International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 10, October - 2020, Pages: 14-22

Artificial Neural Network for Predicting Diabetes Using JNN

Hussam Hatem Harz, Ahmed Osama Rafi, Musbah Osama Hijazi's, Samy S. Abu-Naser

Department of Information Technology,
Faculty of Engineering and Information Technology,
Al-Azhar University, Gaza, Palestine

Abstract *Diabetes is one of the most common diseases worldwide where a cure is not found for it yet. Annually it cost a lot of money to care for people with diabetes. Thus the most important issue is the prediction to be very accurate and to use a reliable method for that. One of these methods is using artificial intelligence systems and in particular is the use of Artificial Neural Networks (ANN). Therefore, in this paper, we used artificial neural networks to predict whether a person is diabetic or not. The criterion was to minimize the error function in neural network training using a neural network model. After training the ANN model, the average error function of the neural network was equal to 0.01 and the accuracy of the prediction of whether a person is diabetics or not was 987.3% .*

Keywords: diabetes, neural network, ANN, prediction.

1. Introduction

Diabetes is a long-lasting disease that happens when the pancreas fails to create enough insulin, or when the body cannot use the insulin produced efficiently. Insulin is a hormone that controls the level of sugar in the blood. Hyperglycemia or hyperglycemia is a common result of uncontrolled diabetes and, over time, causes severe damage to many organs, particularly nerves and blood vessels. In 2015, 8.5% of adults aged 17 years or older had diabetes. In 2013, diabetes was the cause of 1.5 million deaths, and high blood glucose caused 2.3 million deaths. Diabetes patients have doubled in the last ten years worldwide. More than 200 million people are infected and about seven percent increase in the annual predominance of diabetes in the world. People for a long time suffered from different diseases that in some cases have been able to diagnose diseases and offer them the solution in order to enhance it, but unfortunately, sometimes, due to the lack of diagnosis of symptoms in patients for a long time may even threaten the life of the patient. Therefore, many studies have been done in the field of predicting for several diseases to the extent that today's human take advantage of decision supports models and smart method to predict. One of the decision support models application is in the medical field and diagnosis of illnesses such as diabetes [1, 2]. Deferment in the diagnosis and prediction of diabetes due to insufficient control of blood glucose increases macro vascular and Capillaries difficulties risk, ocular diseases and kidney failure [1, 2]. So we proposed an ANN model to predict diabetes that can be useful and helpful for doctors and practitioners. In this research, we used the following attributes: Number of pregnancies, PG Concentration (Plasma glucose at 2 hours in an oral glucose tolerance test), Diastolic BP (Diastolic Blood Pressure (mm Hg))), Tri Fold Thick (Triceps Skin Fold Thickness (mm)), Serum Ins(2-Hour Serum Insulin (mu U/ml)), BMI (Body Mass Index: (weight in kg/ (height in m)^2)), DP Function(Diabetes Pedigree Function), Age (years), Diabetes (Whether or not the person has diabetes)[15].

Based on the Diabetes Research Center reports, the incidence of diabetes has folded in the last ten years worldwide and more than 200 million people are infected and about seven percent increase in the annual prevalence of diabetes worldwide. Since diabetes is a long-lasting disease and import permanent damage to the limbs and vital organs in the body, using artificial intelligence tools can enhance the detection methods and disease control which will be of a great help to the physicians. According to the Diabetes Research Center, it has been shown that early diagnosis of patients at risk can prevent 80 percent of lasting complications of type II diabetes or deferred them [5]. There are two types of diabetes, type I and type II diabetes. Type I diabetes also named insulin dependent and type II diabetes named relative insulin deficiency [6]. Protracted complications of diabetes are mainly distributed into two categories: vascular and nonvascular complications of diabetes. Vascular complications include micro vascular (eye disease, neuropathy, nephropathy) and macro vascular complications (coronary artery disease, peripheral vascular disease, cerebrovascular disease). Non-vascular complications include gastro paresis, sexual dysfunction, and skin changes [7].

2. The objectives of the study

- To predict and categorize the state of health.
- To identify some appropriate factors that affect health conditions,
- To design an artificial neural network that can be used to predict health performance based on certain pre-defined data for a particular health condition

3. Literature review

Diabetes or diabetes mellitus is a metabolic disorder (metabolic) in the body. This disease destroy the ability to produce insulin in the patient's body or the body develops resistance to insulin the and consequently the produced insulin cannot achieve its normal job. The main role of the produced insulin is to decrees blood sugar by different instruments. There are two key types of diabetes. In Type I diabetes, obliteration of beta pancreatic cells damage insulin construction and in type II, there is a progressive insulin confrontation in the body and ultimately may yield to the obliteration of pancreatic beta cells and faults in insulin production. In type II diabetes, it is known that genetic issues, obesity and lack of physical activity have a vital part in a person [1]. Even though the precise cause of type I diabetes is unidentified, issues that may indicate a greater risk comprise the followings [2]:

- Family history. A person risk upsurges if his parent or sibling has history of type I diabetes.
- Environmental factors. Situations for example contact with a viral illness probably play some role in type I diabetes.
- The existence of harmful immune system cells. Occasionally family members of a person with type I diabetes are examined for the existence of diabetes autoantibodies. If a person has these autoantibodies, he/she has a chance of increased risk for evolving type I diabetes. Nonetheless not every person who has these autoantibodies gets diabetes.
- Geography. Some countries, like Sweden, have bigger rates of type I diabetes.

Researchers don't completely comprehend why certain people develop pre-diabetes and type II diabetes and others don't. It's sure that some factors upsurge the risk like [2]:

- Weight. The more fatty tissue you have, the more resilient a person cells to insulin.
- Inactivity. The less energetic a person is, the more a person has risk. Physical activity assists a person control of his/her weight, consumes glucose as energy and makes a person cells more sensitive to insulin.
- Family history. A person risk upsurges if his parent or sibling has history of type II diabetes.
- Race. Even though it's uncertain why, people of specific races are at higher risk.
- Age. A person risk upsurges as he/she gets older. This may be because a person has a habit to exercise less, lose muscle mass and add weight as he/she gets older. Nonetheless type II diabetes is likewise growing among children, youths and adults.
- Gestational diabetes. If a person developed gestational diabetes when she was pregnant, her risk of emerging pre-diabetes and type II diabetes far ahead upsurges. If she gives birth to a baby weighing more than 4 kilograms, she is also at risk of type II diabetes.
- Polycystic ovary syndrome. For females, having polycystic ovary syndrome increases the risk of getting diabetes.
- High blood pressure. Having blood pressure more than 140/90 millimeters of mercury (mm Hg) is connected to an augmented risk of type II diabetes.
- Abnormal cholesterol and triglyceride levels. If a person has low levels of highdensity lipoprotein, or good cholesterol, his/her risk of type II diabetes is going to be higher. Triglycerides are additional type of fat passed in the blood. A person with greater levels of triglycerides has an augmented risk of type II diabetes.

A practical approach to this type of problem is the application of regression analysis where past data is better combined into some functions. The result is an equation in which both xj inputs are multiplied by wj ; the sum of all these products is constant, and then output $y = \Sigma$ wj xj +, where j = 0..n.

The problem is the difficulty of choosing an appropriate function to have all the collected data and adjust the output automatically when more information is attained, because the candidate's performance is organized by a number of arguments, and this control will not have any clear regression model. The artificial neural network, which emulates the human thinking in solving a problem, is a more common approach that can address this type of problems. Thus, the attempt to develop an adaptive system such as artificial neural network to predict the situation and classification based on the results of these arguments [14].

3.1 Artificial Neural Network

Adaptive Artificial Neural Network is a non-parametric technique to categorize that in the medical field based on input variables to categorize subjects into healthy or unhealthy. Classification and prediction of the patient's condition based on risk factors are an application of artificial neural networks [12]. Furthermore, ANN is an application of Artificial Intelligence [13]. In artificial neural networks is inspired by the diverse structure of the human brain. Billions of nerve cells (neurons) through the communication that with each other (synapses) creates a biological neural network in the human brain that is devoted to human activities like speaking, reading, comprehension, breathing, face detection, movement, voice recognition, also resolve issues and data storage. Artificial neural networks, in fact, mimic a part of brain jobs [13].

3.2 Artificial neural network structure

Neural networks are nonlinear modeling of intelligent computational methods which recently is considered as an advance in computing and information processing tools acquired a significant and advanced position in the science field, and the consequences have been promising. Feedforward neural networks are valuable type of artificial neural networks, since feedforward neural network with a hidden layer, appropriate activation function in the hidden layer and the sufficient hidden layer neurons are able to estimate any function with an arbitrary accuracy. For this aim, in the following section we present a structure of feedforward neural network modeling to prediction diabetes problem. In general, artificial neural networks have three types of layers as follows:

- Input layer: Get the raw data that has been fed to the network.
- Hidden layers: the function of these layers is determined by inputs, weight, the relationship between them, and the hidden layers. Weights between input and hidden units determine when a hidden unit needs to be activated.
- Output layer: output unit function depending on activity and weight of the hidden unit and the connection between hidden units and output.

3.3 The Back-propagation Training Algorithm

- Initialize each wi to some small random value
- Until the termination condition is met, Do
- For each training example Do
- Input the instance (x1,…,xn) to the network and compute the network outputs ok
- For each output unit k: δk=ok(1-ok)(tk-ok)
- For each hidden unit h: δh=oh(1-oh) Σk wh,k δk
- For each network weight wj Do
- wi,j=wi,j+Δwi,j,where Δwi,j= η δj xi,j and η is the learning rate.

3.4 Previous studies

• The author in [16] used Data Mining to develop a model for classifying diabetic patient control level based on historical medical records. The author was motivated by the death caused by diabetes in the world which necessitated avoiding the complication of the disease. He developed a new predictive model using data mining techniques which would classify diabetic patient control level based on historical medical records. The research was carried out using three data mining techniques which are Naïve Bayes, Logistic and J48. The research was implemented using WEKA application. The result showed that Logistic data mining algorithm gave a precision average of 0.73, recall of 0.744, Fmeasure of 0.653 and accuracy of 74.4%. Naïve Bayes gave a precision average of 0.717, recall of 0.742, F-measure of 0.653 and accuracy of 74.2%. J48 gave a precision average of 0.54, recall of 0.735, F-measure of 0.623 and accuracy of 73.5%. This proved that the logistic algorithm was more accurate than the other two. The research was limited in that only diabetes type 2 was considered. They also did not look into the discovery of appropriate features with minimal effort and validation on discovered features.

• The author in [17] developed a prediction model for diabetes Type II treatment plans by using data mining. The author was motivated by the highly dangerous complication of chronic disease as well as the complication which required amputation of one of the parties. He developed a new model for classifying diabetes type 2 treatment plans which could help the control of blood glucose level of diabetic patient. He made use of J48 algorithm in conducting the experiment on 318 medical records which was collected from JABER ABN ABU ALIZ clinic center for diabetes in Sudan. The basic control information showed that 59.1% of the record was considered for Oral Hypoglycemic, 35.5% for Insulin and 5.3% for Diet. The evaluation was done using the WEKA application. The research work did not consider diabetes type 1 patients which could have been included with additional attributes. Also, the nutrition system and exercise could have been included to increase the accuracy of the system.

• The authors in [18] used prediction of diabetes mellitus based on boosting ensemble modeling. They were motivated by the focus of aiding diabetes patients fit themselves into their normal activities of life by early predicting their state and tacking it. They intended to predict the diabetes types of patients based on physical and clinical information using boosting ensemble technique. They made use of boosting ensemble technique which internally uses random committee classifier. The architecture used was supported by integrating data management, learning, and prediction components together. The evaluation result of the technique showed accuracy gave a weighted average TP rate of 0.81, FP rate of 0.198, Precision of 0.81, Recall of 0.81, F-measure of 0.82 and ROC area of 0.82 for diabetes type 1 and 2. The research work is intended to be extended in future the integration into a cloud based clinical decision support system for chronic diseases and the inclusion of a feedback mechanism to increase the level of satisfaction of users.

• Sernyak used logistic regression analysis to calculate odds ratio neuroleptic unusual version and a diagnosis of diabetes in each of the age groups, control the effects of population, and diagnosis [9]. Thirugnanam has improved diabetes prediction using fuzzy neural networks [10]. Hamid and others have offered hybrid intelligent systems for the detection of micro albuminuria in patients with type 2 diabetes without measuring the urinary albumin [11]. Javad and others proposed the method base on automatic learning on type II diabetes to regulate blood sugar [12].

4. Methodology

By looking intensely through literature and soliciting the experience of human experts on pathological conditions, a number of factors have been recognized that have an impact on determining patients' cases in the subsequent period. These factors were prudently studied and coordinated with an appropriate number for coding the computer within the modeling environment ANN. These factors were categorized as input variables and output variables that reflect some possible levels of disease status in terms of the assessment system. The data were entered into the JNN tool environment, determined the value of each of the variables using JNN(the most influential factor on diabetes), then the data were trained, validated, and tested.

4.1 Input variables

The specified input variables are those that can be obtained simply from the file system and the registry of diseases. Input variables are:

Table 1: attributes in the Data set

No.	Attribute Name	Attribute Meaning
1	Pregnancies	Number of pregnancies
2	PG Concentration	Plasma glucose at 2 hours in an oral glucose tolerance test
3	Diastolic BP	Diastolic Blood Pressure (mm Hg)
4	Tri Fold Thick:	Triceps Skin Fold Thickness (mm)
5	Serum Ins:	2-Hour Serum Insulin (mu U/ml)
6	BMI:	Body Mass Index: (weight in kg/ (height in m)^2)
7	DP Function:	Diabetes Pedigree Function
8	Age:	Age (years)
9	Diabetes:	Whether or not the person diabetes

These factors were converted into a format suitable for neural network analysis as shown in Table2 "data set up to 1004", Input characteristics 8 and one output (0 diabetic, 1 healthy) .

4.2 The Output Variable

The output variable represents whether a person has diabetes or not (Sick, Healthy).

Table 2: Output Data Transformation

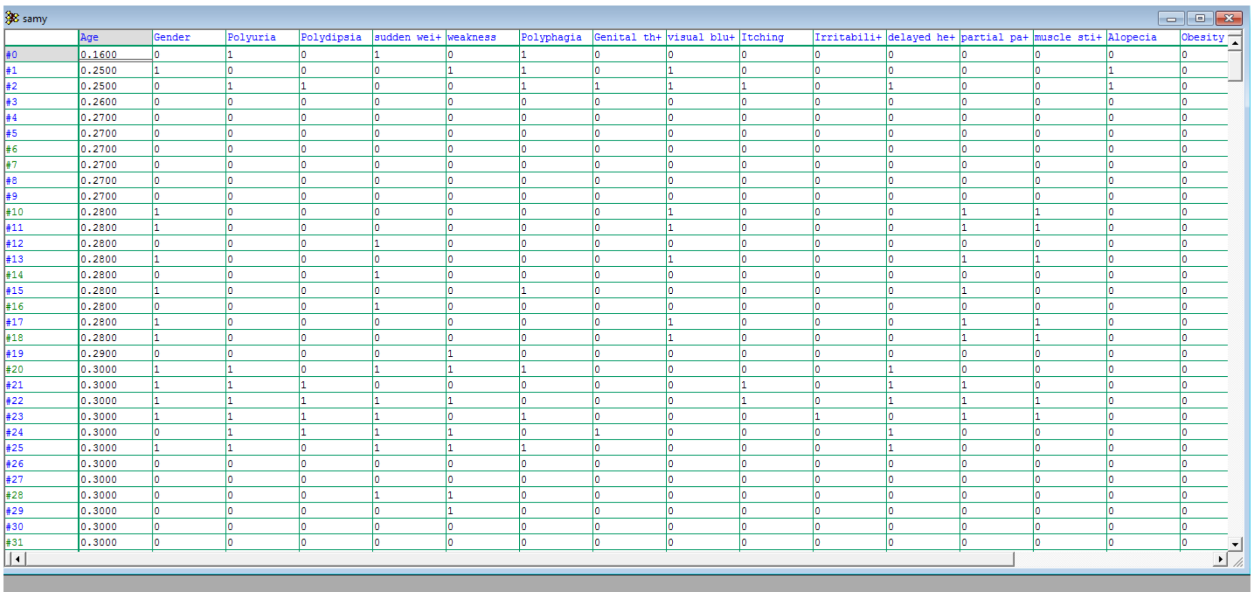
#	Output Variable	Diabetes
1	Healthy "1 "	The person does not have diabetes
2	Sick “0”	The person has diabetes

Table 2 shows the classification of the selected output variable, which is consistent with the classification system, in the identification of disease cases.

4.3. Neural network evaluation

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 10, October - 2020, Pages: 14-22

As mentioned above, the purpose of this experiment was to identify whether or not the person has diabetes. We used Backpropagation algorithm, which provides the ability to perform neural network learning and testing. Our neural network is the front feed network, with one input layer (16 inputs), one hidden layer and one output layer (1 output) as seen in Figure 2. The proposed model is implemented in Just Neural Network (JNN) environment. The dataset for the diagnoses of diabetes were gathered from Early stage diabetes risk prediction dataset in UCI machine learning repository[ ] which contains 520 samples with 17 attributes (as seen in Figure 1). This model was used to determine the value of each of the variables using JNN which they are the most influential factor on diabetes prediction as shown in Figure 3. After training and validating, the network, it was tested using the test data and the following results were obtained. The accuracy of the diabetes predication was (87.3%). The average error was 0.010. The training cycles (number of epochs) were 11,909. The training examples were 347. The number of validating examples was 173 as seen in Figure 4. The control parameter values of the model is shown in Figure 5 and the detail summary of the proposed model is shown in Figure 6.



	Age	Gender	Polyuria	Polydipsia	sudden wei+	weakness	Polyphagia	Genital th+	visual blu+	Itching	Irritabili+	delayed he+	partial pa+	muscle sti+	Alopecia	Obesity
#0	0.1600	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
#1	0.2500	1	0	0	0	1	1	0	1	0	0	0	0	0	1	0
#2	0.2500	0	1	1	0	0	1	1	1	1	0	1	0	0	1	0
#3	0.2600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#4	0.2700	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#5	0.2700	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#6	0.2700	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#7	0.2700	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#8	0.2700	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#9	0.2700	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#10	0.2800	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0
#11	0.2800	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0
#12	0.2800	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
#13	0.2800	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0
#14	0.2800	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
#15	0.2800	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0
#16	0.2800	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
#17	0.2800	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0
#18	0.2800	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0
#19	0.2900	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
#20	0.3000	1	1	0	1	1	1	0	0	0	0	1	0	0	0	0
#21	0.3000	1	1	1	0	0	0	0	0	1	0	1	1	0	0	0
#22	0.3000	1	1	1	1	1	0	0	0	1	0	1	1	1	0	0
#23	0.3000	1	1	1	1	0	1	0	0	0	1	0	1	1	0	0
#24	0.3000	0	1	1	1	1	0	1	0	0	0	1	0	0	0	0
#25	0.3000	1	1	0	1	1	1	0	0	0	0	1	0	0	0	0
#26	0.3000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#27	0.3000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#28	0.3000	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
#29	0.3000	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
#30	0.3000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#31	0.3000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1: Imported data into JNN environment

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 10, October - 2020, Pages: 14-22

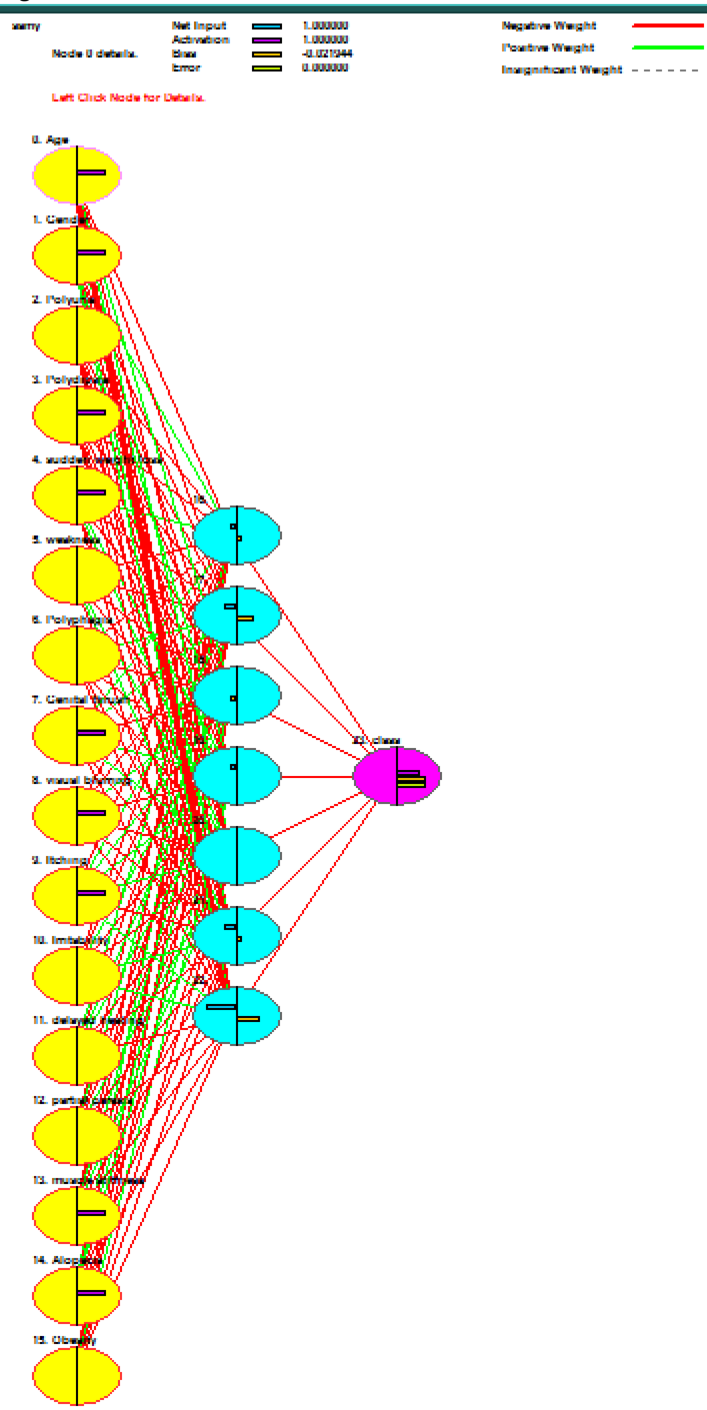
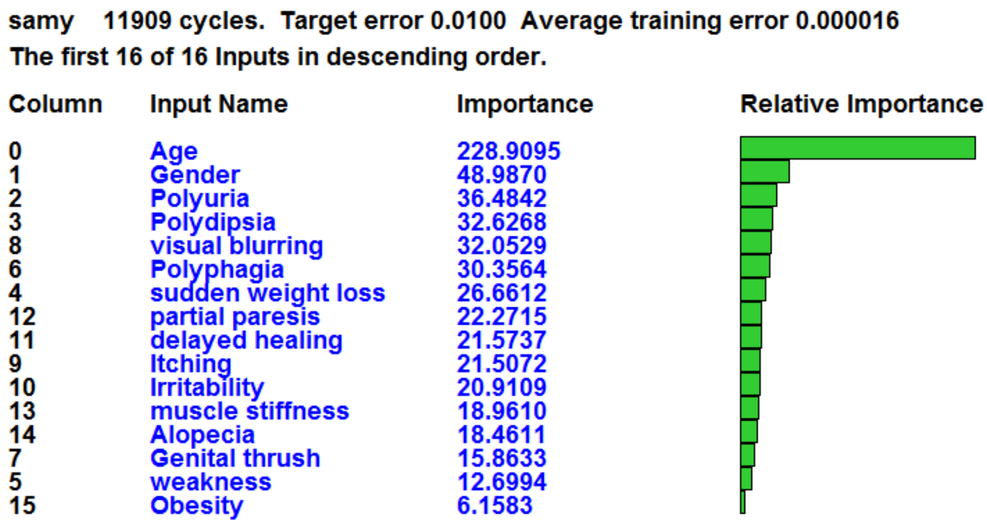


Figure 2: Architecture of ANN model



samy 11909 cycles. Target error 0.0100 Average training error 0.000016
The first 16 of 16 Inputs in descending order.

Column	Input Name	Importance	Relative Importance
0	Age	228.9095	
1	Gender	48.9870	
2	Polyuria	36.4842	
3	Polydipsia	32.6268	
8	visual blurring	32.0529	
6	Polyphagia	30.3564	
4	sudden weight loss	26.6612	
12	partial paresis	22.2715	
11	delayed healing	21.5737	
9	Itching	21.5072	
10	Irritability	20.9109	
13	muscle stiffness	18.9610	
14	Alopecia	18.4611	
7	Genital thrush	15.8633	
5	weakness	12.6994	
15	Obesity	6.1583	

Figure 3: The most influential Features

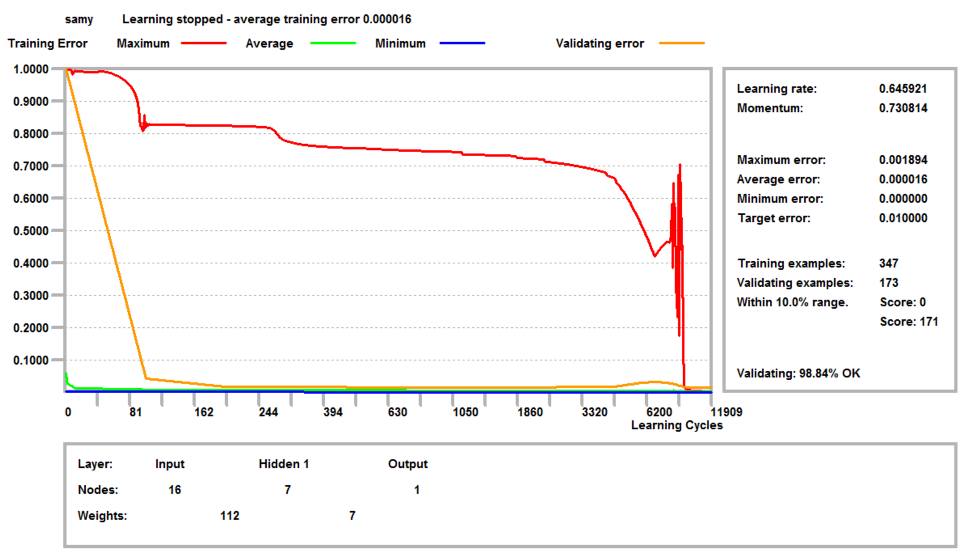


Figure 4: Training and validating of the proposed model

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 10, October - 2020, Pages: 14-22

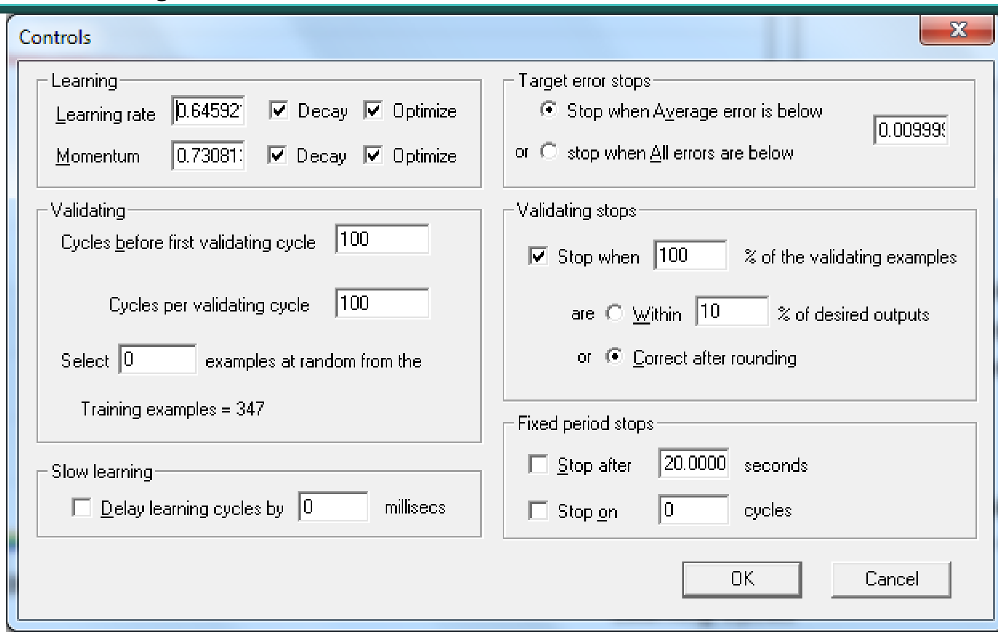


Figure 5: Control parameter of the proposed model

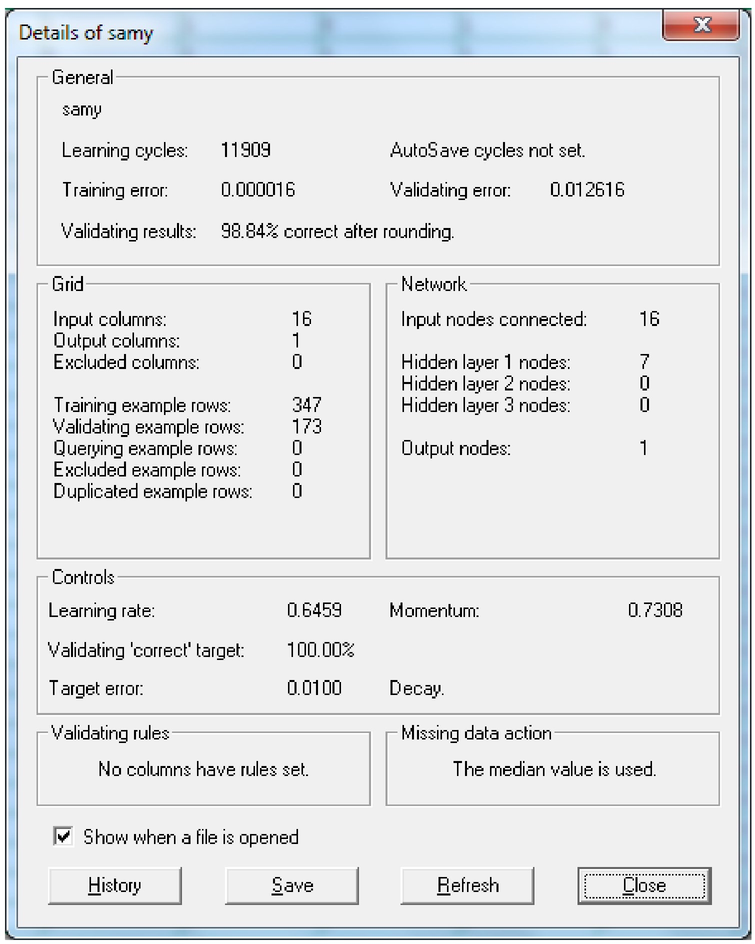


Figure 6: detail Summary of the proposed model

5. Conclusion

In this paper, artificial neural network was used to predict diabetes. Using artificial neural networks model we can design and implement complex medical processes using software. The software systems are more effective and efficient in various medical fields including predicting, diagnosing, treating and helping the surgeons, physicians, and the general population. These systems can be implemented in a parallel way and are distributed in different measures. In general, artificial neural network is a parallel processing system that is used to detect complex patterns in the data. The aim of this study was to determine the effective variables and their impact on diabetes. The proposed model was implemented in JNN environment. The diabetes dataset contains 520 samples with 17 attributes. This model was first used to determine the value of each of the variables using JNN (the most influential factor on diabetes). After training, validating, and testing the dataset, we got (98.84%) accuracy, average error was (0.010), number of epochs was (11,909), number of training examples was (347), and number of validating examples was (173).

References

1. World Health Organization (WHO), "Definition, Diagnosis, and classification of diabetes mellitus and its complications", part 1. WHO/NCD/NCS/2016.2, (2016).
2. H. Temurtas, N. Yumusak and F. Temurtas, "A comparative study on diabetes disease diagnosis using neural networks", Expert System, vol. 36, (2009), pp. 8610–15.
3. A. Chavey, M. Kioon and D. Bailbé, "programming of beta-cell disorders and intergenerational risk of type 2 diabetes Diabetes", Maternal Diabetes, vol.40, no.5, (2014), pp. 323-30.
4. D. Manzella, R. Grella, A.M. Abbatecola and G. Paolisso, "Repaglinide Administration Improves Brachial Reactivity in Type 2 Diabetic Patients", Diabetes Care, Vol. 28, (2005), pp. 366– 71.
5. E. I. Mohamed, R. Linde, G. Perriello, N. Di Daniele, S. J. Pöppl and A. De Lorenzo, "Predicting type 2 diabetes using an electronic nose-based artificial neural network analysis", Diabetes nutrition & metabolism Vol.15, No.4, (2002). pp. 222-215.
6. Halland, C. Igel, and S. Alstrup, “High-school dropout prediction using machine learning: a Danish largescale study,” in Proceedings of the ESANN 2015, M. Verleysen, Ed., pp. 319–324, i6doc.com, Bruges, Belgium, 2015.
7. Lakkaraju, E. Aguiar, C. Shan et al., “A machine learning framework to identify students at risk of adverse academic outcomes,” in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’15, ACM, New York, NY, USA, 2015.
8. K. Ahmadi, Guideline &book review. The internal (endocrine and lung). Ahmadi Cultural Institute, (2009).
9. A. Morteza et al., "Inconsistency in albuminuria predictors in type 2 diabetes: a comparison between neural network and conditional logistic regression", Translational Research, vol. 161, No.5, (2013), pp. 397-405.
10. M. J. Sernyak et al., "Association of diabetes mellitus with use of atypical neuroleptics in the treatment of schizophrenia", American Journal of Psychiatry, (2014).
11. M. Thirugnanam et al., "Improving the Prediction Rate of Diabetes Diagnosis Using Fuzzy, Neural Network, Case Based (FNC) Approach."Procedia Engineering, Vol.38, (2012). pp. 1709-118,.
12. H. R. Marateb et al., "A hybrid intelligent system for diagnosing microalbuminuria in type 2", (2014). pp. 34-42,.
13. J. A. Torkestani and G. P. Elham, "A learning automata-based blood glucose regulation mechanism in type 2 diabetes", Control Engineering Practice, Vol. 26, (2014). pp. 151-159.
14. D. Livingstone and N. J. Totowa, “Artificial Neural Networks Methods and Application. 1th ed. Totowa, NJ: Hummana Press”, (2008).
15. R. A. Dunne, Wiley, J., Inc, S.," A Statistical Approach to Neural Networks for Pattern Recognition", New Jersey: John Wiley & Sons Inc; (2007).
16. Pima Indians Diabetes DataBase, Data Obtained From: http://www.liacc.up.pt/ML/statlog/datasets/diabetes/diabetes.doc.html
17. T. M. Ahmed, Using Data Mining To Develop Model For Classifying Diabetic Patient Control Level Based On Historical Medical Records. Journal of Theoretical and Applied Information Technology, Vol. 87 no. (2), (2016), pp. 316-350.
18. R. Ali, M. H. Siddiqi, M. Idris, B. H. Kang and S. Lee, “Prediction of diabetes mellitus based on boosting ensemble modeling”. In International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 25-28). Springer International Publishing, (2014).