[1]Mohammed Qasim Alazzawi

[2]Mustafa Ridha Al-Yasari

Internet of Things Based-Early Diagnosis System of Diabetes Using Optimized Neural Network

Abstract: - One of the most common diseases today is diabetes. This disease has an initial phase, if it is diagnosed on time, it helps to control and treat this disease to a great extent. But early diagnosis of this disease is very challenging due to the difficulty of repeated tests, especially in some geographical areas. Therefore, in smart health systems, a series of information is frequently taken from users by mobile phone, and by analyzing this information using artificial intelligence and machine learning algorithms and with the help of the Internet of Things, the doctor can diagnose diabetes in the first steps remotely. In this paper, we used a database that has 16 features related to early diabetes. Our proposed method for early diagnosis of diabetes is based on feature ranking along with MLP neural network optimized with whale optimization algorithm (WOA). also, the NCA algorithm is used to rank the features, and the WOA algorithm is used to optimize the parameters of the MLP neural network. Finally, based on the simulation results of the proposed method, we achieved 98.1% accuracy in the early diagnosis of diabetes.

Keywords: Medical data mining, diabetes disease, feature ranking, Whale optimization algorithm

I. INTRODUCTION

The proliferation of innovative technologies in various industrial domains, including healthcare, transportation, agriculture, and logistics, can be attributed to the progress made in information and communication technology (ICT) [1]. With the Internet of Things (IoT) playing a significant role in ICT, the future-focused industry has benefited from innovative tools like automation and decentralised intelligence [2]. IoT devices are used to connect the physical and cyber worlds for automatic data analysis and intelligent decision making, as their capacity to incorporate environmental intelligence grows [3].

IoT-based fog computing has been used in the healthcare industry to offer a number of time-sensitive features like medical recommendations, remote interventions, and intelligent clinical diagnosis [4]. The use of machine learning algorithms in conjunction with IoT technologies to diagnose diabetes early and manage it at home is the specific focus of this study.

In 2018, the World Health Organisation (WHO) reported that 422 million people worldwide suffer from diabetes, one of the chronic diseases with the fastest growth rates. Diabetes has a relatively long asymptomatic phase, so early diagnosis is always preferred for a successful clinical outcome. The lengthy asymptomatic phase of diabetes causes about 50% of patients to go undiagnosed. Only by accurately evaluating both common and uncommon symptoms—which can appear at various stages from the disease's onset to diagnosis—is an early diagnosis of diabetes achievable. Researchers accept data mining classification techniques for models of disease risk prediction. To forecast the likelihood of diabetes, a data collection that contains the data of new diabetic patients is required. In this study, we utilised a dataset of 520 cases gathered by direct surveys from patients of Sylhet Diabetes Hospital in Sylhet, Bangladesh.

In this work, we create a monitoring system for the management of chronic diseases of diabetic patients with the help of IOT and machine learning. This system has the ability to diagnose diabetes in the early stages of the disease and can send information to a mobile phone. This platform can provide information for patient-doctor interaction.

In this paper, our goal is to provide a monitoring system for diabetic patients on the IoT platform, which, the most important part of this system, which has a direct impact on the decision-making about the patient's condition is the machine learning part that runs on the server. Two important factors must be considered in the machine learning algorithm, which are very critical, namely the accuracy of data classification and the speed of classification. In this paper, we use Multi-layer perceptron (MLP) neural network for the machine learning algorithm, which is optimized

[1]*Assistant lecturer at Al-Mustaqbal University, mohammedqassim20@gmail.com , https://orcid.org/0009-0007-2348-6807

[2]Iraq, babel, must6021@gmail.com

by meta-heuristic whale algorithm in order to increase the accuracy and speed of this network. In this paper, to determine the weights and biases in the MLP neural network, meta-heuristic whale algorithms are optimally used. Also, along with this combined machine learning algorithm, a method based on K-nearest neighborhood (KNN) is also used to rank features. This algorithm increases the speed of the proposed method and also examines the impact of each feature on the proposed method, which can also increase the accuracy of the diagnosis.

This paper's remaining parts are organized as follows: Researchers' contributions to the field of diagnosis of diabetes are covered in Section 2; the proposed method is examined in Section 3; the results are discussed in Section 4; and the study is concluded in Section 5.

II. RELATED WORKS

In this section, various research works that have been presented to predict diabetes using data mining are reviewed. Diabetes detection was carried out by Zolfaghari [5] using a feedforward neural network and an ensemble of SVM. The majority vote approach was used to integrate the results from each separate classifier. With 88.04% success rate, the ensemble method outperformed the individual classifiers. Sneha and Gangil [6] used a variety of machine learning techniques, including logistic regression, SVM, and naïve Bayes (NB), to predict diabetes. SVM produced the greatest accuracy, at 77.37%. For the PIMA dataset, the authors also used feature selection. Low correlation characteristics were eliminated. In order to predict diabetes, Edeh et al. [7] tested four machine learning algorithms: Bayes, decision tree (DT), support vector machines (SVM), and random forest (RF) on two distinct datasets. SVM achieved the maximum accuracy of 83.1% in the PIMA experimental results. Chen et al. [8] used preprocessing to reorganise the PIMA data using the k-means method to exclude the data that had been incorrectly categorised (data reduction). After that, they used DT to classify the reduced data. The study's findings led to a 90.04% prediction accuracy for diabetes. To predict diabetes, Dadgar and Kaardaan [9] suggested a hybrid method. First, the UTA algorithm was used to pick features. Next, the two-layer neural network (NN) with the chosen features was fed, and the genetic algorithm (GA) was used to update the NN's weights. Thus, an accuracy of 87.46% for diabetes estimate was obtained. DT, RF, and NN models were used by Zou et al. [10] to predict diabetes. To further decrease dimensionality, they used minimal redundancy maximum relevance (mRMR) and principal component analysis (PCA). RF outperformed the rest in forecast success, with an accuracy rate of 77.21%.

A network featuring an input layer, completely linked layers, dropouts, and an output layer architecture was constructed by the authors in [11]. By feeding the PIMA dataset characteristics straight into this developed MLP, it was able to finish the application with an accuracy of 88.41%. Long short-term memory (LSTM) (LSTM-AR) was used in [12] to classify this data after artificial records were produced. With prior cross-validation, the LSTM-AR classification result, which was reported as 89%, outperformed both LSTM and the multi-layer perceptron (MLP). The authors [13] created a deep neural network that uses softmax to classify diabetes and stacked autoencoders to extract information. The correctness of the deep architecture that was created was 86.26%. The convolutional long short-term memory (Conv-LSTM) model was introduced by the authors in [14]. In order to compare the outcomes, they also tested using conventional LSTM and CNN. They used the grid search technique to optimise hyperparameters in deep models. The input layer was one dimensional (1D) for every model. Conv-LSTM fared better than other models for test data after training and test separation, with an accuracy of 91.38%. created a 1D CNN architecture in [15] to forecast diabetes. Outlier detection, however, repaired missing values. The imbalance in the data was then eliminated by preprocessing it using the synthetic minority oversampling method (SMOTE). After processing the data, they put it into the 1D CNN architecture, achieving an accuracy of 86.29%.

III. PROPOSED METHOD

A. In this section, we will examine the method proposed in this article for the early prediction of diabetes. In this research, a special database related to the early diagnosis of diabetes was used. The full description of the database is given in the next section. Our proposed method in this work is a combination of feature ranking to use effective features along with optimized multi-layer perceptron neural network with whale meta-heuristic algorithm (WOA). In this section, the steps of the proposed method are explained. The diagram of proposed method is shown in figure (1)

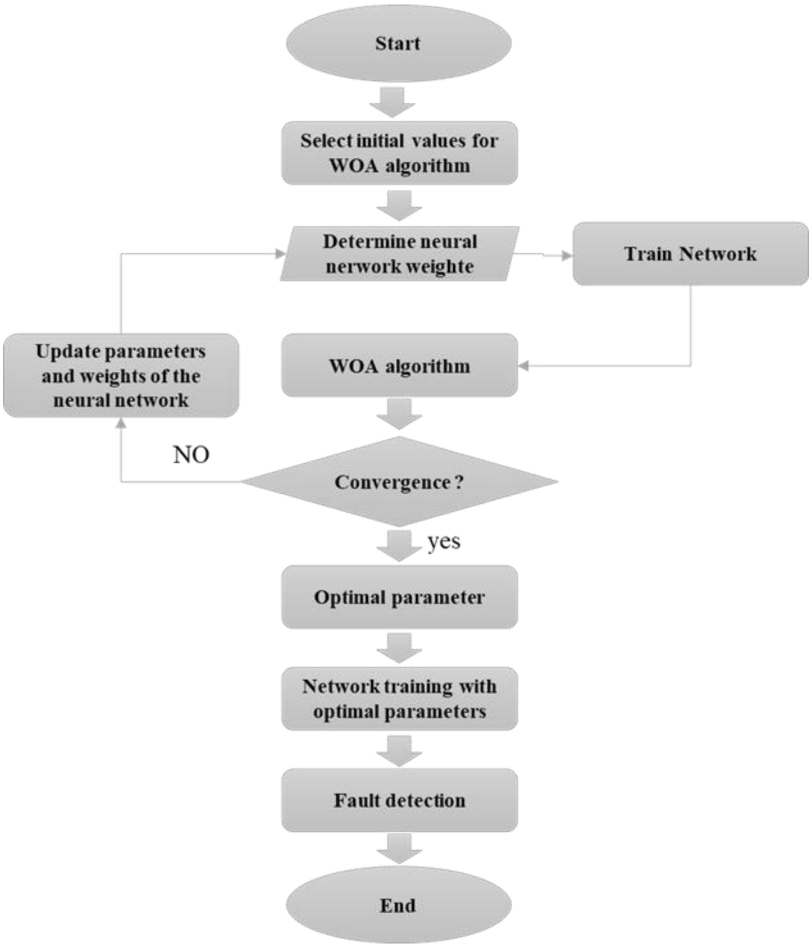

Start
Select initial values for WOA algorithm
Determine neural nerwork weighte
Train Network
Update parameters and weights of the neural network
WOA algorithm
NO
Convergence ?
yes
Optimal parameter
Network training with optimal parameters
Fault detection
End

Figure 1. Diagram of proposed method

B. Preprocessing

Usually, in the pre-processing stage, operations such as normalization, removal of missing data and labeling are performed on the data. In this research, the database used is completely text, so that a series of features such as obesity, sudden weight loss, fatigue, etc. have been answered with yes and no answers for different people. As a result, the most important pre-processing applied to this database was converting text to Numerical values. Also, the labels of the database were positive and negative, which belong to diabetic and healthy people. We considered zero for diabetic people and one for healthy people.

Feature ranking with NCA algorithm

After pre-processing the data and quantifying them, it is time to process the features. As explained above, the database used in this work contains a number of features that were asked from the respondents. These features have been answered with yes and no, which we have converted into numbers zero and one. This database has a total of 16 features, based on these features, diabetes is predicted in the audience. In fact, the neural network must find a relationship between these features with their corresponding labels, which are healthy and diseased states. Now, some of these characteristics have a greater effect and some less effect in identifying the disease. If we use all the features, it becomes more difficult for the neural network, because it is difficult to find a relationship between the 16 features with labels, some of which also have no effect on the response, and the recognition accuracy reduces. For this purpose, we will use the neighborhood component analysis (NCA) algorithm, which is an algorithm based on distance, to rank the features and select the effective features in finding the answer.

How the NCA algorithm works in this research is as follows:

1- Defining the weight of features: First, NCA determines a weight for each data sample and features. This weighting of each feature indicates how important features are that have significant similarities with other features.

2- Objective function definition: NCA tries to optimize an objective function called "chosen kernel function". This kernel function determines for each pair of data samples how similar they are in a feature space. This kernel function is defined as follows:

K_ij = exp(-x_i- 〖x_j〗 ^2)   (1)

where x_i and x_j are samples of two data points and x_i- x_j represents the Euclidean distance between them.

3- The objective function of NCA: The objective of NCA is defined as a criterion for determining the quality of the weight of features in separating samples from each other:

J(W) =∑_i▒∑_j▒ 〖p_ij K_ij 〗   (2)

Here, W represents the weight of features, p_ij represents the probability of two data samples i and j to belong to the same class, and K_ij is the chosen kernel function.

4- Optimization of the objective function: The objective of NCA is to optimize the function J(W) so that the weight of the features is determined in such a way that important differences and similarities are expressed in the data.

Classification with optimized MLP

MLP is a model for supervised learning that uses the back-propagation algorithm. It consists of two steps. In phase 1, the error is calculated based on the predicted outputs corresponding to the given input (forward phase) and in phase 2, the obtained error is re-propagated to the network and the network weights are adjusted to minimize the error (re-propagation phase). MLP structure is shown in figure 2.

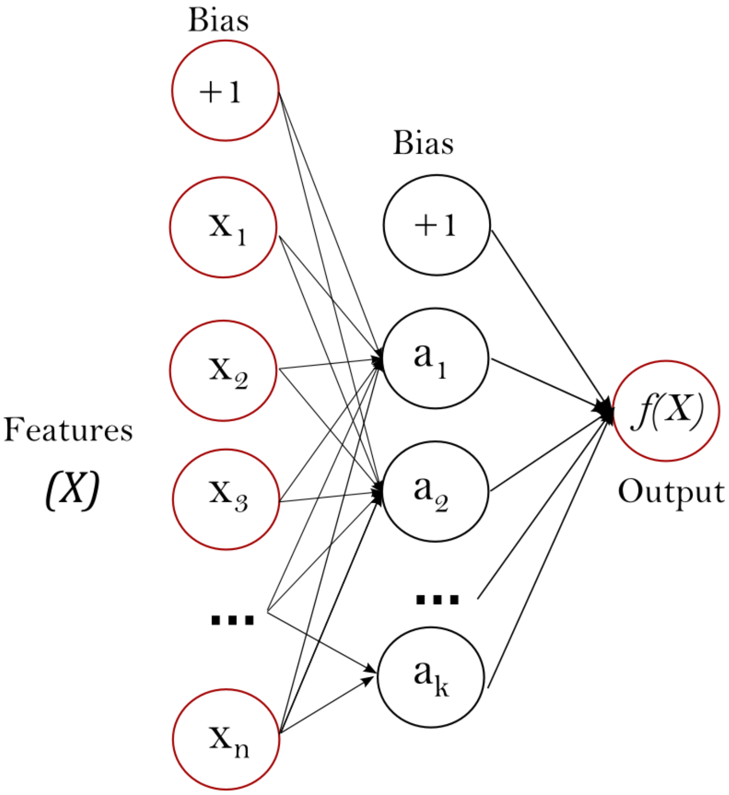


Figure 2: MLP structure.

To train the MLP neural network, the error value is calculated in line with the maximum slope of the error function and this value is sent to the previous layers to reduce the error value by resetting the weights of the neurons according to the delta rule:

W_ij^new=w_ij^old-η ∂E/(∂w_ij)   (3)

In the above equation, W_ij^new and w_ij^old are respectively the weight between neurons i and j before and after a certain repetition, η is the learning rate and E is the error function.

The error reduction continues until convergence is reached in the MLP neural network. Neural networks usually have several hidden layers. But in many researches, it has been shown that feedforward neural networks, even with a hidden layer, can obtain a suitable approximation of any type of nonlinear function.

In the learning process, weights and biases are determined in the network. The better the learning process in a neural network, the more accurately the weights and biases are determined and the accuracy of the output prediction increases. There are many learning processes such as Hebb process or reinforcement learning process.

But another way to optimally determine weights and biases is to use meta-heuristic algorithms. These algorithms find the best solution using iterative search. In this paper, we will use Whale Optimization Algorithm (WOA) to fine-tune the weights and biases.

Optimization with WOA algorithm

Three major processes comprise the whale optimization algorithm [16]: encircling the prey, attacking using a spiral bubble net, and randomly searching for prey. (1) Encircling the victim. The best candidate individual position in the current whale group is considered the target prey position by the WOA algorithm since the target prey location is unknown. Other whale group members then update their positions based on the position of the best candidate individual. Specifically:

█$(D=|C\cdot X^{*}(t)-X(t)|)$ (4)

$X(t+1)=X^{*}(t)-A\cdot D$ (5)

where A and C are the vectors representing the coefficients, X is the position vector of the current solution, t is the number of iterations, and X^{*} is the position vector of the ideal solution within the current whale population. The ways that A and C compute are:

█$(A=2ar-a)$ (6)

$C=2r$ (7)

where r is any vector between 0 and 1, and an is the number of iterations with an increment that reduces linearly from 2 to 0.

(2) Not attacking with a spiral bubble. The WOA algorithm replicates the humpback whale's spiral movement in order to determine the distance between the particular whale and the intended prey:

█$(D'=|X^{*}(t)-X(t)|)$ (8)

$X(t+1)=D'\cdot e^{bl}\cdot\cos(2\pi l)+X^{*}(t)$ (9)

where: b is constant coefficient of spiral shape, l is a random number in [−1, 1].

(3) Hunting prey at random. In order to enhance the algorithm's global search capability, individuals within the whale population choose prey at random by using each other's locations when A is larger than or less than −1.

$D=|C\cdot X_{\text{rand}}-X|$ (10)

$X(t+1)=X_{\text{rand}}-A\cdot D$ (11)

where: X_randis randomly selected position vector of the current whale group.

IV. RESULTS AND DISCUSSION

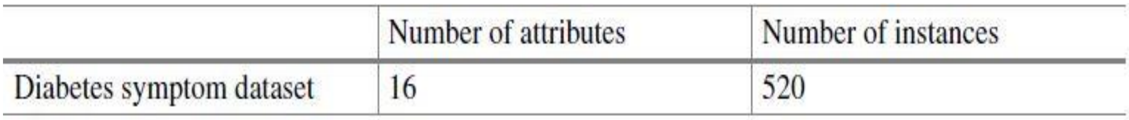
According to the proposed method for predicting early diabetes, in this section the simulation results of the proposed method will be presented. All the simulations in this research have been done using MATLAB 2021 software. In the rest of this section, the database used in the simulations and evaluation criteria, as well as the characteristics of the desired neural network, will be explained. Finally, the obtained results will be compared with other works.:

A. Database

This database includes reports of symptoms related to diabetes of 520 people [17]. It contains data about people including symptoms that may cause diabetes. This database was created from a direct questionnaire for people who have recently become diabetic or are still non-diabetic but have few or more symptoms. Data were collected from patients using direct questionnaires from Sylhet Diabetes Hospital in Sylhet, Bangladesh.

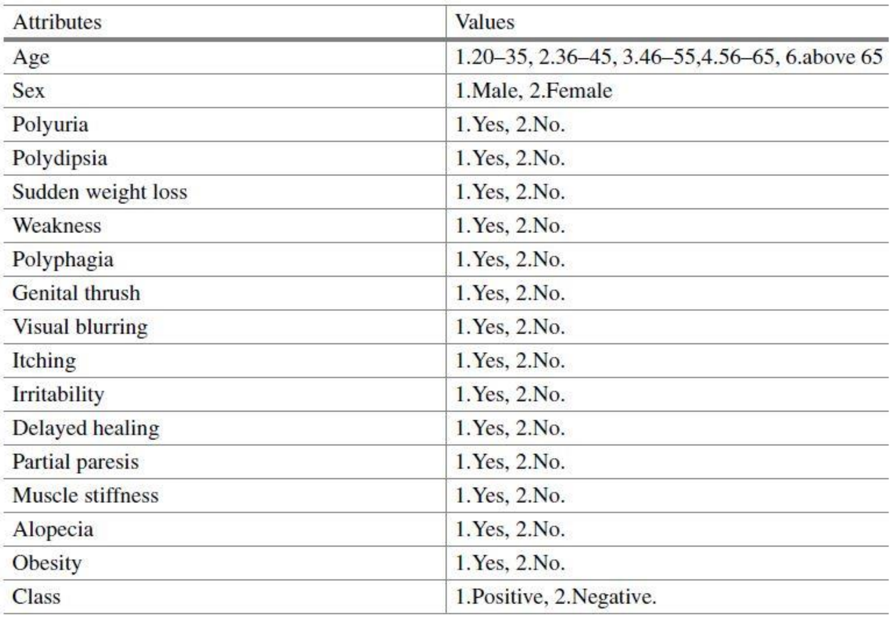
Data preprocessing is done by handling missing values following the technique of ignoring tuples with incomplete values. After preprocessing, 500 samples are left in total. Among them, 314 numbers are positive and 186 numbers are negative. A detailed description of the dataset and features are shown in Tables 5-1 and 5-2. Two class variables are used to find out whether a patient is at risk for diabetes (positive) or not (negative).

Table 1: Database description



	Number of attributes	Number of instances
Diabetes symptom dataset	16	520

Table 2: Feature description



Attributes	Values
Age	1.20–35, 2.36–45, 3.46–55,4.56–65, 6.above 65
Sex	1.Male, 2.Female
Polyuria	1.Yes, 2.No.
Polydipsia	1.Yes, 2.No.
Sudden weight loss	1.Yes, 2.No.
Weakness	1.Yes, 2.No.
Polyphagia	1.Yes, 2.No.
Genital thrush	1.Yes, 2.No.
Visual blurring	1.Yes, 2.No.
Itching	1.Yes, 2.No.
Irritability	1.Yes, 2.No.
Delayed healing	1.Yes, 2.No.
Partial paresis	1.Yes, 2.No.
Muscle stiffness	1.Yes, 2.No.
Alopecia	1.Yes, 2.No.
Obesity	1.Yes, 2.No.
Class	1.Positive, 2.Negative.

As it is clear in the table (2), most of the features are in the form of yes and no, in the pre-processing part, we have converted all the data into the format of zero and one numbers. Also, the labels are zero and one.

B. Evaluation Criteria

In this research, the criteria of accuracy, precision, recall and F-score have been used to evaluate the efficiency of the proposed method, which are calculated with the following equations.

$$Accuracy(acc) = \frac{TP+TN}{TP+TN+FP+FN} \qquad (12)$$

$$precision = \frac{TP}{TP+FP} \qquad (13)$$

$$recall = \frac{TP}{TP+FP} \qquad (14)$$

$$F1score = 2 * \frac{recall * precision}{recall + precision} \qquad (15)$$

Simulation setting

According to our goal in this work, which is the prediction of early diabetes and it is a two-class classification, we continue to examine the simulation parameters.

First, 520 vectors with 16 columns of features are considered and then we give 70% of these vectors to the neural network for the learning process. Of course, in feature ranking mode, the desired features are selected before dividing the data into train and test, which is explained in the next section. Then the learning process is completed using the train data and the weights and biases are adjusted using the Whale optimization algorithm.

The number of neurons in the hidden layer of the network is 10 and the maximum iteration is 100. The parameters of Whale algorithm are: Vmax=0.5 and Vmin=-0.5.

Results

In this section, numerical simulation results of the proposed method are presented. We obtained the results for two different cases. The first case without feature ranking and with the presence of all features and in the second case with feature ranking and optimal feature selection. Also, all the results obtained in this section have been obtained from test data. And the train data is only used for the learning process of the proposed network.

1. Results without Ranking of Features

In this case, the proposed neural network is trained with data that has all the features and no feature has been removed. And after network training, it has been evaluated by test data, which is shown in figure (3) of the confusion matrix for this case. As can be seen, it is a two-class classification where the first class (0) belongs to diabetic people and the second class (1) belongs to healthy people. The overall detection accuracy in this case is 88.5%. Of course, this accuracy is for one execution, and since the accuracy value is different in each simulation execution, for a correct evaluation, an average should be taken from a large number of executions.

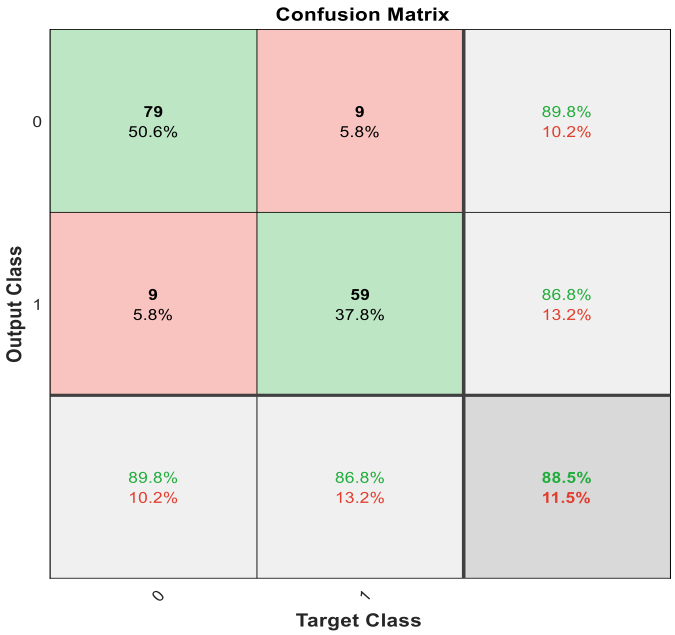


Figure 3: Confusion matrix for the mode without feature ranking

2. Results with Ranking of Features
In this section, the complete results of the proposed method, which includes the ranking of features, are presented. As stated, we used the NCA algorithm to rank the features, which works based on the K-nearest neighbor algorithm. Figure (4) shows the ranking chart of the features of this database. We omitted the features number 2, 7, 8, 10 and 14 because they are less important and we trained the network with other features. Figures (5) and (6) show the confusion matrix and ROC diagram for this mode, respectively. As it is known, the detection accuracy for a run in this mode is equal to 95.5%. Finally, in figure (7) the convergence diagram of Whale optimization algorithm is shown. It is used to find the optimal values of MLP neural network parameters. As shown, the WOA algorithm has converged after approximately 40 iterations.

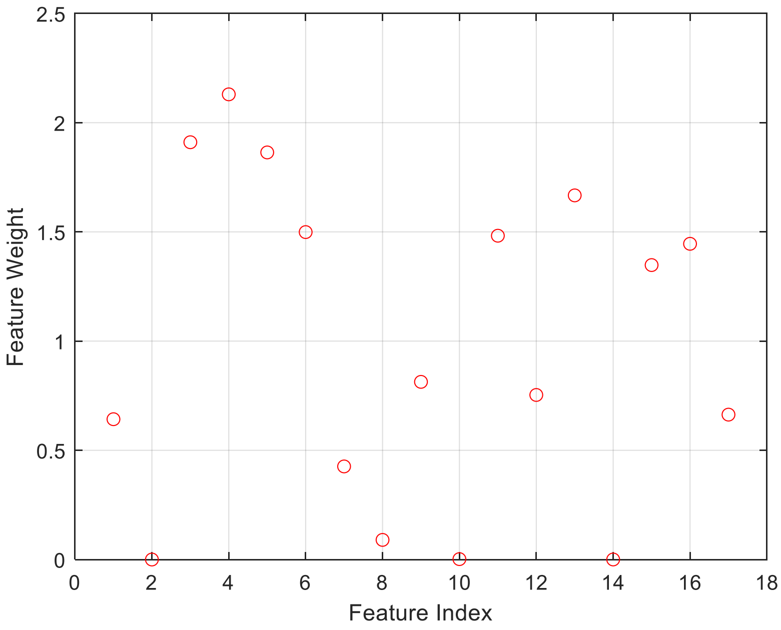


Figure 4: Features Ranking Chart

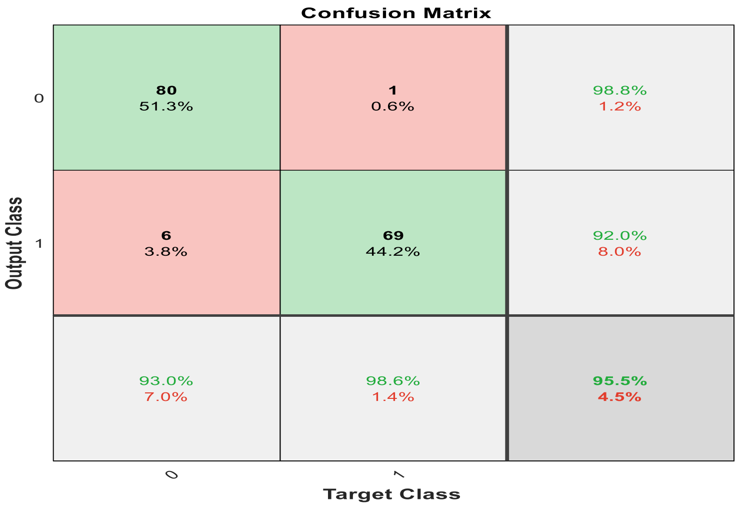


Figure 5: Confusion matrix for feature ranking case

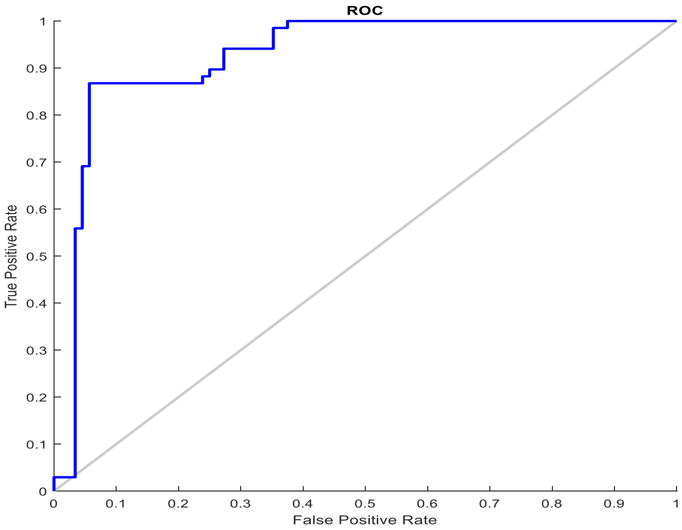


Figure 6: ROC plot for feature ranking mode

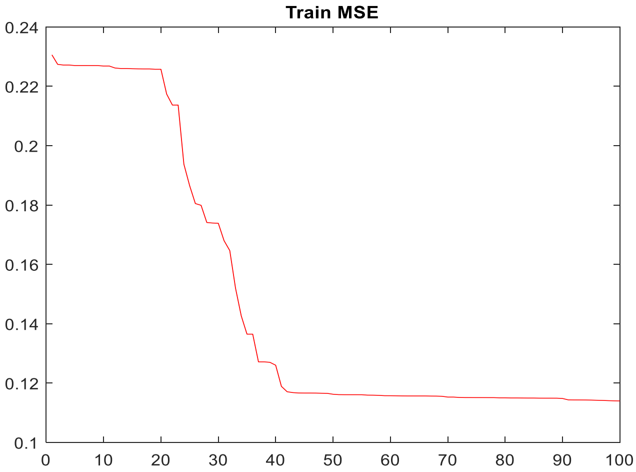


Figure 7: WOA algorithm convergence diagram

3. Comparison of Results

Finally, in Table (3) a comparison between the proposed method and some other methods is presented in terms of accuracy. For the validity of the results, the values presented in this table are the results of 30 averaging of different run of simulation. As can be seen, the proposed method is superior to other methods in terms of the accuracy of early diagnosis of diabetes.

Table 3: Classification results of different methods on Sylhet diabetes database.

Method	**Accuracy**
Naïve Bayes	87.4
Logistic Function	92.4
Decision Tree (J48)	95.6
Random Forest	97.4
Proposed method	**98.1**

V. CONCLUSION

Early evaluation of symptoms can lead to early diagnosis of diabetes. Since many people in the world do not have easy access to doctors and laboratories, it should be possible to predict diabetes as soon as possible through a simple and convenient system with a very low cost. The system should be designed for specific target users so that it is easily accessible to the public. In this paper, we tried to diagnose diabetic patients by means of optimized MLP neural network and feature ranking method, as well as examine the advantages and disadvantages of some of these methods. Our proposed method in this work is a combination of feature ranking to use effective features along with recurrent neural network optimized with meta-heuristic Whale optimization algorithm (WOA). By evaluating the results, we find that the proposed model has increased the efficiency of early diagnosis of diabetes by considering advantages such as high accuracy.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, \The internet of things: A survey," Computer networks, vol. 54, no. 15, pp. 2787{2805, 2010.

[2] G. Manogaran, R. Varatharajan, D. Lopez, P. M. Kumar, R. Sundarasekar, and C. Thota, \A new architecture of internet of things and big data ecosystem for secured smart healthcare monitoring and alerting system," Future Generation Computer Systems, vol. 82, pp. 375{ 387, 2018.

[3] M. Bhatia and S. K. Sood, \Game theoretic decision making in iot-assisted activity monitoring of defence personnel," Multimedia Tools and Applications, vol. 76, no. 21, pp. 21 911{21 935, 2017.

[4] "The Internet of Things: How the Next Evolution of the Internet (April 2011). Cisco. Retrieved 4 September 2015.

[5] Zolfaghari, R. Diagnosis of diabetes in female population of pima indian heritage with ensemble of bp neural network and svm. Int. J. Comput. Eng. Manag/ 2012, 15, 2230–7893.

[6] Sneha, N.; Gangil, T. Analysis of diabetes mellitus for early prediction using optimal features selection. J. Big Data 2019, 6, 13.

[7] Edeh, M.O.; Khalaf, O.I.; Tavera, C.A.; Tayeb, S.; Ghouali, S.; Abdulsahib, G.M.; Richard-Nnabu, N.E.; Louni, A. A Classification Algorithm-Based Hybrid Diabetes Prediction Model. Front. Public Health 2022, 10, 829519.

[8] Chen, W.; Chen, S.; Zhang, H.; Wu, T. A hybrid prediction model for type 2 diabetes using K-means and decision tree. In Proceedings of the 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 24–26 November 2017; pp. 386–390.

[9] Dadgar, S.M.H.; Kaardaan, M. A Hybrid Method of Feature Selection and Neural Network with Genetic Algorithm to Predict Diabetes. Int. J. Mechatron. Electr. Comput. Technol. (IJMEC) 2017, 7, 3397–3404.

[10] Wang Y, Zhang L, Niu M, Li R, Tu R, Liu X, Hou J, Mao Z, Wang Z, Wang C. Genetic risk score increased discriminant efficiency of predictive models for type 2 diabetes mellitus using machine learning: cohort study. Frontiers in public health. 2021 Feb 17;9:606711.

[11] Ashiquzzaman A, Kawsar Tushar A, Rashedul Islam M, Kim JM. Reduction of Overfitting in Diabetes Prediction Using Deep Learning Neural Network. arXiv e-prints. 2017 Jul:arXiv-1707.

[12] Madan P, Singh V, Chaudhari V, Albagory Y, Dumka A, Singh R, Gehlot A, Rashid M, Alshamrani SS, AlGhamdi AS. An optimization-based diabetes prediction model using CNN and Bi-directional LSTM in real-time environment. Applied Sciences. 2022 Apr 14;12(8):3989.

[13] Khamparia A, Saini G, Pandey B, Tiwari S, Gupta D, Khanna A. KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network. Multimedia Tools and Applications. 2020 Dec;79:35425-40.

[14] Rahman, M.; Islam, D.; Mukti, R.J.; Saha, I. A deep learning approach based on convolutional LSTM for detecting diabetes. Comput. Biol. Chem. 2020, 88, 1073 Chowdary PB, Kumar RU. An effective approach for detecting diabetes using deep learning techniques based on convolutional LSTM networks. International Journal of Advanced Computer Science and Applications. 2021;12(4).

[15] Alex, S.A.; Nayahi, J.; Shine, H.; Gopirekha, V. Deep convolutional neural network for diabetes mellitus prediction. Neural Comput. Appl. 2022, 34, 1319–1327.

[16] Mirjalili, S.; Lewis, A. The Whale Optimization Algorithm. Adv. Eng. Softw. 2016, 95, 51–67.

[17] https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+dataset