Performance Analysis of Neural Network Based Classifiers for the Prediction of Diabetes

J. PRADEEP M. HARIKRISHNAN, K.VIJAYAKUMAR
Department of Electronics & Communication Engineering Sri Manakula Vinayagar Engineering College
Puducherry, INDIA

Abstract— Diabetes is the most harmful diseases to consider in recent years since it causes severe damage to human beings in the form of elevated sugar levels. In a recent survey, it was projected that over 385 million public were affected in the entire world. Several investigators were conducted various experiments for prediction of diabetes using various classification techniques. This paper deals with a neural classifier based prediction system to recognize diabetes. Two learning algorithms namely, Levenberg Marquardt back propagation (LM), and gradient descent with variable learning rate are is investigated for different architecture and the best architecture with good accuracy was identified. The data are together from the Government Hospital of Pondicherry and it is formed as a database. Totally, datasets of 500 have been together, out of which 350 datasets as training sets for training process and 150 datasets as testing sets for the testing process. The recognition accuracy is obtained. For comparison, k-Nearest Neigourhood and the K- nearest neighbor and Radial Basis Function (RBF) network are also implemented and it is trained and tested with the same datasets. The result shows that Neural Network outperforms well with other classifiers.

Keywords: Neural Network, Gradient descent with variable rate Sigmoid Activation Function, Prediction, Diabetes, k- Nearest Neigourhood.

Received: May 15, 2021. Revised: February 18, 2022. Accepted: March 20, 2022. Published: April 26, 2022.

1. Introduction

Diabetes is the disease that causes severe harm to human beings, which elevated sugar levels at a high rate [1]. It causes severe continuing problems such as heart disease, kidney disorders, ulcers, and spoil eyes. At present, the kinds of diabetes, namely insipidus, and Mellitus. Insipidus is due to turn out inadequate insulin. In Mellitus, the cells dose not reacts to the creation of insulin. At present, the diabetic patient uses a Fingerstick device with lab tests for testing the elevated sugar level. However, this method is more painful and it consumes more time to detect the elevated sugar level of an individual. In order to defeat this drawback of the existing model, Neural Network-based classifiers [2] is introduced in the literature

A Multi-Layer Perception is often used for the prediction. Multi-Layer uses supervised learning and back propagation for training process. In the neural network, it has layers and nonlinear activation, which distinguish the linear perception. It can also distinguish whether it is linearly or non-linearly independent. It also focused, mainly on computer techniques to perform clinical diagnoses and the prediction with suggestions for the treatment.

Several kinds of research shown attention for diabetes prediction using machine learning and deep learning methods. The following reviews were studied in the literature. Thirugnana *et al.,* [3] proposed improved diabetes prediction using fuzzy neural networks. Afsaneh Morteza *et al.,* presented a neural network predicted albuminuria in type II diabetes compared the condition logistic regression [4].

Kevin *et al*.,[5] suggested a Machine Learning method for diabetes treatment to Predict Blood Glucose Levels. The proposed model has outperformed diabetes experts at estimate blood glucose rates and it can forecast 23% of hypoglycaemic cases 30 minutes. Sneha Joshi *et al*., [6] introduced MATLAB built-in forecasting method that can determine whether a in dividable is diabetes. The GUI is designed to make application user friendly so that even in the absence of a doctor, patients can get test result from assistants. The BPNN results used for predicting diabetes is 76%, which indicates the progress in the previous research.

Zahed Soltani and Ahmad Jafarian [7] proposed a neural network method for identifying the diabetes. The maximum training accuracy is 89.56% and testing accuracy 81.49% is obtained for the proposed framework.

Takoua Hamdi *et.al.*,[8] used an Neural Network for predict insipidus diabetes in the blood sugar levels Experimental tests showed that it was used for detect hyperglycemia or hypoglycemia quarter-hour well in advance. The key concept of ANN is to use the previous N steps to forecast subsequent steps. The Predetermined calculation is then used as reference with the previous (N-1) measurements to estimate consequential meaning and so forth. The calculation of the consequential values as am benefit is cumulative, elastic and nonlinear.

Surajini *et al.,* [9] proposed a prediction model of diabetes with the support of the Probabilistic Neural Network. He trained the prediction model using the Back propagation algorithm. PNN achieved the prediction model with minimal error and it shows the diabetic prediction.

Quanzou *et.al.* [10] used a decision tree and random

forest to predict diabetes mellitus. They randomly selected 68994 healthy people and diabetic patient's data as the training set. In this study, the proposed utilized principal component analysis and minimum redundancy maximum relevance to reducing the dimensionality.

Suresh Kumar *et.al.*,[11] implemented data mining strategies to determine the type of diabetes and its intensity degree for each individual from the data gathered including clustering and grouping. A base k-means algorithm is used to segment the whole dataset, classifiers the risk level of each patient as mild, moderate and server.

Vrushali Balpande *et al.,* [12] discussed the detailed review of existing data mining methods used for the prediction of diabetes. The K- Nearest Neighbor Algorithm, Bayesian Classifier, Naïve Bayesian Classifier methods are used for the prediction of diabetes, which gives patient's condition of Normal, Pre-diabetes, and diabetes.

Suyash Srivastava *et al*.,[13] proposed and presented a diabetes prediction with the help of Neural Network method and it archives 92% accuracy for predicting diabetes.

The above-discussed kinds of literature are the inspiration to initiate this paperwork. From the studies of various related and existing models, the idea for creating a prediction model of diabetes with the help of an artificial neural network is achieved with valuable knowledge. This paper proposes a diabetes prediction with a neural network classifier. The datasets are collected and created as a database. The datasets are used for the trained and tested process. The result of this prediction method is obtained and it is evaluated with the existing systems.

This paper is organized as follows. Section II describes the pre-processing of the data and the benefits of the proposed system, overcoming the disadvantages of the previous system. Section III confers the results of the prediction method. Section IV describes the comparisons of the proposed prediction model. Section V ends with a conclusion and Section VI provides the acknowledgment respectively.

TABLE.1 SAMPLE OF COLLECTED DATASET

Sl .No	NAME	GENDER	AGE	RBS	FBS	PPBS	UREA	CREATIVE	HBAIC	OUTCOME
1	Rajendiren	1	77	120	89	108	28	0.8	5.2	0
2	Soundarajan	1	54	174	92	126	34	2.4	5.5	0
3	Devanathan	1	55	145	150	281	77	2.6	7.2	1
4	Krishna	1	78	234	150	276	70	1	8.3	1
5	Velu	1	40	138	82	114	24	0.4	5.4	0
6	Rajaramam	1	40	210	84	135	38	1.6	7.3	1
7	Nedunchezian	1	47	113	84	154	20	0.9	6.2	0
8	Sanjeevi	0	46	84	82	172	20	0.9	5.8	0

2. Proposed Model

2.1 Database Description

In the proposed model, real-time diabetes dataset is collected from the Government Hospital of Pondicherry. The data consists of medical details of 500 instances, out of which 350 datasets are used as training sets for the training process and 150 datasets are used as testing sets for the testing process. The collected datasets consist of 10 attributes namely Random Blood Sugar, Fasting Blood sugar, Pre/Post Pradinal Blood Sugar, Urea, Creatinine, Glycated haemoglobin, Age, Gender, and Outcome. The value of Outcome '0' is considered as non-diabetic and the value of Outcome '1' is considered as diabetic. The collected dataset samples are shown in below Table.1

2.2 Data Preprocessing

For enhanced perceptive about the dataset and to obtain a high-quality result with a low error rate as possible from the prediction model, the data pre-processing and data visualizations are done. The data pre-processing are used on the dataset is listedbelow.

2.3 Neural Network Based Classifiers

A Neural Network (NN) technique shows the potential solution for the classifying for the prediction of diabetes. The features are the input to the different classifiers. The ability of the classification is determined from architecture of the network and the rule of learning. The architectures used in this paper are feed-forward, radial basis function and nearest neighborhood architecture. The prediction of diabetes is evaluated using the NN based classifier technique. Totally 500 datasets were collected, out of which 350 datasets are used as training sets for the training process, and 250 datasets are used as testing sets for the testing process. The prediction model of diabetes has been implemented using Matlab software. The feed- forward back propagation classifier is introduced and investigated. For comparison of accuracy, the K- nearest neighbor and Radial Basis Function (RBF) network are also designed with the help of a real-time diabetes dataset. The prediction models have been designed using all the various system as mentioned below.

2.3.1. Feed Forward Neural Network

In order to obtain the maximum recognition accuracy, different neural network architecture with two learning algorithms namely, Levenberg Marquardt back propagation (LM), and gradient descent with variable learning rate are investigated. It is observed from Table.2, that the hidden layer with 65 neurons gives the result with maximum accuracy. Thus, the two hidden layers with 65 neurons in each are used. For the testing process, the testing dataset is given to the trained neural network Architecture and it has perform is obtained. From the result, the Recognition accuracy is also determined.

TABLE.2 RECOGNITION PERFORMANCE OF NEURAL NETWORK ARCHIECTURE FOR DIFFERRENT LEARNING ALGORITHM

Sl.No	NN Architecture	Training Algorithm	Recognition accuracy (%)
1	8:30:2	LM	76
		GD with VLR	79
2	8:30:30:2	LM	82
		GD with VLR	84
3	8:40:2	LM	79
		GD with VLR	82
4	8:40:40:2	LM	65
		GD with VLR	75
5	8:65:2	LM	75
		GD with VLR	86
6	8:65:65:2	LM	82
		GD with VLR	96.47
7	8:75:2	LM	79
		GD with VLR	82
8	8:75:75:2	LM	69
		GD with VLR	65

2.3.2. k- Nearest Neighbor Network

The k-nearest neighbor algorithm is a technique used for classifying the neighborhood in the feature space [15]. The training stage consists of storing only data from the function vectors with class labels. The same features are computed from the test data at the classification level. To get closest neighbours, the Euclidean distance between the test data and the entire cumulative vector is measured input together, and the distances obtained are listed in ascending order the smallest distance is taken.

The k-nearest neighbor algorithm is applied in this paper. The Simulation results are obtained for the 3rd nearest neighbor which yields better accuracy and the results are tabulated in the subsequent section.

2.3.3. Radial Basis Function Network

Radial Basis Function (RBF) network has better quality and it is used in wide variety of functions [16]. RBF network has Gaussian function as nonlinearity for the transmission elements of hidden layers. The Gaussian function only refers to a specific area where the Gaussian is located in the input space. The key to successful implementation of these networks is to find appropriate centres for the Gussian functions. The basic architecture for an RBF is a 3-layer network that is investigated and the best architecture is obtained. The hidden layer neurons are 100 in the RBF network. For the classifying the prediction of the diabetes, two neurons are used in the output layer. The feed-forward neural network, RBF and the k-nearest neighbor network classifier are used for investigation and the performance study is carried out in the next section.

3. Result and Discussion

A real-time database is generated and the data's are collected from diabetic patients in the Government Hospital of Pondicherry. Totally 500 datasets were collected, out of which 350 datasets are used as training sets for the training process, and 250 datasets are used as testing sets for the testing process. From the proposed 8x65x65x2 neural network Architecture, the output is obtained and theaccuracy is 96.47%.

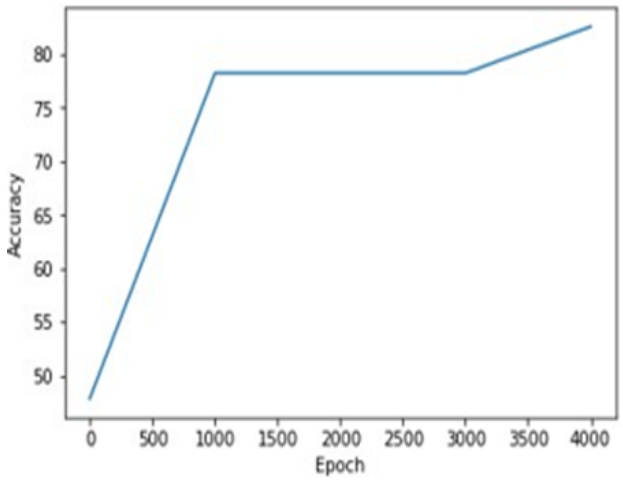

Fig.3 Performance Illustration of Gradient Descent Optimization

Fig. 3 shows that accuracy is steadily increased, when the epoch rate increases. It states that the epoch and the accuracy are directly proportional to each other. The accuracy achieved for the proposed prediction model is shown in Table.3 and it is compared with the other two classifiers.Table.3 shows the reduction of error rate concerning the epochs for the gradient descent optimization. The error rate was obtained for every 1000 iterations in the training process and it is shown as Table.3.

3.1 Performance Comparison of the Classifier

For the performance comparison, the k-nearest classifier and the logistic regression classifier is used. The training and testing process is carried with the help of the created same real-time database. After the training, the classifiers tested with the testing samples, and the results areobtained. The result obtained for the k-nearest and logistic regression classifiers are illustrated in the below table. This reveals from the table that the k-Nearest Neighbor classification offers 85.65 %accuracy and 89 % accuracy for the RBF network in classification. Table.3 shows that the average accuracy obtained for the neural network classifier of architecture 8:65:65:2 with

Gradient descent and variable rate accuracy is 96.47 % and the proposed neural classifier leads 10% of accuracy when it compared with other classifiers. Therefore, the neural network classifier outperforms well compared with the other two classifiers in terms of accuracy.Moreover, the proposed system is more suitable and efficient for real-time applications.

TABLE.3 DIABETES ACCURACY WITHDIFFERENT NN CLASSIFERS

S. No	Dataset	Classifiers	Accuracy
1	Real time dataset (AGE, GENDER, RBS, FBS, PPBS, UREA, CREATINE, HBA1C)	Neural Network	96.47%
2	Real time dataset (AGE, GENDER, RBS, FBS, PPBS, UREA, CREATINE, HBA1C)	Radial basis Function network	89%
3	Real time dataset (AGE, GENDER, RBS, FBS, PPBS, UREA, CREATINE, HBA1C)	k-Nearest Neighbour	85.65%

4. Conclusion

The prediction of diabetes using a Neural Network classifier is proposed. For the proposed system, the data collected from the Government Hospital of Pondicherry and the database is created using the data collected. Totally 500 datasets were collected, out of which 350 datasets are used as training sets for the training process and1 50 datasets are used as the testing set for the testing process. After the training process, the classifier is tested. The NN classifier with 65 neurons in the two hidden layers gives the result with maximum accuracy finally; the NN is obtained with maximum accuracy. The obtained result of the neural network classifier is evaluated with the otherclassifiers. The best performance among these classifiers is found and the proposed neural network classifier with 96.47% of accuracy is obtained. It shows that the performance of the proposed NN classifier outperforms well with the remaining classifier with respect to accuracy.

5. Acknowledgement

Our thanks to Dr. M. Sivakamy (General Duty Medical Officer) belonging to Government Hospital of Pondicherry for helping us by providing the relevant and most required real time diabetes dataset is used for train and test in the proposed system.

References

[1] Ishwarya, R., Gayathri, P., Jaisankar, N., 2013. A Method for Classification Using Machine Learning Technique for Diabetes. International Journal of Engineering and Technology (IJET) 5, 2903–2908

[2] Temurtas, H., Yumusak, N., Temurtas, F., "A comparative study on diabetes disease diagnosis using neural networks", Expert Syst, Vol. 36, pp. 8610–15 2009.

[3] Thirugnanam, Mythili, et al., "Improving the Prediction Rate of Diabetes Diagnosis Using Fuzzy, Neural Network, Case Based (FNC) Approach."Procedia Engineering, Vol.38, pp. 1709-118, 201

[4] Morteza, Afsaneh, et al., "Inconsistency in albuminuria predictors in type 2 diabetes: a comparison betweenneural network and conditional logistic regression", Translational Research, Vol.161, No.5, pp. 397-405, 2013.

[5] Kevin phlis, Razvan Bunescu, and Cindy Marling "A Machine Learning Approach in Predicting Blood Glucose Levels for Diabetes Management", *Modern Artificial Intelligence for Health Analytics*, pp: 35-37, 2014.

[6] Sneha Joshi and MeghaBorse, "Detection and Prediction of Diabetes Mellitus Using Back- Propagation Neural Network", Proceedings *International Conference on Microelectronics and Telecommunication Engineering*, pp: 110-113, 2016.

[7] ZahedSoltani and Ahmad Jafarian, "A New Artificial Neural Networks Approach for Diagnosing Diabetes Disease Type II", *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 6, pp: 85-94, 2016.

[8] Takoua Hamdi and Jaouher Ben Ali, "Artificial Neural Network for Blood Glucose Level Prediction", Proceedings *International Conference on Smart, Monitored and Controlled Cities*, pp: 17-19, 2017.

[9] P. Sujarani and Dr. K. Kalaiselvi, "Prediction of Diabetes Using Artificial Neural Networks: A Review, "*International Journal of Computer Applications & Information Technology*, Vol. 10, No. 4, pp: 67-72, 2018.

[10] Quanzou, Kaiyang Quo and Yamei Luo, "Predicting Diabetes Mellitus with Machine Learning Techniques", *Frontiers in Genetics*, Vol.9, No.10 pp:1-9,2018.

[11] P.Suresh Kumar and V. Umatejaswi, "Diagnosing Diabetes using Data Mining Techniques", *International Journal of Scientific and Research Publications*, Vol.7, No.6, pp: 705-707, 2017.

[12] Vrushali Balpande and Rakhi Wajgi, "Review on Prediction of Diabetes using Data Mining Technique", *International Journal of Research and Scientific Innovation*, Vol.4, No.1, pp: 43-46, January 2017.

[13] Parastoo Rahimloo and Ahmad Jafarian, "Prediction of Diabetes by Using Artificial Neural Network" *Logistic Regression Statistical Model and Combination of Them*, Vol. 85, pp: 1148-1164, 2016.

[14] Islam MR, Kamal ARM, Sultana N, Islam R, Moni MA et al, "Detecting depression using k-nearest neighbors (KNN) classification technique.", *International conference on computer, communication, chemical, material and electronic engineering* (IC4ME2). IEEE, pp 1–4. 2018.

[15] Amal H. Khaleel et al, "*A Weighted Voting of K- Nearest Neighbor Algorithm for Diabetes Mellitus*," International Journal of Computer Science and Mobile Computing, Vol.6 Issue.1, January- 2017, pg. 43-51

[16] Cheruku R., Edla D., and Kuppili V., "*Diabetes Classification using Radial Basis Function Network by Combining Cluster Validity Index and BAT Optimization with Novel Fitness Function*," International Journal of Computational Intelligence Systems, vol. 10, no. 1, pp. 247-265, 2017.