Prediction of Diabetes Using Artificial Neural Networks: A Review

P. Sujarani, Research Scholar, Department of Computer Science, Vels Institute of Science Technology and Advanced Studies (VISTAS), (Formerly Vels University) Chennai. E-mail:sujiraji873@gmail.com

Dr.K. Kalaiselvi, Associate Professor & Head, Department of Computer Science, Vels Institute of Science Technology and Advanced Studies (VISTAS), (Formerly Vels University), Chennai. E-mail:kalairaghu.scs@velsuniv.ac.in

Abstract--- Diabetes is the one of the major disease for all types of age group. Two major factors are to increase the diabetes and growing incidence at child or younger aged people and even born babies this is because of obesity and legacy of diabetes. The main aim of this paper is to predict the diabetes for the age less than 25 years. This pared a way to propose comparative study with previous studies based on Neural Networks. According to this study, Multilayer Perceptron (MLP) and Probabilistic Neural Network (PNN) were proposed .The performance and accuracy compared to PNN and MLP which was trained by Back propagation, PNN achieves high accuracy for the prediction of diabetes and whether the patient is diabetic or prediabetic.

Keywords: Diabetes, Multilayer Perceptron, Neural Network, Probabilistic Neural Network.

I. Introduction

Diabetes is a disease, it will occurs depending on the blood glucose level. When the glucose level is very high, it confirms the patient suffer with diabetes disease and it causes so many health problems such as kidney disease, heart disease and blood vessel damage etc. Now a days mainly children and young people are suffered from the diabetes disease. The one of the most common type of diabetes is Type II which was developed by the age of 40 and above. To predict this type of diabetes, the two classifiers are used. They are c4.5 and Neural network. Compared to the both classifiers accuracies, Neural network produced best accuracy to predict the diabetes[7].Risk estimation is one of the main source of prediction of diabetes. According to the risk factors whether the patient was suffered with type I and Type II Diabetes. Depending on the person body characteristics, Feed forward Neural Network Model to estimates the risk factors. They are weight, Blood sugar, Family history etc[9]. Diabetes is the one of the major problem in the present days. Various data mining techniques and different algorithms are used for finding diabetes and provide a better result to patients who are suffered from the diagnosis of diabetes. One of the best model proposed are decision tree and artificial neural network using diabetic data, when compared to both performance and implementation on weka software, DTA performed better than ANN [11]. ANN was the best approach which it is used for the data of diabetic patients the age between 25 to 78 years of old by using Bayesian regulation it provided the best performance than the other regulation techniques [5]. Back propagation algorithm is one of the neuro rule method was introduced and performed for more accurate. It will be reducing the redundant rules for future purpose [10]. Data mining used for various purposes in many application and various data mining techniques and different algorithms used for finding the diabetes. Neural network ,artificial fuzzy inference system, K-Nearest–Neighbor, Genetic algorithm and back propagation algorithm are provided the better result to the patients who are suffered from the diagnosis of diabetes[12]. The accuracy of prediction was compared to the Decision Tree,ID3 and other different classification algorithms against the ANN approach on pima Indian diabetes data set .These are implemented on weka , J48 shown highest accuracy[6].The ANN techniques are back propagation, Multilayer perceptron, Radial Basis Function are best suitable for prediction of diabetes[15].The statistical model are artificial neural network and logistic regression methods are used to analyze the data modeling and analyzes the data with different diabetes parameters[13].To improve the classification accuracy, ANN requires complete set of data for accurate classification and various missing value techniques. It was applied on pima Indian data set by using combination of preprocessing techniques[3].

II. Related Works

According to the previous studies, a comparative analysis was prepared for diabetes prediction. It can be predicted by used various data mining techniques, algorithms and the comparison noted down and we will find an accuracy easily, and different attributes can be used.

Probabilistic neural network approach can be used for diagnosing diabetes type II and the performance of the training and testing accuracy is 89.56% and 81.49%. Both accuracies are used to identify the diagnosing diabetes type II and we can use different approaches with combination of ANN approach in future[14]. For achieved best results and an enhanced the one of the best model was ANN to predict the diabetes disease by used Genetic Algorithm(GA) and trained with the Back Propagation Algorithm (BP).when compared to the other models and algorithms, ANN proved the performance and to improvise the accuracies using some techniques for further use[4]. The authors Xue-Hui Meng, Yi-Xiang Huang, Dong-Ping Rao,Qiu Zhang, Qing Lu[8] study to predict the diabetes or prediabetes by using risk factors to compare the performance of Logical regression(LR), ANN, and Decision Tree models. When compared to the classification accuracies of these three models the Decision Tree had the best accuracy of 77.87%, lowest accuracy achieved by ANN. This model is very useful to reducing diabetes shortages. In this study, General regression Neural Network(GRNN)model can be used for diagnosing the Pima Indian diabetes. when the performance compared to the Multilayer Perception(MLP), Radial basis function(RBF) and GRNN, which are applied to the pima Indian diabetes medical data. The GRNN gives the best to classify a medical data[1]. The authors Hasan Temurtas, Nejat Yumusak, Feyzullah Temurtas[2] suggested to the MLP structure which was enhanced by Levenberg–Marquardt(LM) algorithm and Probabilistic Neural Network structure are focused on diagnosis of diabetes disease, for this UCI learning data set used and compared to the previous research studies. The classification accuracies are mentioned for the comparative study on Pima Indian diabetes diagnosis and used same data set. The multilayer neural networks(MLNN) has been used two methods which are cross validation and conventional when compared to MLNN with levenberg-marquardt (LM)Algorithm, it achieve79.62% of the classification accuracy(10x FC). Compared to the MLNN with LM by used conventional valid . It gives better estimation results for diagnosis of diabetes disease.

III.   Proposed Model

3.1 Methodology

Multilayer Perceptron (MLP)

Multilayer perceptron(MLP) is a one of the feed forward neural network. The network connected with one more layers from input layer to output layer. Feed forward means that the information pass through in one direction from input layer to output layer.
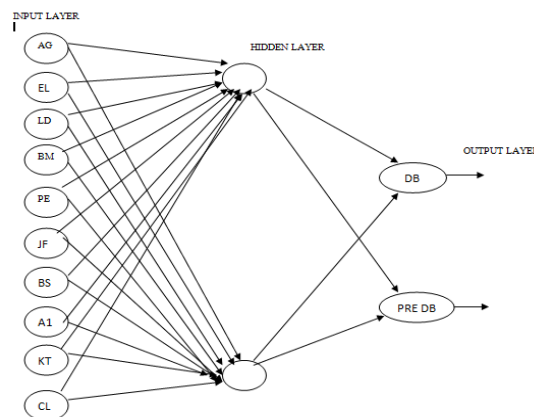

Fig. 1: Block Diagram of Multilayer Perceptron

MLP is widely used for prediction. [4] the author proposed that it has three layers. They are Input layer, Hidden layer, and Output layer. The Fig.1 shows the Block diagram of Multilayer perceptron with ten attributes of diabetes. Each input layer connected with hidden layer and each layer connect to the units. The inputs from the network that are correspond to the attributes which they measured for each training tuple. All inputs are connected into the units and it change as a input layer. These inputs transfer through the input layer and the weighted attributes are connected into the second layer. This second layer is called Hidden layer. The weighted outputs of the last hidden layers are input units and it changing as output. This Multilayer perceptron network was enhanced by Back propagation learning algorithm. Back propagation is a supervised learning feed forward topology that propagates back the errors, adjusting the weights till the model is adequately trained and comparing the prediction for each and every tuple with its actual target value. For every training tuple, the weights that are changed then it will minimize the mean squared

error between both the networks prediction and its actual target value. All changes are processed in backwards direction. Calculate the net input, and every input is connected from the unit which it is multiplied by its corresponding weight and I is the input and j is the unit and the net input is computed as:

$$I_j = \sum_i W_{ij} O_i + \theta_j$$

and O is the output , the unit j is calculated as:

$$O_j = \frac{1}{1+e^{-I_j}}$$

Two affective features of this algorithm are simplicity and its proven effectiveness for a wide range of problems. It can be implemented in any applications and without any problem. The main disadvantage is, it is very slow in networks with multiple hidden layers and it requires high processing time for large neural networks. So we can avoid these problems, introduce the one of the best neural network model is probabilistic neural network.

Probabilistic Neural Network (PNN)

Probabilistic Neural Network is one of the best methodologies in neural networks. In this paper proposed PNN for the best accuracy and prediction by using Bayes theorem. The following fig.2 shows the Block diagram of PNN.
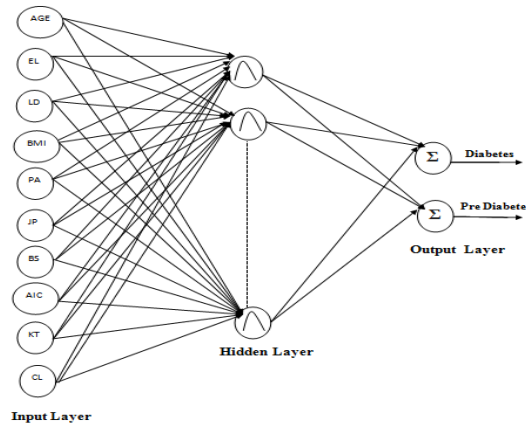

Fig. 2: Block Diagram of Proposed Probabilistic Neural Network

PNN consists of four layers. [16]the author suggested that the layers are input layer, hidden layer, summation layer, output layer. The prior probability of any pattern which it is belong to a class and it is measured by fraction of patterns in that class assuming number of patterns in the training set. [16] The first layer calculates the distance from the input vectors when the input is present. It will produces the vector where the attributes show how it will close the input to the training input and the second layer adds the contribution for that each class of inputs. It gives the net output as a vector of probabilities. At last a complete transfer function on the output of the second layer takes the maximum of these probabilities. Finally it shows '1' for targeted class '0' for non targeted class. The probabilities as described by bayes theorem. The probabilities of two classes and the probability density function of the form is:

$$P(x_i) \; = \; \frac{1}{\sqrt{2\pi}\sigma_{i2}}\exp\left(-\frac{(x_i - \bar{x}_i)^2}{2\sigma_{i2}}\right)$$

Where the mean $\bar{x}_i$ and the variance σ_{i2} satisfy:

$$\bar{x}_i = \; E[x_i] = \int_{-\infty}^{\infty} x_i p(x_i)dx_i$$

$$\sigma_{i2} = E[(x_i - \bar{x}_i)^2] = \int_{-\infty}^{\infty}(x_i - \bar{x}_i)^2 \; p(x_i)dx_i$$

The main benefits of PNN are

PNN is much faster than Mutilayer perceptron.

PNN produces high accuracy than Multilayer Perceptron networks.

PNN model Bayes optimal classification.

It generate accurate predicted target probability scores.

PNN suitable for prediction problems than MLP

3.2 Data Set

The Dataset contains ten Attributes which are suitable for the age of less than 25 years. The Attributes are Age, Education level, Legacy of diabetes, Body mass index, Physical Exercise, Junk foods, Blood sample,A1c test, Ketone test and cholestral. The following Table.1 shows the attributes of prediction of diabetes and their description.

Table.1: Attributes and Description for Prediction of Diabets

Attribute No	Attribute Name	Description
1	Age	0 to 10 yrs 10 to 20 yrs 20 to 25 yrs
2	Education Level	Less than 5 yrs Between 5 to10 years Between 10 to 12 years
3	Legacy of Diabetes	Yes(Diabetes) No(pre Diabetes)
4	Body Mass Index	Weight in kg. Less than 20 and greater than equal to20 (0-10 years) less than 30 and greater than equal to 30 (10-25 years)
5	Physical Exercise	Less than per a day or Four days per a week one hr or more a day or at least four days in a week
6	Junk Foods	Less than two times in a week At least two times in a week
7	Blood Sample	70-100 mg (before past) 140-200mg(after food)
8	A1C test	Three months or One year(Test report)
9	Ketone Test	Glucose level high or low
10	Cholestral	100 mg/dl or less(LDL) 45mg/dl or more (HDL)

3.3. Proposed Architecture

The diabetes Dataset collect from the UCI and select the suitable attributes from the diabetic patients. It is extracted and applied the models of MLP and PNN Neural Networks using the selected data .Finally Probabilistic Neural Network achieves highest accuracy better than MLP. The following fig.3 shows proposed architecture for prediction of diabetes.
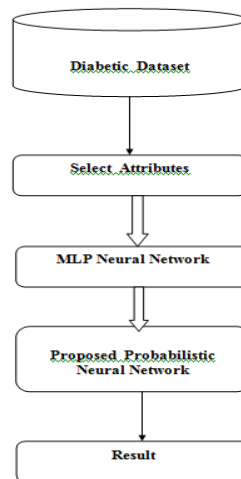

Fig. 3: Proposed Flow Diagram

IV.   Results and Discussion

The following Table.2 shows the comparative study of research done by many authors and different attributes, Methodologies and their Accuracies. Compared to the all neural network methodologies, probabilistic neural network achieved highest accuracy better than all. According to this performance, proposed PNN gives best accuracy with different attributes when compared to the previous studies accuracies.

Table 2: Comparative Study for Predicting Diabetes

S.no	Attributes	Methodologies	Accuracy
1	Plasma, Pressure, Skin, Insulin, Pregnancy, Mass, Pedigree, Age.	PNN	89.56%
2	Plasma, Pressure, Skin, Insulin, Pregnancy, Mass, Pedigree, Age.	ANN with GA	73.4%
3	Age, Family history ,Martial status, Educational level, Work stress, Sleeping Duration, Physical fitness, food with Salty, Gender, Eat fish, Drinking of Coffee, BMI.	Logistic Regression ANN C5.0	76.13% 73.23% 77.87%
4	Plasma, Pressure, Skin, Insulin, Pregnancy, Mass, Pedigree, Age.	GRNN	80.21%
5	Pregnancy, Plasma, Pressure, Skin, Insulin, BMI, Pedigree, Age.	MLNN with LM PNN	79.62% 78.05%

V.   Conclusion

The main aim is to compare with other research works done using Neural Networks. Different parameters were considered by the researchers to predict the diabetes disease. Here the proposed model mentioned based on the different attributes related to diabetes patients at the age less than 25 years.PNN was the best Neural Network model than MLP and other Artificial Neural Networks. It achieves high accuracy to predict the diabetes and PNN has fast training process so we will predict whether the patient is diabetic or prediabetic information or not in a quick process.

References

[1]   Kayaer, K. and Yıldırım, T. Medical diagnosis on Pima Indian diabetes using general regression neural networks. *Proceedings of the international conference on artificial neural networks and neural information processing (ICANN/ICONIP)*, 2003, 181-184.

[2]   Temurtas, H., Yumusak, N. and Temurtas, F. A comparative study on diabetes disease diagnosis using neural networks. *Expert Systems with applications* **36** (4) (2009) 8610-8615.

[3]   Jayalakshmi, T. and Santhakumaran, A. A novel classification method for diagnosis of diabetes mellitus using artificial neural networks. *IEEE International Conference on Data Storage and Data Engineering (DSDE),* 2010, 159-163.

[4]   Pradhan, M. and Sahu, R.K. Predict the onset of diabetes disease using Artificial Neural Network (ANN). *International Journal of Computer Science & Emerging Technologies* **2** (2) (2011).

[5]   Sapon, M.A., Ismail, K. and Zainudin, S. Prediction of diabetes by using artificial neural network. *Proceedings of the International Conference on Circuits, System and Simulation, Singapore*, 2011.

[6]   Ahmad, A., Mustapha, A., Zahadi, E.D., Masah, N. and Yahaya, N.Y. Comparison between Neural Networks against Decision Tree in Improving Prediction Accuracy for Diabetes Mellitus. *Digital Information Processing and Communications*, 2011, 537-545.

[7]   Priya, S. and Rajalaxmi, R.R. An improved data mining model to predict the occurrence of type-2 diabetes using neural network. *International Conference on Recent Trends in Computational Methods, Communication and Controls,* 2012.

[8]   Meng, X.H., Huang, Y.X., Rao, D.P., Zhang, Q. and Liu, Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *The Kaohsiung journal of medical sciences* **29** (2) (2013) 93-99.

[9]   Jain, D. and Singh, D. A Neural Network based Approach for the Diabetes Risk Estimation. *International Journal of Computer Applications* **73** (10) (2013).

[10]   Pushkar, S. Data mining using Artificial Neural Network Rules. *International Journal of Innovations in Engineering and Technology* **3** (1) (2013).

[11]   Alhassan, J.K., Attah, B. and Misra, S. Performance Analysis of Artificial Neural Network with Decision Tree in Prediction of Diabetes Mellitus. *World Academy of Science, Engineering and Technology, International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering* **9** (9) (2015) 669-672.

[12]   Singh, S. and Kaur, K. A Review on Diagnosis of Diabetes in Data Mining. *International Journal of Science and Research (IJSR)* (2012).

[13]   Rahimloo, P. and Jafarian, A. Prediction of Diabetes by Using Artificial Neural Network, Logistic Regression Statistical Model and Combination of Them. *Bulletin de la Société Royale des Sciences de Liège* 85 (2016) 1148-1164.

[14]   Soltani, Z. and Jafarian, A. A New Artificial Neural Networks Approach for Diagnosing Diabetes Disease Type II. *International Journal of Advanced Computer Science & Applications* **1** (7) (2016) 89-94.

[15]   Usha, S. Suitability of neural network for disease prediction: A comprehensive Literature Review. *Research Journal of Computer and Information Technology Sciences* **5** (6) (2017) 12-20.

[16]   Specht, D.F. Probabilistic neural networks. *Neural networks* **3** (1) (1990) 109-118.