms to generate optimal rules which help to improve the classification accuracy. For the classification of diabetes disease, the input database is segmented into two sets namely, training dataset ( ) to generate the optimal fuzzy rules and testing dataset ( ). The process of optimal rule generation using FFBAT algorithm is explained using following steps.

Discretization: The discretization function is defined as the conversion of input data into corresponding discrete data. Let $D^{TR} = \{A_{ij}; 0 \le i \le m \ and \ 0 \le j \le n\}$ be the input data with n number of attributes is given to the discretization function. In general, discretization is the primary step in data processing applications in which it converts the data into specified interval.

$$Dev^j = \left[\frac{\max(A_j) - \min(A_j)}{3} \right] \forall j \qquad (1)$$

$$D^L = \min(A_j) \le \left[\min(A_j) + Dev^j \right] \qquad (2)$$

$$D^M = \left[\min(A_j) + Dev^j \right] \le \left[\min(A_j) + 2 \bullet Dev^j \right] \qquad (3)$$

$$D^H = \left[\min(A_j) + 2 \bullet Dev^j \right] \le \max(A_j) j \qquad (4)$$

Using Eq. (2), (3) and (4), we can adjust all the attribute values in the specified interval. Then the input data values within specified range get discretized by replacing the interval values. Subsequently, the training dataset is also converted into the discretized format with the data elements of L, M, and H.

Solution Encoding: Since for optimizing the solution quickly, FFBAT algorithm initially generates arbitrary population of solution with randomly selecting solutions of the whole data in the dataset. Every individual solution performs two logical rules, namely AND and OR operations. Moreover, dimension reduced dataset produces the number of random solutions. A logical rule based on the solution encoding diagram is given in Fig. 2. The dimension of the solution is M X N. Here N represents twice the number of classes whereas M value is calculated using Eq. (5)

$$M = F_{Red} + C_N + O_p \qquad (5)$$

Where, F_{Red} denotes reduced attribute, C_N represents Number of Classes, O_p is operator.

Fitness Evaluation: The selection of the fitness function is a crucial aspect in FFBAT algorithm. It is used to evaluate the aptitude (goodness) of candidate solutions. Here, classification accuracy is the main criteria used to design a fitness function. The fitness computation is executed for each solution. For each iteration, the fitness is calculated using Eq. (6),

$$Fitness = \frac{T_1 + T_2}{(T_1 + T_2 + F_1 + F_2)} \qquad (6)$$

where, T_1 is True positive, F_1 is False positive, T_2 is True negative, F_2 is False negative. T_1 is one that detects the condition when the condition is present and T_2 is one that does not detect the condition when the condition is absent. In addition, F_1 is one that detects the condition when the condition is absent and F_2 is one that does not detect the condition when the condition is present.

Hybrid FF-BAT Algorithm: The novel hybrid FFBAT optimization algorithm produces optimal rules to fuzzy logic classifier. The fundamental step of the hybrid FFBAT algorithm is discussed below:

Step 1: Randomly generate N number of initial solutions.

Step 2: Evaluating the fitness function in Eq. (6).

Step 3: Set the number of iterations required to produce the best solution in firefly algorithm and then update the solution using Eq. (7):

$$S_i^{t+1} = S_i^t - \beta_0^{at} e^{-D_{ij}^2}(S_j^t - S_i^t) + \sigma_t \mu \qquad (7)$$

In the above equation S_i^{t+1} denotes the updated i^{th} solution, S_i^t denote the current i^{th} solution and S_j^t denotes the jth solution which is the brighter firefly. Moreover σ_t represents the randomized parameter for every iteration and μ denotes random number vector obtained from the Gaussian distribution; β_0^{at}

is the constant value correlated with the firefly attractiveness.

Step 4: After reaches the first iteration of firefly algorithm, N number of fitness values are obtained.

Step 5: Set the number of iterations required to generate best solution in BAT algorithm. The equation to update the position of ith solution is as follows:

$$S_i^{t+1} = S_i^t + V_i^{t+1} \qquad (8)$$

and the velocity of ith solution is to be updated as follows:

$$V_i^{t+1} = V_i^t + \left(S_j^t - S_i^c\right) \times F_i \qquad (9)$$

The above equation V_i^{t+1} denotes the velocity update of ith solution; and V_i^t represents the velocity of ith solution; and S_j^t denotes the current best solution; S_j^c denotes the current i^{th} solution; F_i denotes the frequency of each solution.

Simultaneously F_i is also calculated using Eq. (10), where β indicates the random vector of range [0,1]. Initially, each bat is randomly assigned a frequency which is drawn uniformly from f^{min} and f^{max}.

$$F_i = f^{min} + \left(f^{max} - f^{min}\right) \times \beta \qquad (10)$$

Step 6: After reaches the first iteration of BAT algorithm; N numbers of fitness values are obtained.

Step 7: For every next iteration, N number of fitness is selected using step 4 and 6.
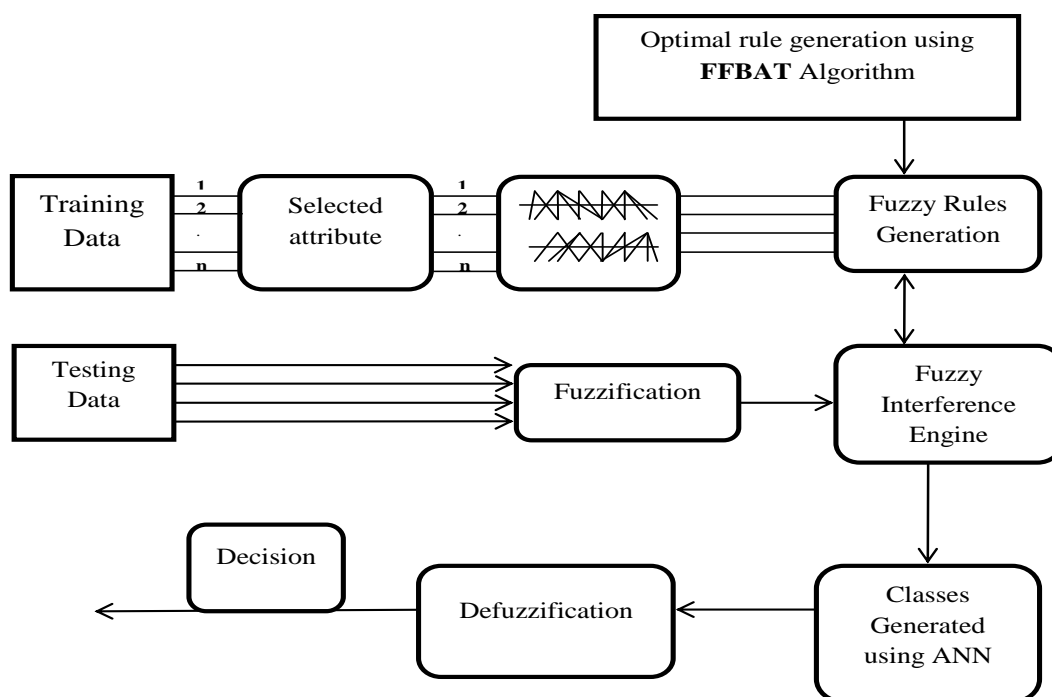

Figure .3   FFBAT-ANN Algorithm

Step 8: Termination criteria: While the maximum iteration reaches, the FF-BAT algorithm stops. Finally, it selects the best optimal fitness

value and it is given to the classifier as best rule. Collectively, the proposed FF-BAT algorithm can solve the explorations. Even though, FF is an

intelligent algorithm, it does promote sufficient searching ability in some cases. Under such circumstances, BAT algorithm provides the required solutions and vice-versa.

Stage 2: Classification Using Artificial Neural Networks Classifier

The block diagram representation of the proposed system is illustrated in Fig. 3.

Design of Fuzzy System: The fuzzy system has been classified as three significant stages namely a) Fuzzification b) fuzzy inference engine c) Defuzzification. Fuzzification is the initial stage in which the crisp input is converted into a linguistic variable with the relationship of the membership function collected from the fuzzy knowledge base. In Fuzzy inference engine the fuzzified input is further modified into the fuzzy output by means of If-Then type fuzzy rules. Defuzzification is the process of transforming the fuzzy output obtained from the inference engine to the original crisp with the help of similar membership function that is employed by the fuzzifier stage. In this paper, we considered the triangular membership function for fuzzification of the crisp input. The reduced feature (N) of the testing data (D^{TE}) or the crisp input is forwarded to the fuzzification stage of the fuzzy system, where the fuzzy value is obtained in accordance with the membership function. Then, the fuzzified input is given to the fuzzy inference engine stage, where it is examined with the fuzzy rules and gets converted to the fuzzy score using the average weighted method. Furthermore, the classification decision of diabetes disease from the dataset is produced from the obtained fuzzy score, using Fuzzy Neural Networks classifier.

Artificial Neural Networks: Artificial Neural Network (ANN) is the mathematical representation of human neural architecture by reflecting human "knowledge" and "generalized aptitudes". Since the medical diagnosis with the complex variable relationship of non-linear systems, ANNs are successfully used. The general structure of neural network is designed by a sequence of neurons or nodes that are arranged in layers. The connection between the neurons in different layer is weighted and so it is called a weighted link. The strength of the link which connects the ith neuron in a layer and the jth neuron in the next layer is represented as weight w_{ij}. These weights can be tuned to obtain the appropriate output. Fig. 4 explains about the general structure of Artificial Neural Networks.

4. Results and Discussion

The comprehensive analysis of the proposed FFBAT-ANN algorithm for diabetes disease classification is described in this section. The experimental analysis in our proposed system were implemented in MATLAB 8.4 and the simulations were carried out with i5 processor of 4 GB RAM.

4.1 Dataset Description

The raw database from September 2014 to January 2015 for the analysis of diabetes classification has been collected from the Sree Diabetic Care Center, Kurnool in Andhra Pradesh, India including the information of recorded patient's age, sex, BMI, BP, cholesterol, LDL, HDL, Triglycerides, FBS, Smoking, alcohol, family history of diabetics, heart disease, and status of diabetes (class label) over the age limit of 18-77. The details of features to be considered from the dataset is provided in the review papers.

4.2 Evaluation Metrics

Metrics such as sensitivity, specificity and accuracy of the proposed method is discussed and estimated from the following link: https://en.wikipedia.org/wiki/Sensitivity_and_specificity

Since the diabetes classification is carried out with respect to the metrics, it has been divided into two major classes of positive tuples (tuples with Diabetes = Yes) and negative tuples (tuples with Diabetes = No). True Positives (TP) and False Negatives (FN) are the positive tuples with which they are correctly and wrongly marked by the classifier respectively. while True Negatives (TN) and False Positives (FP) are the negative tuples that were correctly and wrongly marked by the classifier respectively. In addition, the False Positives (FP) and False Negatives (FN) correspond to the false prediction of diabetes by the classifier. We can say that FP refers to the tuples of class diabetes=No at which the classifier predicted diabetes=Yes. In a similar way, FN gives information about the tuples with class of diabetes='Yes' while the classifier predicted diabetes='No'. The performance of classifier can be analysed with the aforementioned terms. The statistical measure of sensitivity and specificity shows the binary classification test performance. The proportion of correctly labelled positives and negatives are calculated by corresponding sensitivity and specificity [43-50].

4.3 Performance Evaluation

In this section, three well known metrics such as accuracy, sensitivity and specificity are analysed with the performance of the proposed FFBAT-ANN. We assumed that the input dataset with the population size of 10, 15, 20 samples from which 9, 14, 19 samples are indicated as training dataset and the testing dataset indicates the remaining 1 sample of each population respectively. Fig. 5 shows the accuracy performance, Fig. 6 shows the sensitivity performance and Fig.7 shows the specificity performance of diabetes classification for an existing (FFBAT-Rule Based Fuzzy Classifier(FFBAT-RBFL) and proposed approaches respectively (FFBAT-Artificial Neural Networks Classifier) with respect to aforementioned population sizes. As such the three metric measures such as Accuracy, Sensitivity, and Specificity of the proposed approach outperforms the existing approach is demonstrated in the following figures.

As shown in Fig. 4, 5, 6, the metrics of the proposed system namely accuracy, sensitivity and specificity with respect to varying population range and its comparison with existing system. That is to say, the accuracy value of proposed system corresponding with population size 10, 15 is 3.5% improvement than the existing system. As in the case of Sensitivity 2.66% get increased with respect to the population size 20 and in specificity metric 2.9% and 3.44% better than previous system for the corresponding 10, 15, 20 population size. Finally all the metric values are increasing in proportional to the population size.

Table 1 (A) shows the accuracy values of FF, BAT and proposed FF-BAT methods for Neural Network as well as the standard RBFL. As shown the maximum percentage increase of accuracy values in the 4th, 5th and 6th position are 6.25%, 7.6%, 5.2% when comparing simple BAT with FF method and also 14.2%, 6.8%, 4% improvement in proposed FF-BAT with simple BAT method. Moreover, Table 1 (B) illustrates the performance of proposed system with respect to the sensitivity metric on diabetes. When comparing the proposed with the standard, the sensitivity is high enough. However, in Table 1 (C) indicate the relative rise in specificity values of FF, BAT and our proposed FF-BAT of neural network than existing RBFL. Behind that from the neural network we compared the values of accuracy, sensitivity and specificity for simple FF, BAT and proposed FF-BAT methods with RBFL logic.
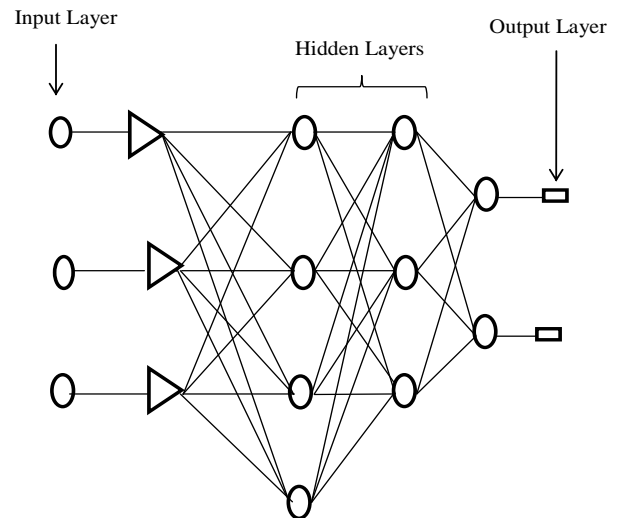

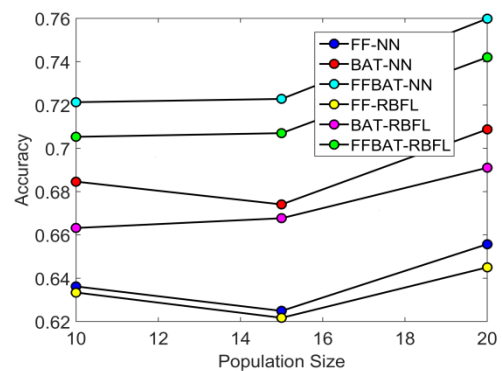Figure .4   General Structure of a Neural Network with different Layer of nodes

Figure 5. Accuracy of FFBAT-RBFL and proposed FFBAT-NN approach for varying population size on diabetes.
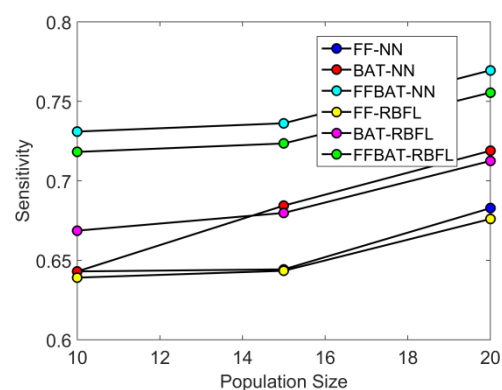

Figure 6. Sensitivity of FFBAT-RBFL and proposed FFBAT-NN approach for varying population size on diabetes.
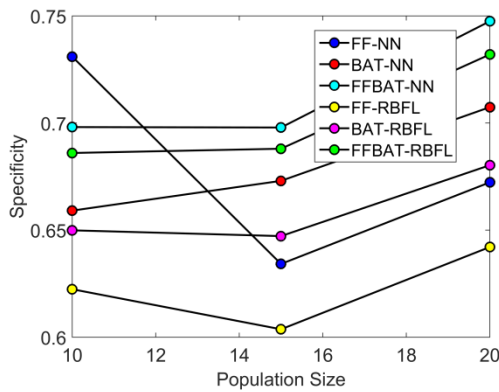
Figure 7. Specificity of FFBAT-RBFL and proposed FFBAT-NN approach for varying population size on diabetes.

Table 1. Performance comparison between proposed FFBAT-ANN and FFBAT-RBFL algorithm with ANN and RBFL trained by FF and bat

(a) Accuracy

NN				RBFL		
K-fold	FF	BAT	FF-BAT	FF	BAT	FF-BAT
1	0.705	0.63	0.665	0.567	0.6	0.65
2	0.75	0.678	0.705	0.63	0.66	0.703
3	0.716	0.64	0.681	0.63	0.645	0.67
4	0.72	0.646	0.68	0.58	0.64	0.67
5	0.795	0.712	0.743	0.654	0.7	0.73
6	0.72	0.65	0.7	0.623	0.64	0.68
7	0.744	0.71	0.722	0.675	0.68	0.705
8	0.7	0.63	0.685	0.623	0.65	0.67
9	0.71	0.65	0.695	0.63	0.64	0.68
10	0.75	0.645	0.705	0.612	0.644	0.702

(b) Sensitivity

NN				RBFL		
K-fold	FF	BAT	FF-BAT	FF	BAT	FF-BAT
1	0.62	0.657	0.69	0.57	0.58	0.67
2	0.68	0.72	0.74	0.64	0.67	0.73
3	0.658	0.679	0.715	0.63	0.67	0.69
4	0.64	0.68	0.72	0.646	0.687	0.71
5	0.678	0.73	0.78	0.65	0.67	0.76
6	0.653	0.687	0.715	0.633	0.657	0.7
7	0.568	0.647	0.72	0.654	0.689	0.713
8	0.615	0.657	0.71	0.577	0.644	0.68
9	0.59	0.674	0.693	0.654	0.67	0.68
10	0.66	0.715	0.73	0.68	0.695	0.72

(c) Specificity

NN				RBFL		
K-fold	FF	BAT	FF-BAT	FF	BAT	FF-BAT
1	0.65	0.65	0.705	0.63	0.67	0.69
2	0.64	0.64	0.75	0.67	0.64	0.74
3	0.634	0.634	0.716	0.65	0.68	0.705
4	0.63	0.63	0.72	0.643	0.675	0.719
5	0.68	0.68	0.795	0.687	0.702	0.775
6	0.61	0.61	0.72	0.68	0.7	0.715
7	0.622	0.622	0.744	0.65	0.67	0.729
8	0.62	0.62	0.7	0.57	0.64	0.68
9	0.664	0.664	0.71	0.566	0.6345	0.699
10	0.68	0.68	0.75	0.645	0.675	0.73

On considering the accuracy, FF attains 79% as best performance and 70% as worst performance, BAT attains 71% as best performance and 63% as worst performance and FF-BAT attains 74% as best performance and 66% as worst performance regarding the NN- based classification. On the contrary while using the proposed RBFL classification, best of 65% and 56% of worst performances in FF, 70% of best and 64% of worst performances in BAT and 73% of best and 67% of worst performance is occurred.

However in case of sensitivity, 68% of best and 56% of worst performances in FF, 73% of best and 64% of worst performances in BAT and 78% of best and 69% of worst performance are resulted using NN classification. To the next, FF attains 68% as best and 57% as worst performances, BAT attains 69% as best and 58% as worst performances and FF-BAT attains 76% as best and 67% as worst performances using RBFL classification.

Regarding the specificity, FF and BAT attains 68% as best and 61% as worst performances and FF- BAT attains 79% as best and 70% as worst performances using NN- based classification. Subsequently, best of 68% and 57% of worst performances in FF, 70% of best and 63% of worst performances in BAT and 77% of best and 69% of worst performance is occurred.

5. Conclusion

In this paper a new diabetes classifier called, FFBAT -ANN classification algorithm has been proposed by merging FFBAT algorithm with Artificial Neural Networks Classifier to classify Diabetes dataset. The hybrid FFBAT-ANN algorithm involves three stages namely dimension reduction with reduced attributes, rule generation using fuzzification process and then classification. Here, LPP algorithm is used to reduce the dimension of dataset, then fuzzy rules are generated and then the classification is carried out by novel FFBAT-ANN algorithm. Simulation results have proved that the proposed classification framework outperforms the existing method by achieving better accuracy, specificity and sensitivity metric measures.

Further, rule strength and overall performance of our proposed system can be improved by applying various statistical measures and meta-heuristic optimization techniques. The proposed rule based classifiers for clinical diagnosis has good future scope for further increasing the accuracy by introducing more advanced optimization algorithms. The advent of optimization algorithms is increasing and they are developed in a globally acceptable

manner. As a result, the latest optimization algorithms can improve the training performance of the classifier and as a result, the classification accuracy can be substantially improved.

Acknowledgment

References

[1] J. Han and M. Kamber, "*Data Mining: Concepts and Techniques*", 3rd ed.Morgan kaufmann, 2011.

[2] K. Mali and S. Bhattacharya, "Soft Computing on Medical-Data (SCOM) for a Countrywide Medical System using Data Mining and Cloud Computing Features", *Global Journal of Computer Science and Technology Cloud and Distributed*, Vol.13, No.3, 2013.

[3] J. C. Prather, D.F.Lobach, L.K.Goodwin, J.W.Hales, M.L.Hage and W.E.Hammond, "Medical data mining: knowledge discovery in a clinical data warehouse," *in Proc. American Medical Informatics Association Annu. fall Symp.*, pp. 101-105, 1997.

[4] A. Antoniadis, S. Lambert-Lacroix and F.Leblanc, "Effective dimension reduction methods for tumor classification using gene expression data", *Bioinformatics*, Vol.19, No.5, pp.563–70, 2003.

[5] I. Guyon, J.Weston, S.Barnhill and V.Vapnik, "Gene selection for cancer classification using support vector machines. *Machine Learning*, Vol.46, pp.389–422, 2002.

[6] J-S. Yu, S. Ongarello, R. Fiedler, X-W. Chen, G. Toffolo and C.Cobelli, " Ovarian cancer identification based on dimensionality reduction for high-throughput mass spectrometry data", *Bioinformatics*,2005.

[7] A. Iyer, S. Jeyalatha and R.Sumbaly, "Diagnosis of Diabetes using Classification Mining Techniques", *International Journal of Data Mining & Knowledge Management Process* (IJDKP) Vol.5, No.1, 2015.

[8] N. Esfandiari, M. R. Babavalian, A-M E.Moghadam and V.K. Tabar "Knowledge discovery in medicine: Current issue and future trend", *Expert Systems with Applications,* pp.4434–4463, 2014.

[9] N.Wang and G. Kang, "A monitoring system for type 2 diabetes mellitus", *IEEE 14th International Conference on e-Health Networking, Applications and Services* (Healthcom), 2012.

[10] H. Lu, R. Setiono, and H. Liu, "Effective Data Mining Using Neural Networks", *IEEE Transactions on Knowledge And Data Engineering*, 1996.

[11] M. Seera and C.P. Lim "A hybrid intelligent system for medical data classification" *Expert Systems with Applications*, pp.2239–2249, 2014.

[12] F.Amato, A. López, E.M. Peña-Méndez, P. Vaňhara, A. Hampl and J. Havel, "Artificial neural networks in medical diagnosis" *J Appl Biomed*. Vol.11, pp. 47–58, 2013.

[13] O.Y. Atkov, S. G. Gorokhova, A. G. Sboev, E.V. Generozov, E.V. Muraseyeva, S.Y. Moroshkina and N.N. Cherniy "Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters" *Journal of Cardiology*, pp.190—194, 2012.

[14] P. Herron, "Machine learning for medical decision support: evaluating diagnostic performance of machine learning classification algorithms,"*INLS 110*, Data Mining, 2004.

[15] G. T.Reddy and N.Khare, "An Efficient System for Heart Disease Prediction using Hybrid OFBAT with Rule-Based Fuzzy Logic Model", *Journal of Circuits, Systems and Computers*, 2016.

[16] D.Delen, G.Walker and A.Kadam , "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial Intell. in Medicine*, Vol. 34, No. 2, pp. 113-127, 2005.

[17] N. Rathore, D.Tomar and S.Agarwal, "Predicting the survivability of breast cancer patients using ensemble approach," *in Proc. IEEE Int. Conf. Issues and Challenges in Intelligent Computing Techniques*, pp. 459-464, 2014.

[18] A. R. Webb, "*Statistical Pattern Recognition*" John Wiley & Sons, 2003.

[19] A. K. Jain, R.P. W. Duin and J.Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, No. 1, pp. 4-37, 2000.

[20] H. Jianchao, J.C. Rodriguez and M.Beheshti, "Diabetes data analysis and prediction model discovery using RapidMiner," *in Proc. 2nd IEEE Int. Conf. Future Generation Commun. and Networking*, Vol. 3, pp. 96-99, 2008.

[21] V. Nedic, S. Cvetanovic, D. Despotovic, M. Despotovic and S. Babic, "Data mining with

various optimization methods", *Expert Systems with Applications,*pp. 3993–3999, 2014.

[22] G. ThippaReddy and Neelukhare, "FFBAT-Optimized Rule Based Fuzzy Logic Classifier for Diabetes", *International Journal of Engineering Research in Africa*, Vol. 24, pp 137-151, 2016.

[23] I. Fister, I. Fister, X-S. Yang and J. Brest, "A comprehensive review of firefly algorithms" *Swarm and Evolutionary Computation*, Vol. 13, pp. 34-46, 2013.

[24] T. Hassanzadeh and M. R. Meybodi, "A New Hybrid Approach for Data Clustering using Firefly Algorithm and K-means" *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing* AISP, 2012.

[25] V. P. ArchanaLinnetHailey and N. Sudha, "An Optimization Approach of Firefly Algorithm to Record Deduplication", *International Journal of Engineering Research & Technology* (IJERT), Vol. 2 No. 9, 2013.

[26] E. Sadeghipour, N. Sahragard, M.-R. Sayebani and R. Mahdizadeh, " Breast Cancer Detection Based On A Hybrid Approach Of Firefly Algorithm And Intelligent Systems", *Indian Journal of Fundamental and Applied Life Sciences*, Vol.5 (S1), pp. 468-472. 2015.

[27] E. Sarac and S.A.Ozel, "Web page classification using firefly optimization", Innovations in Intelligent Systems and Applications (INISTA), *IEEE International Symposium*, 2013.

[28] X-S. Yang and X. He, "Firefly Algorithm: Recent Advances and Applications", *Int. J. Swarm Intelligence*, Vol. 1, No. 1, pp. 36–50. 2013.

[29] X-S. Yang , "Firefly Algorithm, Levy Flights and Global Optimization", *Research and Development in Intelligent Systems* XXVI, Springer-Verlag London Limited 2010.

[30] S.K. Pal , C.S Rai and A.P.Singh , "Comparative Study of Firefly Algorithm and Particle Swarm Optimization for Noisy NonLinear Optimization Problems", *I.J. Intelligent Systems and Applications*, 2012.

[31] T. Apostolopoulos and A. Vlachos, "Application of the Firefly Algorithm for Solving the Economic Emissions Load Dispatch Problem", *Hindawi Publishing Corporation International Journal of Combinatorics*, 2011.

[32] H. Banati and M.Bajaj, "Fire Fly Based Feature Selection Approach", *IJCSI International Journal of Computer Science Issues*, Vol. 8, No 2, 2011.

[33] X-S. Yang, "Bat Algorithm: Literature Review and Applications" *International Journal of Bio-Inspired Computation*, Vol. 5, No. 3, pp. 141–149, 2013.

[34] S. Mishra, K.Shaw and D. Mishra, "A New Meta-heuristic Bat Inspired Classification Approach for Microarray Data"*, Procedia Technology,* pp. 802-806, 2012.

[35] A. M. Taha, A.Mustapha, and S-D. Chen, "Naive Bayes-Guided Bat Algorithm for Feature Selection", *The ScientificWorld Journal*, 2013.

[36] X-S. Yang and A.H. Gandomi, "Bat algorithm: a novel approach for global engineering optimization", *Engineering Computations: International Journal for Computer-Aided Engineering and Software* Vol. 29 No. 5, 2012.

[37] G. Das, "Bat algorithm based Softcomputing Approach to Perceive Hairline Bone Fracture in Medical X-ray Images", *International Journal of Computer Science & Engineering Technology* (IJCSET), Vol. 4 No. 4, 2013.

[38] O. Hasancebi, T. Teke and O. Pekcan, "A bat-inspired algorithm for structural optimization" *Computers and Structures,* Vol. 128, pp.77–90, 2013.

[39] D. Rodrigues, L.A.M. Pereira, R.Y.M. Nakamura, K.A.P. Costa, X-S. Yang, A.N. Souza and J.P.Papa "A wrapper approach for feature selection based on Bat Algorithm and Optimum-Path Forest", *Expert Systems with Applications,* pp.2250–2258, 2014.

[40] A. H. Gandomi and X-S. Yang, "Chaotic bat algorithm" *Journal of Computational Science,* pp. 224–232, 2014.

[41] X. S.Yang, "Firefly algorithms for multimodal optimization," *Stochastic Algorithms: Foundations and Applications*, *Lecture Notes in Computer Sciences,* Vol.5792. pp. 169–178. 2009.

[42] H. B. Sandya, P. H. Kumar, H, Bhudiraja and S,K. Rao, "Fuzzy Rule Based Feature Extraction and Classification of Time Series Signal", *International Journal of Soft Computing and Engineering* (IJSCE), 2013.

[43] K. Srinivas, G. R.Rao and A. Govardhan, "Rough-Fuzzy Classifier: A System to Predict the Heart Disease by Blending Two Different Set Theories", *Research article - computer engineering and computer science*, 2014.

[44] C. L. Chowdhary and D.P. Acharjya."A Hybrid Scheme for Breast Cancer Detection using Intuitionistic Fuzzy Rough Set Technique", *International Journal of Healthcare*

Information Systems and Informatics. Vol. 11(2), pp. 38-61, 2016.

[45] C. L. Chowdhary, G. V. K. Sai and D. P. Acharjya, "Decrease in False Assumption for Detection using Digital Mammography", *In Springer Proceedings under AISC series, International Conference on Computational Intelligence in Data Mining (ICCIDM-2015)*, Vol. 2, pp. 325-333, 2016.

[46] K. Lakshmanna and N. Khare, "Constraint-Based Measures for DNA Sequence Mining using Group Search Optimization Algorithm." *International Journal of Intelligent Engineering & Systems,* Vol. 9, No.3, pp. 91-100, 2016.

[47] K. Lakshmanna, R.Kaluri, G.T.Reddy, G. Nagaraja and D.V Subramanian, "An Enhanced Algorithm For FPM From Biological Sequences" *International Journal Of Pharmacy & Technology,* Vo.8, No.2,pp.12776-12784, 2016.

[48] K. Lakshmanna and N. Khare " FDSMO: Frequent DNA Sequence Mining Using FBSB and Optimization", *International Journal of Intelligent Engineering & Systems,* Vol. 9, No.4, pp.157-166, 2016.

[49] R. Kaluri and C.P.Reddy, "A framework for sign gesture recognition using improved genetic algorithm and adaptive filter", *Cogent Engineering*, Vol.3, No.1, 2016.

[50] R. Kaluri and Ch. Pradeep Reddy, "Sign Gesture Recognition Using Modified Region Growing Algorithm and Adaptive Genetic Fuzzy Classifier", *International Journal of Intelligent Engineering & Systems*, Vol.9, No.4, pp.225-233, 2016.