An Adaptive Neural Network Based Deep Learning Model for Efficient Diabetes Prediction Using Genetic Algorithm

Dr. Priya Vij[1], Akanksha Mishra[2]

[1]*Associate Professor, Faculty of CS & IT, Kalinga University, Naya Raipur, Chhattisgarh, India, ku.priyavij@kalingauniversity.ac.in*
[2]*Assistant Professor, Faculty of CS & IT, Kalinga University, Naya Raipur, Chhattisgarh, India, ku.akankshamishra@kalingauniversity.ac.in*

The problem of diabetes prediction is well studied, there exists number of approaches in literature which handles the problem. The most approaches uses only limited features like BMI, blood sugar, and HbA1C in predicting the diabetes. However, the methods produces poor accuracy in diabetes prediction as they consider only limited features. To solve this issue, an efficient Adaptive Neural Network Based Deep Leaning Model for Diabetes Prediction (ANN-GA-DP) is presented in this article. The method is focused on applying deep learning technique in the problem of diabetes prediction. In this way, the medical data set has been preprocessed to get the maximum features and removing the noisy records from the set. Further, the features are converted into ensembles and trained with neural network. The network is constructed with number of intermediate layers to reduce the missing problem. The neurons of the layer are initialized with the number of populations produced by genetic algorithm from the ensemble given. Further, the neurons are designed to measure the Diabetes Carrying Weight (DCW) on each size of mutation given. The output layer generates number of DCW weight measures against two different classes. Based on the value of DCW weights, the method computes the Diabetes Prone Weight (DPW) to decide on the class of sample at the test phase.
The proposed ANN-GA-DP introduces higher performance in diabetes prediction.

Keywords: Deep Learning, Diabetes Prediction, Genetic Algorithm, ANN, ANN-GA-DP.

1. Introduction

The changing lifestyle of human society introduces various health issues to them. In this way, there are number of diseases identified every year which affect the health of human society. Some of the diseases are not harmful and some of them are not. However, predicting the possibility of specific symptom according to the presence of set of symptoms would help the person to manage the disease effectively. Diabetes is the most popular disease which is being identified in human society in a higher rate. The presence of diabetes is identified based on the features like Blood Sugar, HbA1C and Body Mass Index (BMI) and so on.

Diabetes prediction is the process of predicting the possibility of the disease according to the presence of different symptoms. There exist number of approaches which uses different features and methods towards the problem. For example, the blood sugar based approach considers the blood sugar recorded before and after the meal to perform prediction. Similarly, there are methods which consider body mass index as the key factor in predicting the disease. Number of approaches can be named to support the problem, but the performance of the method gets vary according to the feature consider and the method used.

The ensemble based approaches maintains number of ensembles of diabetes based on which the method predict the disease. Similarly, the support vector machine would be used in the problem which predicts the disease according to the support value measured for the input sample. On the other side, the decision tree has been used which classifies according to the number of features at sibling level to perform the prediction. The Bayesian classification can be used in this problem, which uses rule sets to perform classification. However, the methods suffer to achieve higher performance in diabetes prediction. With this consideration, an efficient Adaptive Neural Network based Diabetes prediction with genetic algorithm is presented in this article.

The performance of diabetes prediction is depending on the volume of data set used. The machine learning algorithms have the efficiency in improving the performance of diabetes prediction. By incorporating neural network with the model and adapting genetic algorithm in mutation and cross over, the performance of the model can be improved. The proposed model incorporates both NN and GA towards the problem of diabetes prediction. The neurons of the model are designed to measure the diabetes carrying weight (DCW) and diabetes prone weight (DPW) to perform diabetes prediction.

2. Related Works:

The problem of diabetes prediction is approached with different techniques and this section details set of methods around the problem.

A fundus image based automatic diabetic foot prediction model is presented in [1], which uses the radiomics features obtained from fundus image and perform feature selection and perform classification with support vector machine. A fused machine learning based diabetes prediction is presented in [2], which uses SVM and ANN for classification. A blockchain based diabetic cardio prediction is presented in [3], which perform clustering of data with rule based clustering and classification is performed using feature selection based adaptive neuro-fuzzy inference system (FS-ANFIS). A machine learning based blood glucose prediction is presented in [4], which uses biosensors in predicting blood glucose level. An mobile net CNN model is presented in [5], towards identifying the vulnerability against fast gradient sign methods (FGSM) adversarial attacks.

Machine learning based blood glucose prediction is presented in [6], which uses nutritional factors in predicting blood glucose. A guided neural network based readmission prediction is presented in [7], which uses ANN for classification. A deep learning model is presented in [8], to predict major adverse cardiovascular events (MACE).

Improved invasive weed bird swarm optimization algorithm (IWBSOA) enabled hybrid deep learning classifier is presented in [9], which uses SVM, RNN for classification. A generative adversarial network based model is presented in [10], towards diabetic macular edema. A temporal multi head attention based short term blood glucose prediction is presented in [11], which uses intensity correlation module to perform diabetes prediction. A nonlinear autoregressive neural network based blood glucose prediction is presented in [12], which uses continuous glucose monitoring (CGM) data to predict future blood glucose level. A prospective cohort study based foot ulcer risk prediction model is presented in [13]. Fuzzy inference system based covid-19 risk prediction model is presented in [14], which uses different influential symptoms of diabetes patients to perform prediction. A machine learning based prediction model is presented in [15], which uses ratio of insulin segregation and insulin utilization to perform prediction.

3. Adaptive Neural Network Based Diabetes Prediction Using Genetic Algorithm (ANN-DP-GA) Model:

The proposed adaptive neural network based diabetes prediction model with genetic algorithm (ANN-DP-GA) model read the medical data set and perform preprocessing to get the maximum features and removing the noisy records from the set. Further, the features are converted into ensembles and trained with neural network. The network is constructed with number of intermediate layers to reduce the missing problem. The neurons of the layer are initialized with the number of populations produced by genetic algorithm from the ensemble given. Further, the neurons are designed to measure the Diabetes Carrying Weight (DCW) on each size of mutation given. The output layer generates number of DCW weight measures against two different classes. Based on the value of DCW weights, the method computes the Diabetes Prone Weight (DPW) to decide on the class of sample at the test phase.

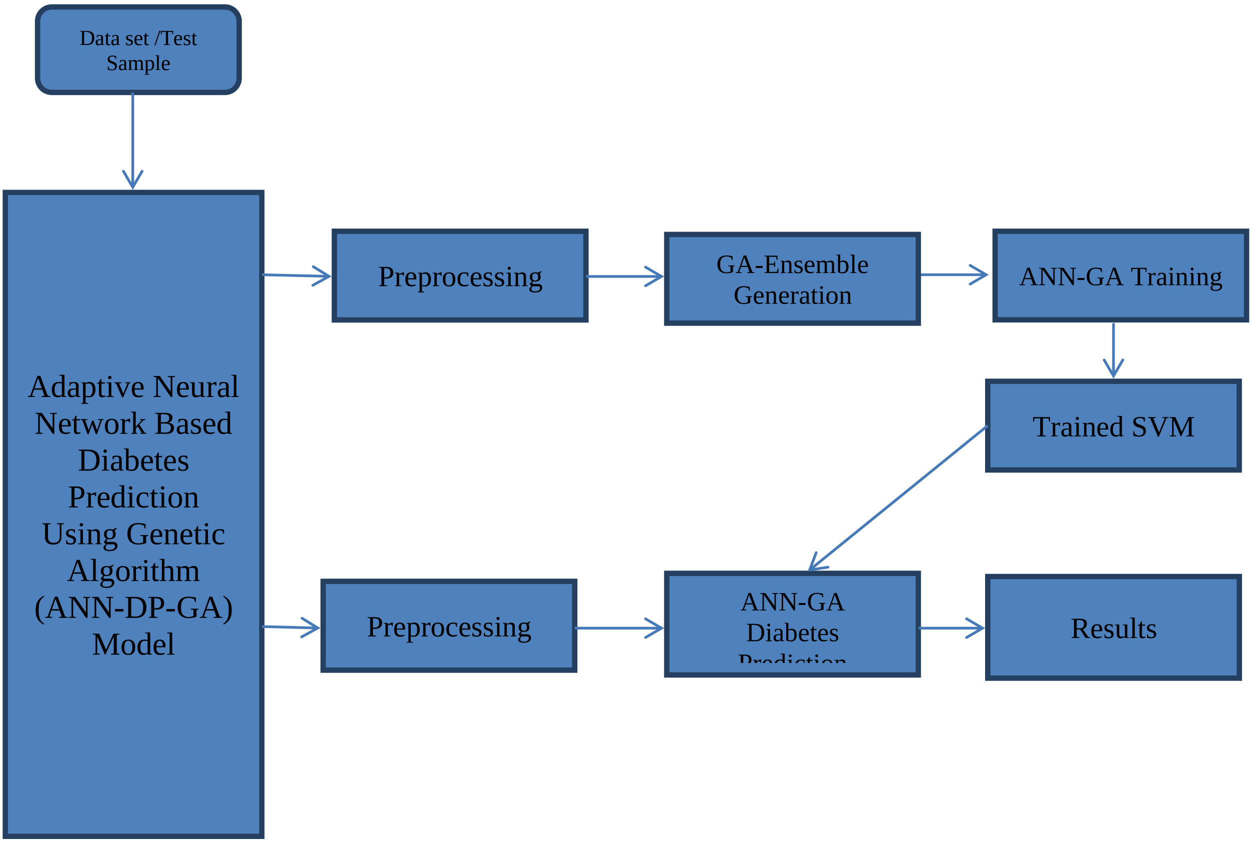

Figure 1: Architecture of proposed ANN-DP-GA Diabetes Prediction

The functional diagram of ANN-DP-GA scheme has been presented in Figure 1, which have number of stages and explained in this section.

Preprocessing:

The diabetes data set contains number of features and records. Each record have features like blood sugar, BMI, Age, Sex, HbA1C, Food Type, Physical Exercise, Smoking, and so on. This process initially identifies the set of features in the data set, which is being verified in each record of the data set. If the data point has covered all the features, then it has been included in the process, otherwise, it has been neglected from the set. The noise removed data set has been used to generate the ensembles.

Algorithm:

Given: Diabetes Data set Dds

Obtain: Pds

Start

Read Dds

Feature set Feas = $\mathrm{Feas} \cup (\sum_{i=1}^{\mathrm{size(Dds)}} \mathrm{Features(Dds(i))} \ni \mathrm{Feas})$

For each record r

If r∈ Features(∀Feas) then

Pds = Pds∪ r then

Else

Dds = Dds ∩ r

End

End

Stop

The preprocessing algorithm finds the features present in the data set and verifies each record for the possession of all features. Accordingly, the incomplete records are eliminated from the data set.

GA-Ensemble Generation:

The preprocessed data set has been used to generate the ensembles. From the preprocessed data set, the method read each record and identifies the set of features. With the set of feature and values, the method applies genetic algorithm to generate mutation and population. With at each mutation, a set of patterns are populated by computing Mutation Fitness Score (MFS). Populated mutations are generated as ensembles. Generated ensembles are added to the ensemble set. The minimum size of mutation is 4 and maximum size of mutation is size of feature set. Generated ensemble set has been used to perform training and testing.

Algorithm:

Given: preprocessed data set Pds

Obtain: ensemble set Es.

Start

Read Pds.

Initialize Ensemble set Es.

For each sample sa

For each size s

Mutation set Ms = Apply Genetic Algorithm to Generate Mutation.

Ms = $\sum_{i=1}^{s} \sum_{j=1}^{\mathrm{size(s)}} \mathrm{Mutate(sa, s(i))}$

For each mutation m

Compute Mutation Fitness Score MFS.

$$MFS = \frac{\mathrm{Sum}_{i=1}^{Size(pds)}\left(\frac{\mathrm{Count}_{j=1}^{size(m)}\left(Dist(m(j),pds(i)(j))<Dth\right)}{size(m)}\right)}{Size(Pds)}$$

If MFS > Th then

Add to ensemble set.

$Es = (\sum mutations \in Es) \cup m$

End

End

End

End

Stop

The GA-Ensemble generation algorithm generates number of mutations from the feature vector obtained from the data sample. The mutations are computed for mutation fitness score (MFS), based on the value of MFS, a set of mutations are populated and added to the ensemble set.

ANN-GA Training:

The proposed method trains the neural network with number of intermediate layers. First layer is the input layer and the last layer is the output layer. Intermediate layers are designed to have number of neurons where each would estimate Diabetes Carrying Weight (DCW) according to the size of feature considered for the layer. Similarly, the methods have k number of intermediate layer according to the number of features the tuples has finally. The DCW weight has been carried through number of layers and finally the output layer neurons produces k number of DCW values for a single sample given. Estimated value of DCW has been used to perform diabetes prediction.

ANN-GA Disease Prediction:

The proposed diabetes prediction scheme reads the test sample and extracts the features present in the sample. Further, the method generates the ensemble set from the feature vector by applying genetic algorithm and generated ensemble set has been passed through the network trained. The neurons at various layers compute the value of DCW and forward the same to the next layers. At last, the output layer neurons, returns set of DCW values according to the number of intermediate layer. Obtained values of DCW have been used to compute DPW (Disease Prone Weight) for various diseases. Based on the value of DPW, the method performs diabetic prediction.

Algorithm:

Given: Trained network nn, Test sample Ts.

Obtain: Disease Class Dc

Start

Read Nn, Ts.

Ensemble set Es = GA_Ensemble Generation (Ts)

For each ensemble E

Test with nn.

At each layer l

$$\text{Compute DCW} = \frac{\sum_{i=1}^{size(l)} Sum\left(\frac{\sum_{j=1}^{size(l(E))} Count(\sum l(i)[j]==E(j))}{Size(l(E))}\right)}{Size(l)}$$

Pass DCW to next layer neuron.

End

End

[DCWS] = obtain results from output layer.

$$DPW = \frac{\sum DCW}{size(DCWS)} \times \sum_{i}^{size(DCWS)} Count(DCWS(i) > Th)$$

Disease class DC = choose the disease with maximum DPW value.

Stop

The above algorithm computes the value of DCW and DPW to identify the class of sample given. Based on the value of DPW, a single class is identified as result.

4. Results and Discussion:

The proposed ANN-GA-DP model has been implemented using matlab. The performance of the model is evaluated using the diabetes data set which contains number of records with around 15 features. It covers generic features of diabetes, with lifestyle and physical features. Using the data set, the performance of the methods are analyzed and presented in this section.

Table 1: Evaluation Details

Feature	Value
Tool Used	Matlab
Data Set	Kaggle, PIMA, Lifestyle, Physical Features
Number of records	2000
Number of features	25

The evaluation details considered for the performance evaluation of the proposed model is presented in Table 1.

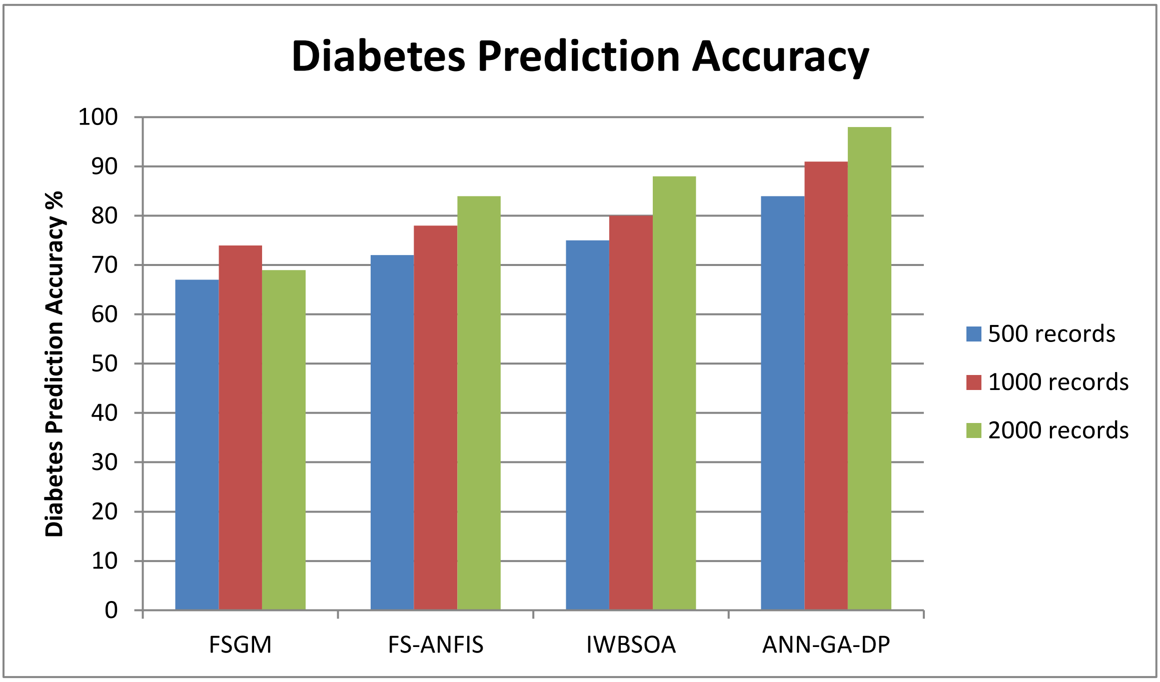

Figure 1: Diabetes Prediction Accuracy

The performance in diabetes prediction is measured and compared in Figure 2. The ANN-GA-DP model introduces higher prediction accuracy.

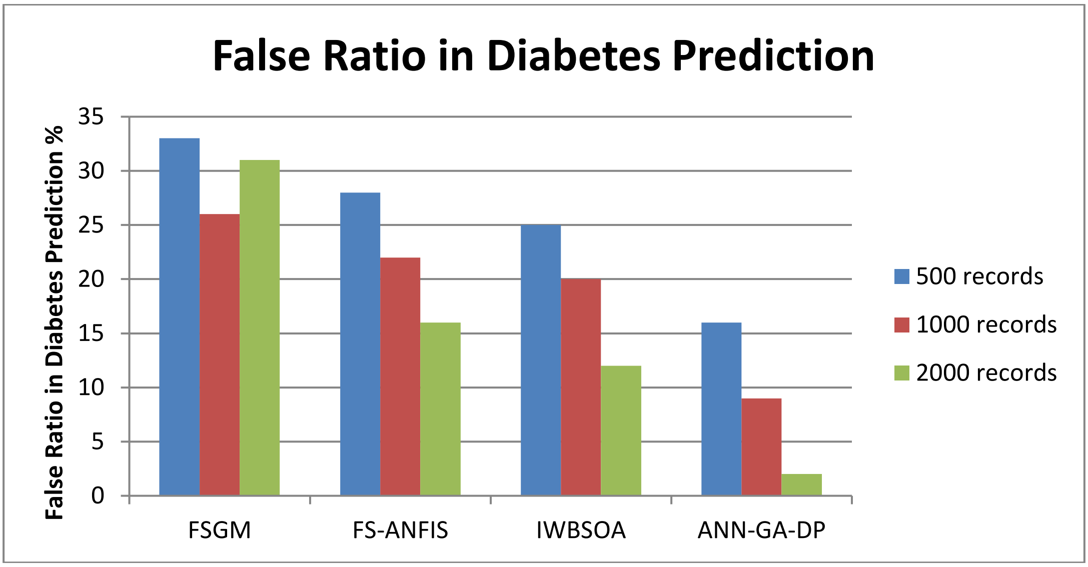

Figure 2: False ratio in Diabetes prediction

The ratio of false prediction is measured and presented in Figure 3, where ANN-GA-DP has introduced less false ratio than other methods.

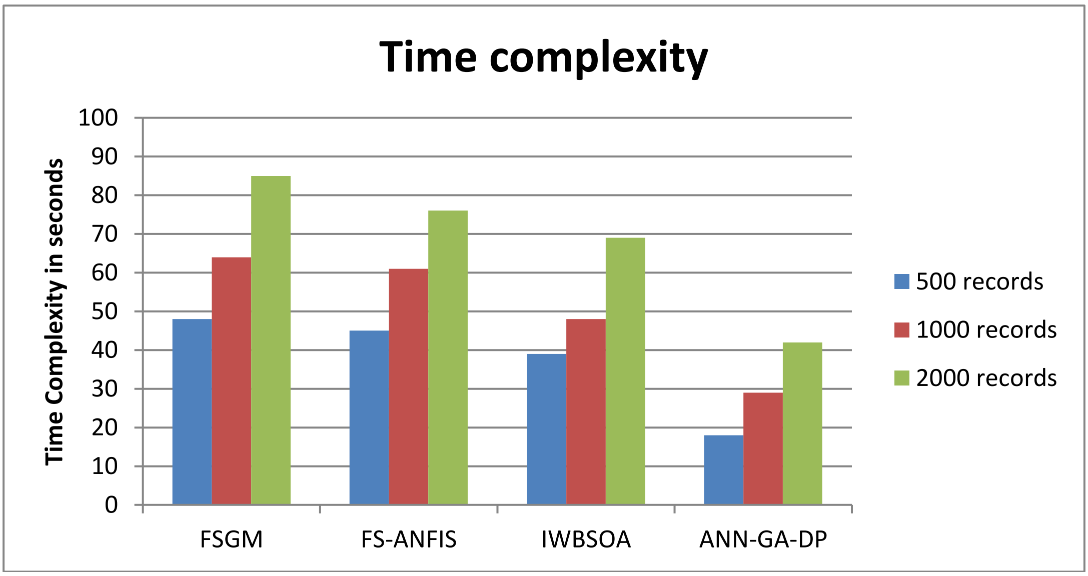

Figure 3: Time complexity

The time complexity introduced by the models are measured and presented in Figure 4. The ANN-GA-DP model introduces less time complexity compare to other methods.

5. Conclusion:

This paper presented a novel adaptive neural network based disease prediction with genetic algorithm. The method read the medical data set and perform preprocessing to get the maximum features and removing the noisy records from the set. Further, the features are converted into ensembles and trained with neural network. The neurons of the layer are initialized with the number of populations produced by genetic algorithm from the ensemble given. Further, the neurons are designed to measure the Diabetes Carrying Weight (DCW) on each size of mutation given. The output layer generates number of DCW weight measures against two different classes. Based on the value of DCW weights, the method computes the Diabetes Prone Weight (DPW) to decide on the class of sample at the test phase. The method introduces higher diabetes prediction accuracy with less time complexity.

References

1. X. Liang, E. N. Alshemmary, M. Ma, S. Liao, W. Zhou and Z. Lu, "Automatic Diabetic Foot Prediction Through Fundus Images by Radiomics Features," in IEEE Access, Volume. 9, pp. 92776-92787, 2021.
2. U. Ahmed et al., "Prediction of Diabetes Empowered With Fused Machine Learning," in IEEE Access, Volume. 10, pp. 8529-8538, 2022.
3. P. G. Shynu, V. G. Menon, R. L. Kumar, S. Kadry and Y. Nam, "Blockchain-Based Secure Healthcare Application for Diabetic-Cardio Disease Prediction in Fog Computing," in IEEE Access, Volume. 9, pp. 45706-45720, 2021.
4. M. Z. M. Shamim, S. Alotaibi, H. S. Hussein, M. Farrag and M. Shiblee, "Diagnostic Accuracy of Smartphone-Connected Electrophysiological Biosensors for Prediction of Blood Glucose Level in a Type-2 Diabetic Patient Using Machine Learning: A Pilot Study," IEEE (ESL), Volume. 14, Number 1, pp. 27-30, 2022.
5. O. Daanouni, B. Cherradi and A. Tmiri, "NSL-MHA-CNN: A Novel CNN Architecture for Robust Diabetic Retinopathy Prediction Against Adversarial Attacks," in IEEE Access, Volume. 10, pp. 103987-103999, 2022.
6. G. Annuzzi et al., "Impact of Nutritional Factors in Blood Glucose Prediction in Type 1 Diabetes Through Machine Learning," in IEEE Access, Volume. 11, pp. 17104-17115, 2023.
7. A. A. Ram, Z. Ali, V. Krishna, N. Nishika and A. Sharma, "A Guided Neural Network Approach to Predict Early Readmission of Diabetic Patients," in IEEE Access, Volume. 11, pp. 47527-47538, 2023.
8. E. Longato, G. P. Fadini, G. Sparacino, A. Avogaro, L. Tramontan and B. Di Camillo, "A Deep Learning Approach to Predict Diabetes' Cardiovascular Complications From Administrative Claims," IEEE (JB&HI), Volume. 25, Number 9, pp. 3608-3617, 2021.
9. Chetan Nimba Aher, "Improved invasive weed bird swarm optimization algorithm (IWBSOA) enabled hybrid deep learning classifier for diabetic prediction", Springer Link (JAI&HC), Volume 14, PP 3929-3945, 2023.
10. Shaopeng Liu, "Prediction of OCT images of short-term response to anti-VEGF treatment for diabetic macular edema using different generative adversarial networks", ELSEVIER (P&PT), Volume 41, Number 103272, 2023.
11. Guanchi Yang, "Short-term prediction method of blood glucose based on temporal multi-head attention mechanism for diabetic patients", ELSEVIER (BSP&C), Volume 82, Number 104552, 2023.

12. Muhammed Asad, “Blood Glucose Level Prediction of Diabetic Type 1 Patients Using Nonlinear Autoregressive Neural Networks”, HINDAWI (JHE), Volume 2021, 2021, doi.org/10.1155/2021/6611091.
13. Jing Lv, Rao Li, “Development and Validation of a Risk Prediction Model for Foot Ulcers in Diabetic Patients”, HINDAWI (JDR), Volume 2023, 2023, doi.org/10.1155/2023/1199885.
14. Alok Aggarwal, “COVID-19 Risk Prediction for Diabetic Patients Using Fuzzy Inference System and Machine Learning Approaches”, HINDAWI (JHE), Volume 2022, 2022, doi.org/10.1155/2022/4096950.
15. Rashmik Manchiraju , Vandana Bhattacharjee, “A Machine Learning Approach to Diabetes Prediction with Feature Selection”, IJERT, Volume 11, Issue 07, 2022.