Prediction of Diabetes Based on Artificial Intelligence Technique

Shameem Hasan

M.Sc. Engineering Student, Department of Electrical & Electronics Engineering, Islamic University of Technology (An organization of OIC), Gazipur-1704, Bangladesh

---

Abstract:- *Diabetes is a chronic disease which has the ability to cause a worldwide health hazard. According to World Health Organization (WHO), about 422 million people are living with diabetes around the world. By 2035, this number will be increased to 592 million. It is a disease caused due to increase in the level of blood glucose. Diabetes affects human organs such as kidney, eye, heart, nerves, foot etc. Various traditional methods are available for diagnosing diabetes. However, early prediction of diabetes is quite challenging task due to its complex interdependence on various factors. Following its development, a new challenging area in the field of artificial intelligence has been opened and an accurate prediction method of diabetes has been targeted by researchers. The objective of this paper is to develop a system that can offer a precise early prediction of diabetes for a patient with the help of artificial intelligence technique. The datasets consist of several medical predictor variables and one target variable. Independent variables include the Body Mass Index (BMI), insulin level, age, number of pregnancies the patient had (for female) and some others. Based on these parameters, prediction of diabetes, applying artificial intelligence technique, described in this article, seems quite satisfactory.*

Keywords: Artificial Intelligence Technique, Body Mass Index (BMI), Human Health, Mean Square Error.

1. INTRODUCTION

Diabetes is a serious chronic disease that occurs either when the pancreas does not produce enough insulin (a hormone that regulates blood sugar or glucose) or when the body cannot effectively use the insulin it produces [1]. Diabetes is an important public health problem, one of four priority non-communicable diseases (NCDs) targeted for action by world leaders. Diabetes mellitus occurs throughout the world, but is more common (especially type 2) in the more developed countries. The greatest increase in diabetes patient is occurring in low- and middle-income countries of Asia and Africa, where most diabetes patients will probably be found by 2030. The increase in incidence in developing countries follows the trend of urbanization and lifestyle changes, including increasingly sedentary lifestyles, less physically demanding work and the global nutrition transition, marked by increased intake of foods that are high energy-dense but poor in nutrition. The number of people with diabetes has risen from 108 million in 1980 to 422 million in 2014.

Type 1 diabetes previously known as insulin-dependent, juvenile or childhood-onset diabetes is characterized by deficient insulin production and requires daily administration of insulin. The cause of type 1 diabetes is not known and it is not preventable with current knowledge. Symptoms include excessive excretion of urine (polyuria), thirst (polydipsia), constant hunger, weight loss, vision changes, and fatigue. These symptoms may occur suddenly.

The risk of getting type 2 diabetes has been widely found to be associated with lower socio-economic position across countries. Hyper glycaemia or raised blood sugar is a common effect of uncontrolled diabetes and over time, leads to serious damage to many of the body's systems, especially the nerves and blood vessels [2,3]. In 2014, 8.5% of adults aged 18 years and older had diabetes. In 2016, diabetes was the direct cause of 1.6 million deaths and back in 2012 high blood glucose was the cause of another 2.2 million deaths. Healthy diet, regular physical activity, maintaining a normal body weight and avoiding tobacco use are ways to prevent or delay the onset of type 2 diabetes.

Diabetes and its complications bring about substantial economic loss to people with diabetes and their families and to health systems and national economies through direct medical costs and loss of work and wages. While the major cost drivers are hospital and outpatient care, a contributing factor is the rise in cost for analogue insulins 1 which are increasingly prescribed despite little evidence that they provide significant advantages over cheaper human insulins [4,5].

Type 1 diabetes cannot be prevented with current knowledge. Effective approaches are available to prevent type 2 diabetes and to prevent the complications and premature death that can result from all types of diabetes. These include policies and practices across whole populations and within specific settings (school, home, workplace) that contribute to good health for everyone, regardless of whether they have diabetes such as exercising regularly, eating healthily, avoiding smoking, and controlling blood pressure and lipids. Taking a life-course perspective is essential for preventing type 2 diabetes, as it is for many health conditions. Early in life, when eating and physical activity habits are formed and when the long-term regulation of energy balance may be programmed, there is a critical window for intervention to mitigate the risk of obesity and type 2 diabetes later in life [6]. In this paper several parameters are selected based on analysis to predict diabetes and then depending on those

parameters with the help of artificial neural network simulations are done to determine the accuracy of the prediction [7-14].

2. STATISTICAL ANALYSIS

In this research work, many statistical approaches such as Bayesian regularization (BR), Levenberg–Marquardt algorithm (LM) and scaled conjugate gradient (SCG) have been used for the evaluation and accuracy of the performance and results. Bayesian classification is the technique to construct the classifiers. Classifiers are nothing but the models that assign the class labels to the problem instance. Levenberg–Marquardt algorithm is also known as DLS that is damped least squares used for solving generic curve fitting problems by finding the local minimum [7,14]. Scaled conjugate gradient is feed-forward and supervised algorithm for neural networks. Feed forward here means that in connections there is no loop between the units. The general equations corresponding to each are mentioned below:

Bayesian regularization:

$$x = \arg_{b\in\{1,....B\}} \max\, p(C_b)\prod_{i=1} np(y_i|C_b) \tag{1}$$

Levenberg–Marquardt algorithm:

$$H(\beta) = \sum_{j=1} m[x_j - f(y_j,\beta)]^2 \tag{2}$$

Scaled conjugate:

$$S_k = \frac{\dot{E}(W_k + \sigma_k P_k) - \dot{E}(W_k)}{2} \tag{3}$$

Bayesian regularization, Levenberg–Marquardt algorithm and scaled conjugate are the various algorithmic parameters and functions used in the neural networks. BR can eliminate or reduce the need for lengthy cross-validations and it is more robust than the standard back-propagation methods, whereas to solve nonlinear least squares problems, the LM technique is considered to be the standard one as it shows lower performance in terms of predictive ability. On the other hand, SCG needs O(n) of memory where n represents the number of weights in the network although it uses second order of information from neural networks [8,14]. Among these three, the BR is considered to be the optimal one as it develops the nonlinear relationships and it has more predictive abilities. To get better and refined results, the data was tested through ten hidden layers, and on observing the results, it can be seen that the BR shows least mean square error (Table 1) for diabetes prediction. In further sections, a brief introduction of neural networks is cited and the results in regard to BR have been shown and explained in the sections after it.

TABLE 1: No. of Hidden layers, Mean Square Error (MSE) and Iterations for Diabetes prediction

Serial No.	No. of hidden layers	Mean Square Error (MSE)	No. of iterations
1	10	0.13341	62
2	15	0.12593	237
3	20	0.14115	87
4	25	0.13725	159
5	30	0.13723	166

2.1 Neural Networks

It is one of the concepts which has been inspired by the functionality of the human brain and its performance in identification of phenomena. Neurons (a single neuron shown in Fig - 1) are placed in different layers in multilayer neuron network. Input layer being the first layer receiving information and till its capability with other neurons, it transfers the information in the form of input signals to the other next layers. Neuron weight is the communication ability of each neuron with other neurons. The number of neurons in each layer depends on the weight of neuron and the previous layers' neurons. In addition to the input layer, the neural network also consists of the hidden layers and the output layers. Some of the advantages of using artificial neural networks(ANN) are its immaculate an on-point accuracy on a wide variety range of problems, less requirement of formal statistical training, offering various multiple training algorithms and having the implicit ability of detecting nonlinear complex relationships between independent and dependent variables. In this, neuron is the main processor and adding neurons to hidden layers will reduce calculation error but will be more time consuming for calculations. In the next section, the methodology used in the paper and results have been discussed.

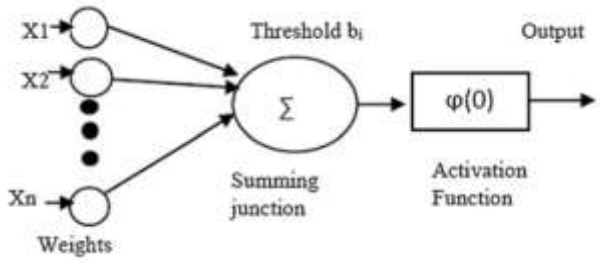

Fig – 1: Single neuron

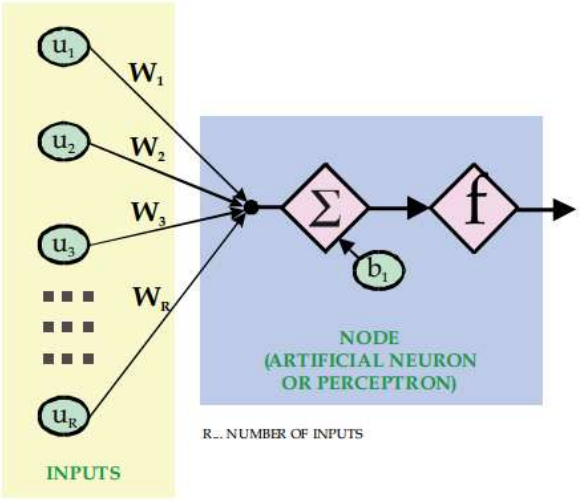

Fig - 2: Node (Artificial neuron or perceptron)

The neural network has 8 input layer, several hidden and one output layer. The input parameters are the values of different effective factors of patients, the auto-regression component and the physical variables for which the statistical analysis has shown strongest influence (Fig - 3). The number of auto-regression components *m* is determined by the correlation coefficients between the predicted value and its previous values persistence of the predicted variable. The goal is forecasting one step ahead. The number of neurons in the first and second layers is determined by the criterion of the minimum squared error(MSE).

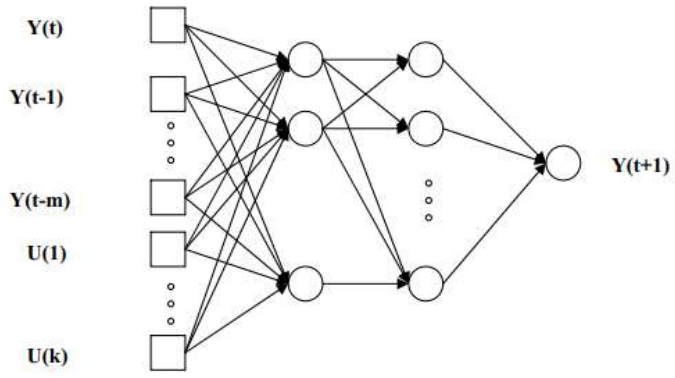

Fig - 3: Architecture of Neural Network

3. DATA AND METHODOLOGY

The dataset used in this paper is of a hospital. The datasets include 8 input parameters. The parameters are

- Pregnancies - number of times pregnant
- Glucose - plasma glucose concentration of 2 hours obtained from an oral glucose tolerance test
- Blood pressure - diastolic blood pressure (mmHg)
- Skin thickness - Triceps skin fold thickness (mm)
- Insulin - 2-hour serum insulin (mu/ml)
- BMI - Body Mass Index (Weight in kg/ height in m^2)
- Diabetes pedigree function
- Age (Years)

There are three portions of the data that act as training set, validation set and testing set in neural network shaving 70%, 15%, and 15% weightage respectively. These values can be changed but these particular values give better results. For this paper, 768 sample data of different patients have been used collected from internet sources.

The nftool (neural fitting tool) of MATLAB has been used in our proposed work to determine the performance and the results. In this tool, the number of input data to layer and the no. of hidden layers have to be de*fi*ned. In this paper 10, 15, 20, 25and 30 hidden layers have been used one by one to get the output and their mean square errors have been compared to get the performance measures. Since layer 15gives most refined results with least value of error, the data has been trained till 15layers.

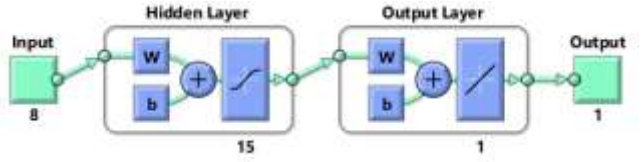

Fig - 4: Neural network working for diabetes prediction

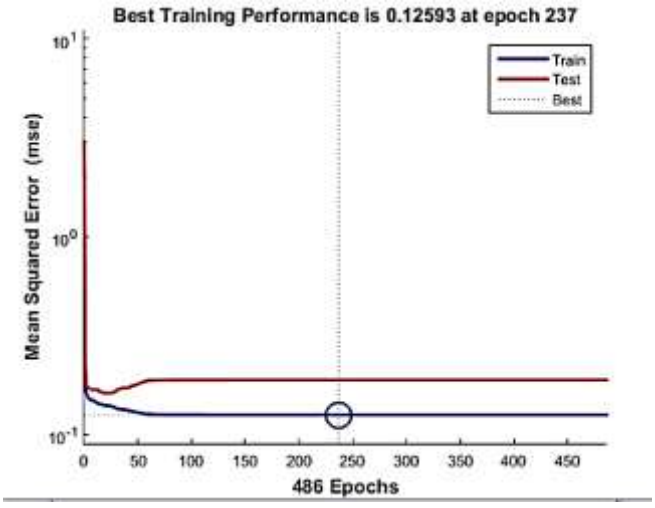

Fig - 5: Performance graph for diabetes prediction showing mean square error

The neural network is given with the inputs such as the values of no. of times pregnant, glucose level, blood pressure measurement, skin thickness, amount of insulin, BMI, diabetes pedigree function, age, the hidden layers, and the output having three desired values such as high chance (1), moderate chance (0.5), and low chance (0) of diabetes. These are levels of the prediction in which the range of its values 300 and above is considered to be serious, range from 150 to 300 is medium, and 0–150 is safe. The harmful impact of

diabetes is explained in the next section. Table 1shows the number of hidden layers with their mean square.

TABLE 1and Fig - 5show that the validation, test data sets and the performance of training with respect to epochs of diabetes prediction. As it is clearly seen the best performance measure is at Serial No. 2 that is with the 15 hidden layers and with minimum mean square value that is 0.12593 for prediction. Fig - 6 shows the 2D plotting of the input and desired output variables for prediction. Further Fig - 7 and Fig - 8show the regression curve between the target data and the output of levels of the diabetes prediction and the error histogram of the analysis respectively.

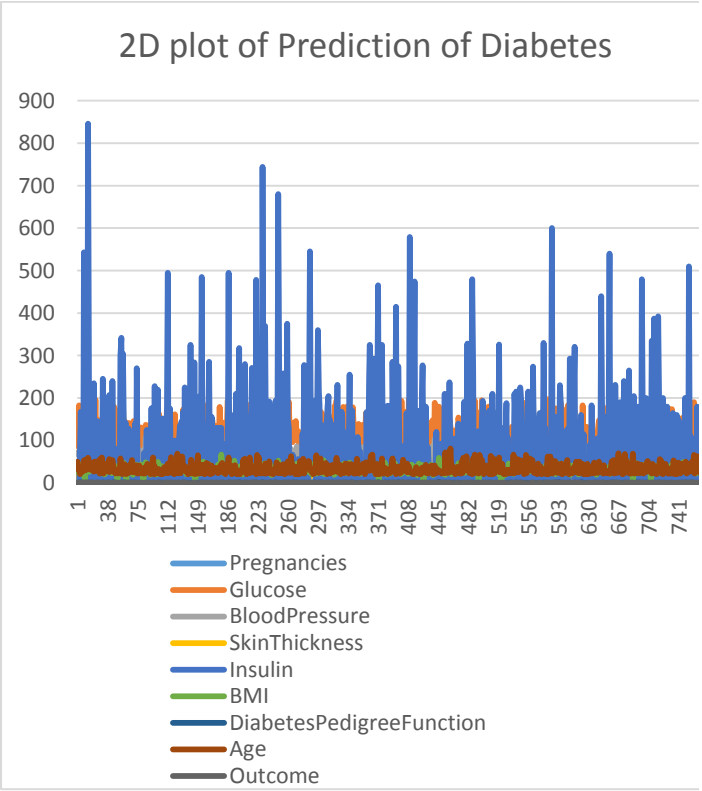

Fig - 6: 2D plot of Prediction of diabetes.

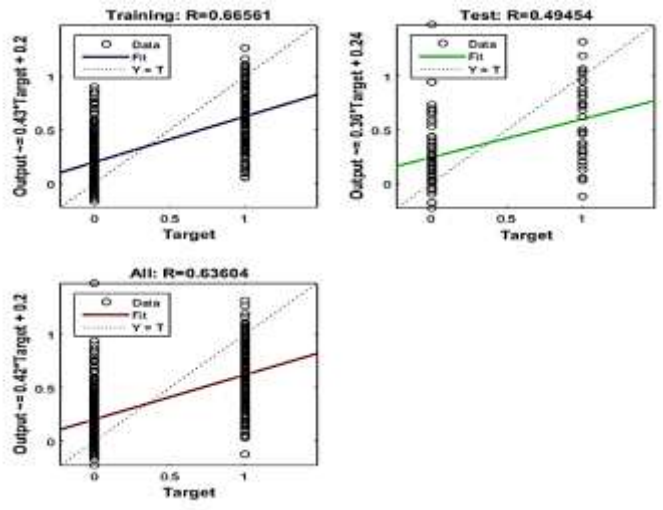

Fig - 7: Regression curve for diabetes prediction

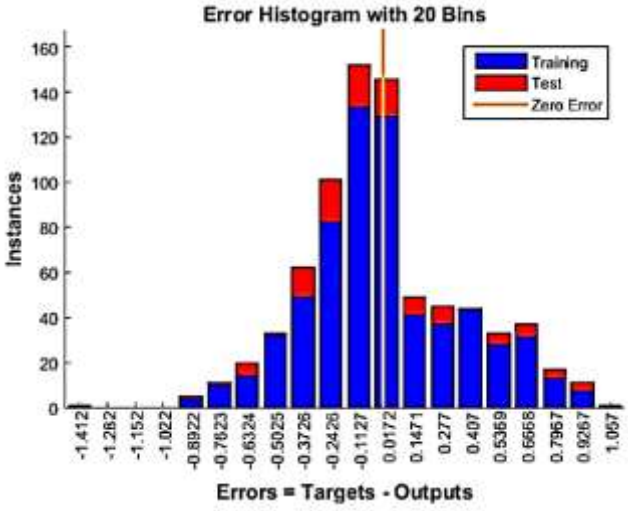

Fig - 8: Error histogram for diabetes prediction

4. IMPACT ON FUTURE GENERATION

DNA is the primary mechanism of inheritance. Kids get half their genes from mom and half from dad.

However, scientists are just starting to understand additional kinds of inheritance like metabolic programming, which occurs when an insult during a critical period of development, either in the womb or soon after birth, triggers permanent changes in metabolism.

In a recent study, researchers looked at the effects of a diet high in saturated fat on mice and their offspring. As expected, they found that a high-fat diet induced type 2 diabetes in the adult mice and that this effect was reversed by stopping the diet. However, if female mice continued a high-fat diet during pregnancy and suckling, their offspring also had a greater frequency of diabetes development, even though the offspring were given a moderate-fat diet. These mice were then mated with healthy mice, and the next generation offspring (grandchildren of the original high-fat fed generation) could develop diabetes as well. In effect, exposing a fetal mouse to high levels of saturated fats can cause it and it's offspring to acquire diabetes, even if the mouse goes off the high-fat diet and it's young are never directly exposed. The study used mice so it is not time to warn women to eat differently during pregnancy and breastfeeding but earlier research has shown that this kind of inheritance is at work in humans. From the above discussion the health hazard related to diabetes is realizable.

5. CONCLUSION

Diabetes has become a major health care problem over the world. As a result, some actions are needed to be taken to reduce the impact of diabetes such as according greater priority to prevention and control including strengthening national capacity to accelerate country response. On the other hand, initiatives such as creating,

sustaining and expanding health promoting environments to reduce modifiable risk factors, monitoring trends and determinants to evaluate progress are also necessary. In this paper, the predictive analysis of diabetes has been done through artificial neural networks. The different parameters of the neural network have been used to get the accurate result. A detailed discussion on the impact of diabetes on human health is also presented.

REFERENCES

[1] Djakouré -Platonoff C, Radermercker R, Reach G, Slama G, Selam J, "Accuracy of the continuous glucose monitoring system in inpatient and outpatient conditions" Diabetes Metabolism, 2003,28, pp.159–62.

[2] Sparacino G, Zanderigo S, Corazza S, Maran A, Facchinetti A, Cobelli C**. "**Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series**",** IEEE Trans Biomed Eng., May 2007, 54(5), pp. 931-937.

[3] Palerm C, Willis J, Desemone J, Bequette B. "Hypoglycemia prediction and detection using optimal estimation", Diabetes Technol Ther, 2005, 7(1), pp. 3–14.

[4] Zarkogianni K, Mougiakakou SG, Prountzou A, Vazeou A., "An insulin infusion advisory system for type 1 diabetes patients based on non-linear model predictive control methods", Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 5972–5975.

[5] Pappada S, Cameron B, Rosman P, Bourey R, Papadimos T, Oloruntu W, "Neural network-based real-time prediction of glucose in patients with insulin dependent diabetes", Diabetes Technol Ther, 2011,13(2), pp. 135–141.

[6] Gniuli D, Calcagno A, Caristo ME, Mancuso A, Macchi V, Mingrone G, Vettor R**, "**Effects of high-fat diet exposure during fetal life on type 2 diabetes development in the progeny**",** The Journal of Lipid Research, 2008; 49 (9), 1936-1945.

[7] Xingang F, Shuhui L, Fairbank M, Wunsch DC, Alonso E., "Training recurrent neural networks with the Levenberg–Marquardt algorithm for optimal control of a grid connected converter", IEEE Trans Neural Network Learn System, 2015, 26(9), pp.1900–1912.

[8] Stulp F, Sigaud O. "Many regression algorithms, one unified model: a review on Neural Network", 2015,69, pp. 60–79.

[9] Hayashia Y, Setiono R. "Combining neural network predictions for medical diagnosis", Computational Biological Med,2002,32, pp. 237–246.

[10] Pérez-Gandia C, Facchinetti A, Sparacino G, Cobelli C, Gomez EJ, Rigla M, "Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring", Diabetes Technol Ther, 2010, 12(1), pp. 81–88

[11] Mougiakakou SG, Prountzou A, Iliopoulou D, Nikita KS, VazeA, Bartsocas CS, "Neural network based glucose-insulin metabolism models for children with type 1 diabetes", Proceeding of 28th Annual International Conference IEEE Engineering Medical Biology Society, 2006, pp. 3545–3548.

[12] Zecchin C, Facchinetti A, Sparacino G, De Nicolao G, CobelliC., "Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration", IEEE Trans Bio-Med Eng., 2012, 59(6), pp. 1550–1560.

[13] Gianola, D.; Okut, H.; Weigel, K.A.; Rosa, G.J.M. Rosa, "Predicting complex quantitative traits with Bayesian neural networks: A case study with Jersey cows and wheat" *BMC Genetics*, *12(87)*, 2011, pp. 1–37.

[14] M.I.A. Lourakis, "A brief description of the Levenberg-Marquardt algorithm implemented by levmar", Technical Report, Institute of Computer Science, Foundation for Research and Technology - Hellas, 2005.