DOI: http://dx.doi.org/10.26483/ijarcs.v13i6.6916

Volume 13, No. 6, November-December 2022

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

AN ANN MODEL FOR EARLY PREDICTION OF DIABETES

Amit Mishra
Assistant Professor
Baze University, Nigeria

Abstract: Diabetes a metabolic disease with the botanical name diabetes mellitus is diagnosed in a person who has high sugar levels in the bloodstream which could be either because of the cells not reacting to the insulin that is created or the pancreas does not deliver insulin by any stretch of the imagination. According to research (World Health Organization, Geneva 2014),Worldwide 194 million people have already been diagnosed with diabetes and this rate is expanding quickly and is estimated to achieve 333 million by 2025. In Africa, more than 5 million people are already diagnosed with diabetes, and it has been estimated that by 2025 diabetes patients in the continent would be up to 15 million. (KMV Narayan).Ogbera also reported that in Nigeria patients confirmed diabetes up to 158 million, and as such, the need to study the prediction of diabetes is very paramount.In this way, there is an incredible need to concentrate on the prediction of diabetes. So that precautions can be taken to control this deadly disease. This research work wishes to present the prediction of diabetes using ANN. The model used in this work considered a group of fifteen factors and identified the factors that are very influential in the diagnosis of diabetes using regression analysis so as to achieve better accuracy of prediction.

Keyword: Diabetes, ANN, Prediction, regression analysis

1. INTRODUCTION

Diabetes a metabolic disease with the botanical name diabetes mellitus is diagnosed in a person who has high sugar levels in the bloodstream which could be either a result of the cells not responding to the insulin that is produced or the pancreas does not produce insulin at all. Over 15.1 million people in 2000 have been diagnosed with type 2 diabetes and this number is estimated to get to 32.3 million by 2025. In 2007, 214 million Africans constitution 6.4% of the total population in the continent were diagnosed with type 2 diabetes. Also, WHO in 2014 reported that an estimated number of 314million i.e., 8.2% of the world population are already in the pre-diabetes stage of which at least one-third would progress to the diabetes stage in a space of ten years. More than 5 million Nigerians are diabetic (Uloko et. al., 2011) of which 90-95% of such cases were adults diagnosed with type 2 diabetes, this type of diabetes affects men and women in the almost same proportion; 2 million men and 1.5 million women are diagnosed,as diabetic. Ojeme (2015) stated in 2015 Health Reform Foundation of Nigeria (HEFRON) declared diabetes as a major cause of death arising from non-communicable diseases with statistics showing death as a result of diabetes accounted for up to 52% of death in that year. As shown in the table below.

TABLE 1 POPULATION OF DIABETES AROUND THE GLOBE

Population of Diabetes People Around the Globe	Years	Rate
Globally, the rate of type 2 diabetes	2000	15.1 million
Estimated increase in number of diabetic patients worldwide	2025	32.3 million
Death through diabetesin Nigeria will be projected to increase	2025	52%

TABLE 2 POPULATION OF DIABETES IN DIFFERENT CONTINENT

Population of Diabetes In the Continent	Years	Rate
world population inpre-diabetes stage	2004	314 million (8.2%)
In African countries	2007	21.4 million or 6.4%
In Nigeria	2011	5 million
Adults with type 2 diabetes.	2011	90 to 95%

TABLE 3 DEATH CAUSED BY DIABETES

Projected Death Cause by Diabetes Gender Base	Years	Gender	Rate
Diabetes in Nigeria	2015	Men	2 million
Diabetes in Nigeria	2015	Women	1.5 million

(Diabetes Mellitus) have three noteworthy causes which are utilized to order them into sorts. The first (Type 1 Diabetes) is because of the powerlessness of the pancreas to deliver insulin which would require licenses with this kind of diabetes to take reliably take insulin shots of wear an insulin siphon and therefore, this structure was at first alluded to as Insulin Dependent Diabetes Mellitus (IDDM) or adolescent diabetes. the subsequent kind (Type 2 Diabetes) is because of a cell issue where cells neglect to utilize the delivered insulin viably and this might be once in a while joined with an outright insulin inadequacy, this structure was at first alluded to as non-insulin dependent diabetes mellitus (NIDDM) or grown-up beginning diabetes. the last one (Type 3 Diabetes), which happens for

the most part in pregnant ladies without a past record of diabetes, who at that point create diabetes because of high glucose levels is called growth diabetes(Liszka,1999), this sort of anyway may prompt the improvement of the sort 2 diabetes mellitusNahla et. al. (2010). Artificial Neural Network has encountered a blast of enthusiasm in the course of the most recent couple of years and has been effectively utilized over a wide scope of issue spaces in assorted territories, for example, account, drug, building, geography and material science. Numerous endeavors have been made to appropriately characterize neural networks. A neural network is a framework made out of numerous straightforward handling component tasks in parallel whose capacity is dictated by network structure, association quality,and the preparing performed at registering component or hub. DARPA Neural Network study (Lippmann, 1987). A neural network is an enormously parallel disseminated processor that has a characteristic property for putting away experience information and making it accessible for use. It takes after the cerebrum in two regards.

2. LITERATURE REVIEW

Recognizable proof of Influential Parameters for Diagnosis of Diabetes utilizing Artificial Neural Network. Diabetes Mellitus, in direct terms is called diabetes, is a metabolic contamination, where an individual is influenced with high blood glucose level. Diabetes is a metabolic issue caused because of the disappointment of body to create insulin or to really utilize insulin. This condition develops when the body does not convey enough insulin, or in light of the fact that the cells don't react to the insulin that is created.

TYPE 1: Diabetes

Type 1 DM additionally alluded to as insulin-dependent diabetes mellitus (IDDM) results from the body's powerlessness to create insulin. Type 1 diabetes is depicted by loss of the insulin-creating beta cells of the islets of Langerhans in the pancreas, prompting insulin inadequacy. This sort can be additionally delegated safe intervenes or idiopathic. The regular of sort 1 diabetes is of the invulnerable intervened condition. In which beta cell misfortune is a T-cell intervened immune system assault. There is no known preventive measure against sort 1 diabetes Most affected people are commonly strong and of a sound weight when beginning occurs. Most influenced people are commonly strong and of sound weight when beginning occurs. Affectability and responsiveness insulin are typically ordinary, especially in the beginning times. Type 1diabetes can impact kids or adults, yet most of the diabetes cases were youthful stars.

TYPE 2: Diabetes.

TYPE 2 DM is in like manner perceived as noninsulin-dependent diabetes mellitus (NIDDM) is because of insulin opposition, a condition where cells fails to use insulin fittingly, at some point joined with an outright insulin insufficiency. Type 2 diabetes mellitus is depicted by insulin opposition, which might be consolidated date with moderately lessened insulin dischargeDerouich, (2002). The viable explicit imperfections are not known. Diabetes mellitus cases because of a realized flaw are requested exclusively. Type 2 diabetes is the most widely recognized sort.

Gestation Diabetes

Gestation Diabetes mellitus (GDDM) take after sort 2 diabetes in few respects, including a blend of generally lacking insulin release and responsiveness. It takes in about 2% to 5% of all pregnant and may improve or disappear after convey. Around 20% to half of influenced ladies create type 2 diabetes sometime down the road.

In any case, it might be transient. Untreated growth diabetes can harm the soundness of the baby or mother. Hazard to the children incorporates high weight, natural cardiovascular and focal sensory system oddities, and skeletal muscle deformations.

Prediabetes

Prediabetes is a neither condition that happens when an individual's blood glucose level is higher than nor mal however not sufficiently high for finding of sort 2 diabetes.

Artificial Neural Network

An Artificial Neural Network (ANN) is an information preparing model that is enlivened in the manner organic sensory system, for example, the cerebrum, process information. The key component of this model is the novel structure of the information preparing framework. Artificial neural networks (ANN) have been created as speculations of numerical models of natural sensory systems. A first rush of enthusiasm for neural networks (otherwise called connectionist models or parallel appropriated handling) developed after the presentation of streamlined neurons by McCulloch and Pitts (1943). The premise preparing components of neural networks are called artificial neurons, or basically neurons or hubs.

The word directed begins from the way that the ideal flag on individual yield hubs are given by an outside educator. The best-realized system happens in the back-proliferation calculation, the delta rule, and the discernment rule. In unsupervised learning (or self-association), a (yield) unit is prepared to react to bunches of examples inside the information. In this worldview, the framework should find measurably notable highlights of the info populace. In contrast to the managed learning worldview, there is no from the earlier arrangement of classifications into which the examples are to be grouped; rather, the framework must build up its very own portrayal of the info upgrades.

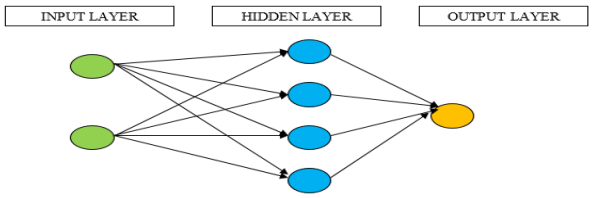

Figure 1: Sample of Neural Network Architecture

Reinforcement learning is learning what to do, and how to map situations to actions so as to maximize a numerical reward signal.

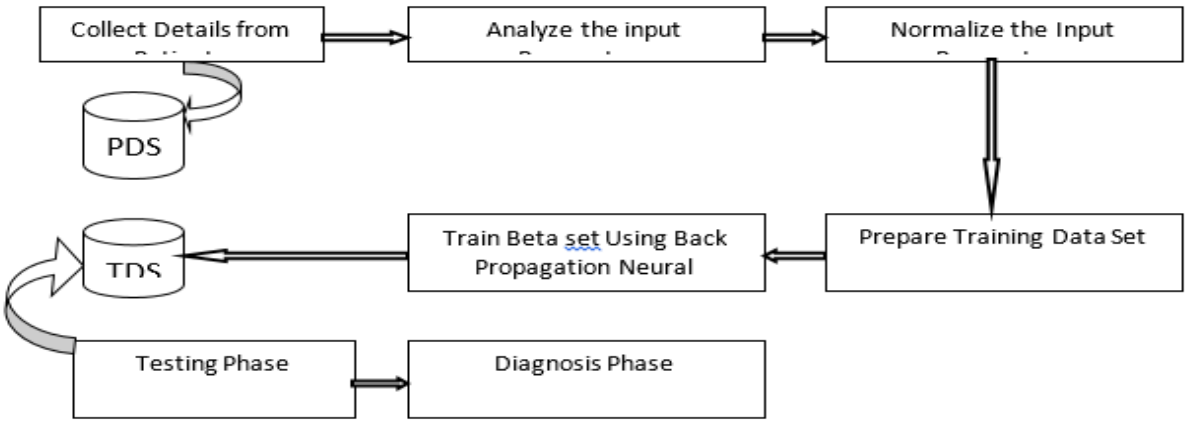

Figure 2: Proposed Framework

3. MATERIALS and METHODOLOGY

Back Propagation Algorithm

Back propagation is a deliberate technique for preparing multi-layer artificial neural network. It is a multi-layer feed forward network that utilizations degree angle plummet based back propagation rule. During the preparation arrange, the preparation information is encouraged to info layer. The information is spread through the shrouded layer and afterward to the yield layer, which is called forward pass.

Thisback-propagation algorithm is appropriate for forecasting.

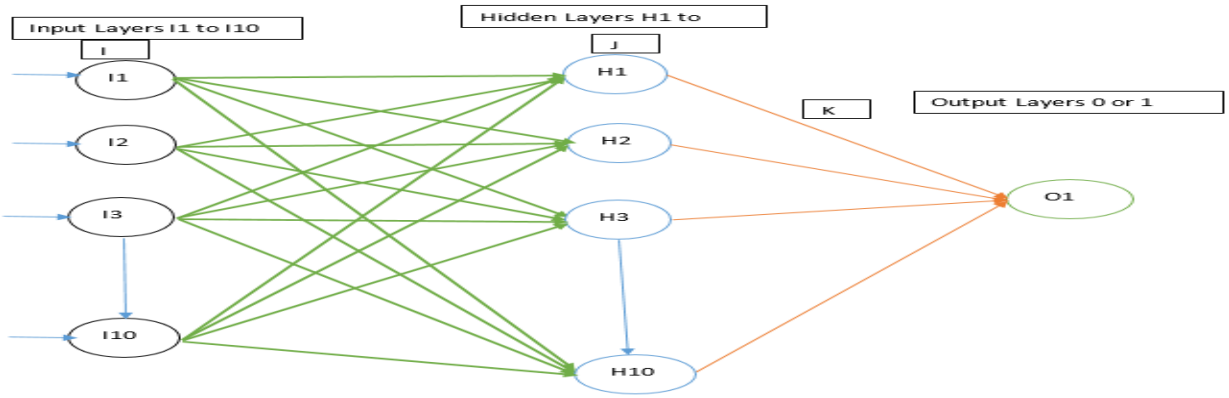

Figure 3:The network Architecture for Training

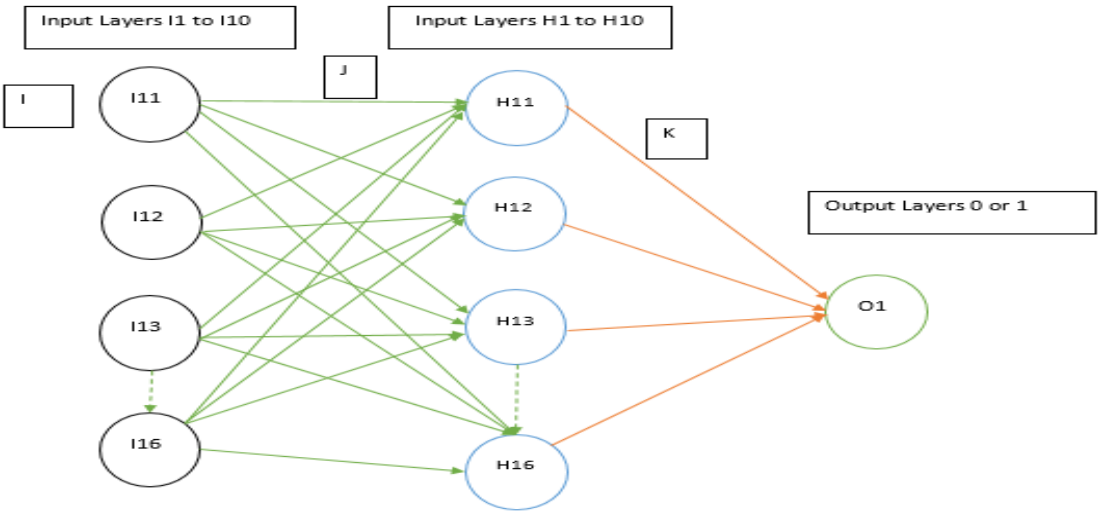

Figure 4: The Network Architecture for Testing

3.1. Data Set:

The data set taken in this study was collected from the local hospitals in Minna, Niger state, Nigeria. The main reasonof using this data set is to predict diabetes in the patient sed on some measures included in the dataset. During the research, many limitations were also found

while selecting the reoccurrence of the from the bigger dataset. The type of dataset and problem is a classic supervised binary classification, (2010).

TABLE 4: COLLECTION OF PATIENT INFORMATION

The information gathering contains fifteen factors, the components estimated are as per the following:

1.	Age	(<45 years)
2.	Family background anybody with diabetes	(Yes/No)
3.	Medication for high blood pressure	(Yes/No)
4.	High blood glucose during illness	(Yes/No)
5.	Smoking or any tobacco items	(Yes/No)
6.	Vegetable or organic product admission consistently)	(Everyday/Not
7.	Physical exercises 30 minutes every day	(Yes/No)
8.	Waist hips proportion	(<0.8)
9.	Increased urination, hunger and thirst	(Yes/No)
10.	Poor injury healing	(Yes/No)
11.	Lifestyle inactive work	(Yes/No)
12.	Frequent intake of non-vegetarian food (more than two times per week)	(Yes/No)
13.	Itching everywhere throughout the body	(Yes/No)
14.	Patients' blood glucose (GRBS) in typical range	(82-110mg/dl)
15.	Body mass index (BMI) weight/height^2	(<24)

TABLE 5:SAMPLE OF DATA COLLECTED FROM THE PATIENT

SN	Parameter	Data
1.	Age (<45 years)	44
2.	Family background-anyone with diabetes(yes/no)	NO
3.	Medication for high blood pressure (yes/No)	NO
4.	High blood glucose during illness(yes/no)	NO
5.	Smoking or any tobacco products (yes/no)	Not Every Day
6.	Vegetable or fruit intake (everyday/not every day)	NO
7.	Physical activities 30 minutes daily (yes/no)	24-29
8.	Waist hip ratio (<0.8)	NO
9.	Increased urination, hunger and thirst (yes/no)	NO
10.	Poor wound healing (yes/no)	NO
11.	Lifestyle-sedentary work (yes/no)	Sedentary work
12.	Frequent intake of non-vegetarian food (more than twice a week) yes/no	NO
13.	Itching all over the body (yes/no)	NO
14.	Patients' blood glucose (GRBS) in normal range (82-110mg/dl)	96 mg/dl
15.	Body mass index (BMI) weight/height^2 (<24)	0.8

TABLE 6: NORMALIZED DATA

SN	Parameter	Data
1.	Age (<45 years)	1
2.	Family background-anyone with diabetes(yes/no)	0
3.	Medication for high blood pressure (yes/No)	0
4.	High blood glucose during illness(yes/no)	0
5.	Smoking or any tobacco products (yes/no)	0
6.	Vegetable or fruit intake (everyday/not every day)	0
7.	Physical activities 30 minutes daily (yes/no)	0
8.	Waist hip ratio (<0.8)	1
9.	Increased urination, hunger and thirst (yes/no)	1
10.	Poor wound healing (yes/no)	1
11.	Lifestyle-sedentary work (yes/no)	0
12.	Frequent intake of non-vegetarian food (more than twice a week) yes/no	0
13.	Itching all over the body (yes/no)	0
14.	Patients' blood glucose (GRBS) in normal range (82-110mg/dl)	1
15.	Body mass index (BMI) weight/height^2 (<24)	0

TABLE 7: SAMPLE OF DATA USED FOR BACK PROPAGATION

SN	Male/female	(x)	X(i)
1.	Age(<45 years)	25	1.66667
2.	Family background-anyone with diabetes	17	1.13333
3.	Medication for high blood pressure	20	1.33333
4.	High blood glucose during illness	5	0.33333
5.	Smoking or any tobacco products	16	0.46667
6.	Vegetable or fruit intake(everyday/not every day)	0	0
7.	Physical activities 30 minutes daily	5	0.33333
8.	Waist hip ratio (<0.8)	20	1.33333
9.	Increased urination, hunger and thirst	7	0.46667
10.	Poor wound healing	6	0.4
11.	Lifestyle-sedentary work	22	1.46667
12.	Frequent intake of non-vegetarian food (more than twice a week)	23	1.53333
13.	Itching all over the body (yes/no)	9	0.6
14.	Patients' blood glucose (GRBS) in normal range (82-110mg/dl)	14	0.93333
15.	Body mass index (BMI) weight/height^2 (<24)	18	1.2

4. Analysis and Data Presentation

Back Propagation Algorithm in Artificial Neural Network

Back Propagation Algorithm in Artificial Neural Network (ANN) was used for the identification of influential parameter for diagnosis of diabetes to predict the high-level performance for prediction of prediabetes.

Different parameter value was use in training the neural and to evaluate the person for diagnosis prediabetes stage to diabetes stage.

Data used for this study was collected by gathering of information from patients for the data group, which was then followed by analyses of the information gathered, this analysis involved normalizing the gathered data. AAN was applied to normalize the data for back propagation algorithm, followed by the generation of graphs and finally identifying factors that are very influential in the diagnosis of diabetes.

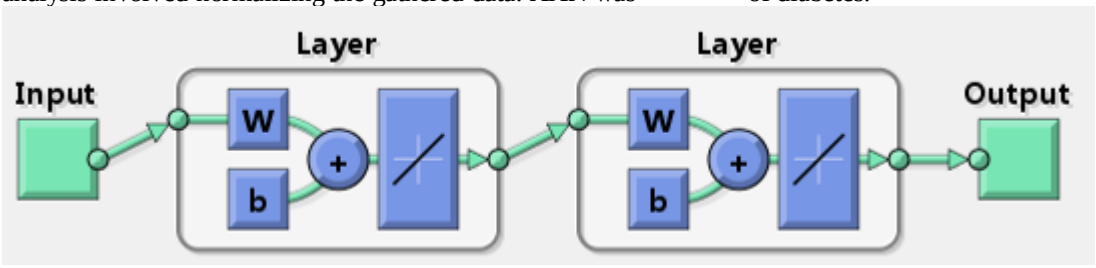

Figure 4: Network Design

Figure 4 is used for the full data set, training data set and testing data set.

TABLE 8: TRAINING DATA USED FOR BACK PROPAGATION

SN	Male/female	(x)	X(i)
1.	Age (<45 years)	12	0.70588
2.	Family background-anyone with diabetes	8	0.47058
3.	Medication for high blood pressure	10	0.58832
4.	High blood glucose during illness	3	0.17647
5.	Smoking or any tobacco products	8	0.47058
6.	Vegetable or fruit intake (everyday/not every day)	0	0
7.	Physical activities 30 minutes daily	5	0.29411
8.	Waist hip ratio (<0.8)	10	0.58823
9.	Increased urination, hunger and thirst	3	0.17647
10.	Poor wound healing	7	0.41176
11.	Lifestyle-sedentary work	11	0.64708
12.	Frequent intake of non-vegetarian food (more than twice a week)	11	0.64708
13.	Itching all over the body (yes/no)	4	0.2352
14.	Patients' blood glucose (GRBS) in normal range (82-110mg/dl)	7	0.41176
15.	Body mass index (BMI) weight/height^2 (<24)	9	0.52941

5. Generation the graphs

Generation of Graphs

Total possibility will be plotted by the use of scatter plot graph to indicate the connection or relationship existing between two variables. With the aid of these graphs, it will be determined the direction between input variable and the output. The direction will help in prediction on the influential parameters.

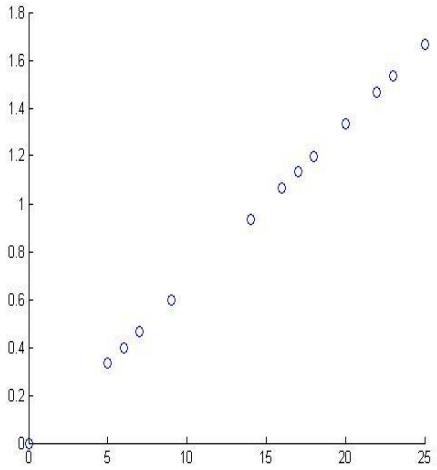

Figure 5: Scatter Graph for Prediction

The most influential factors were determined by concluding data derived from the scatter plots graphs so as to show the best fit point in figure 5. These factors were determined to be:

I.Age
II.Medication of High Blood Pressure
III.Physical Activity for 30 minutes Daily
IV.Increased Urination, Hunger, and thirst

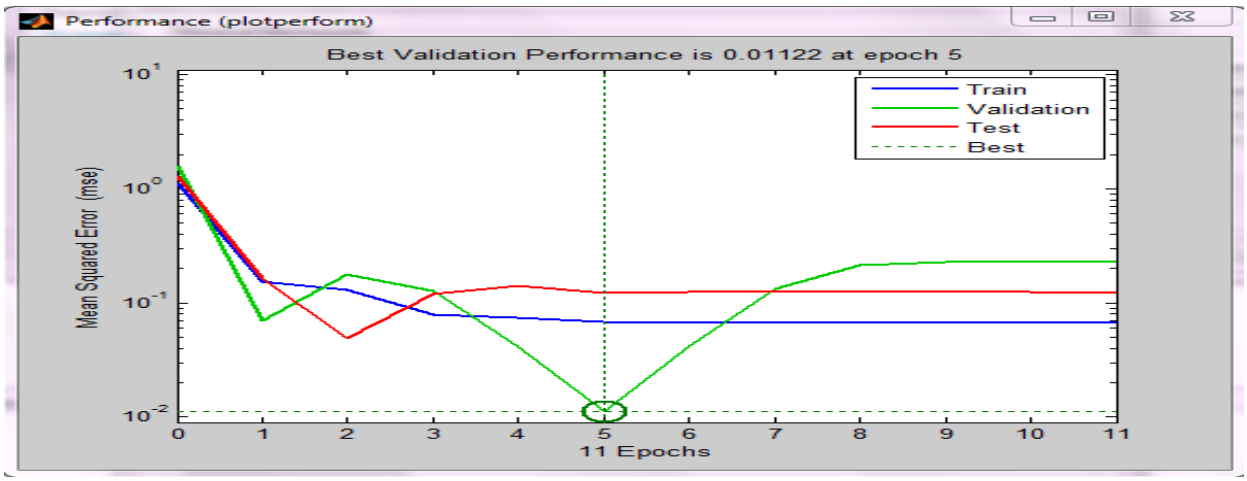

Figure 6: performance of Back Propagation Algorithm

Figure 6 the graph simulates the performance of back propagation, the training, testing, validation and the best for both male and female population X against the number of parameters X(i)

The graph above shows the rate of change of training and testing with the data feed on to the ANN.

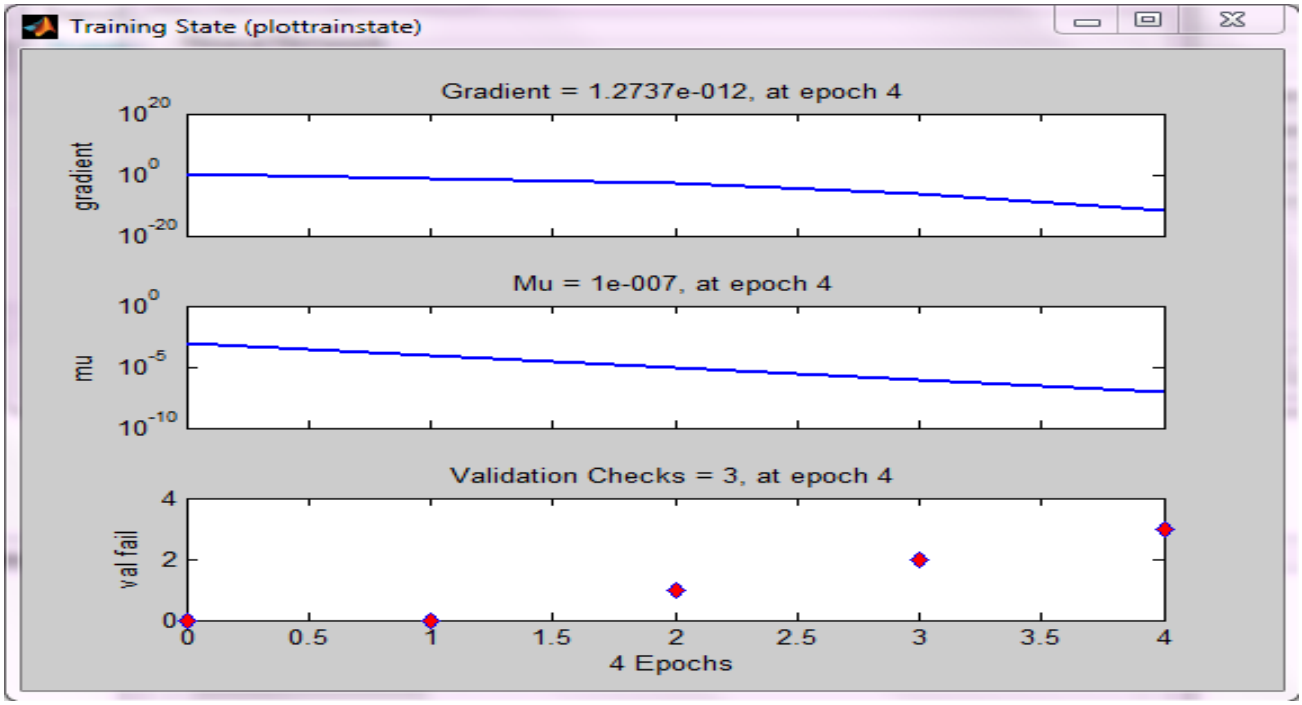

Figure7:Training State

Figure 7 shows thetraining statestageswhich the ANN pass through.

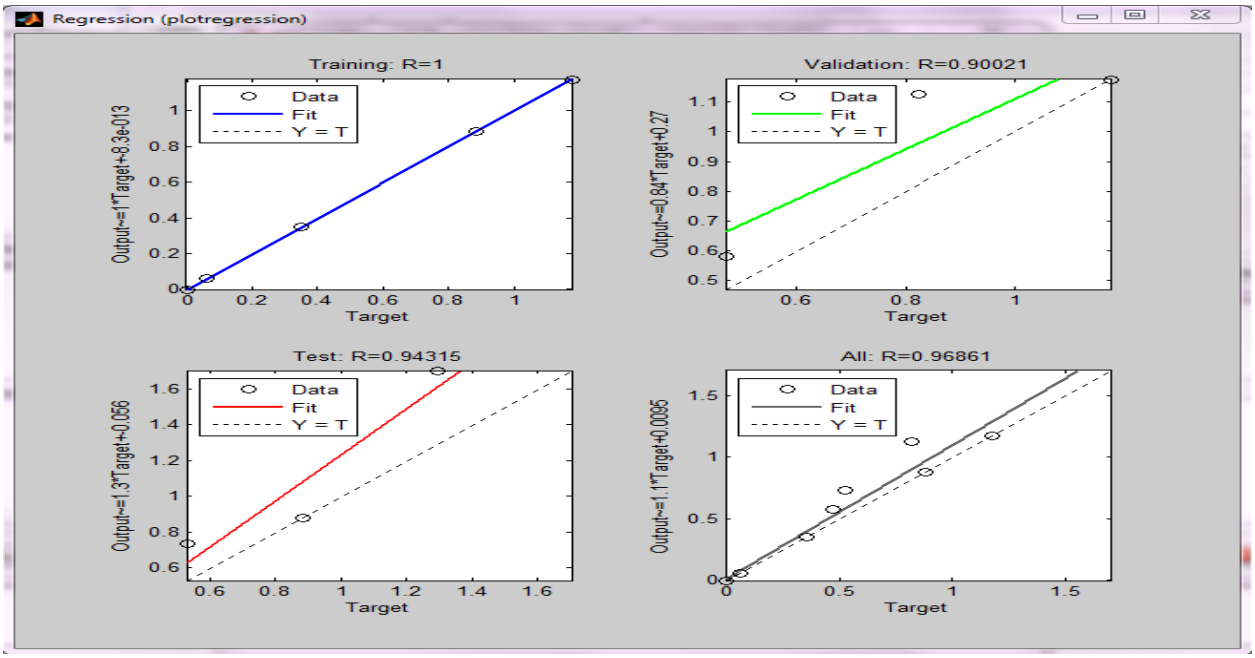

Figure 8: Evaluating the performance using Regression Algorithm

Regression graphs in figure 8 shows the best fit which correlates between the training and testing using thick line, while the dotted line shows perfect fit produced by back propagation algorithm.

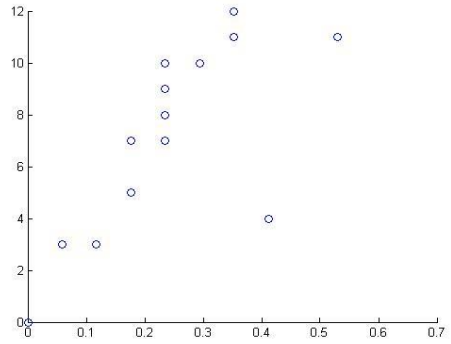

Figure 9: Scatter Graph for Prediction during Training

Figure 9by concluding the data from the scatter plots graph it extract the best fit points and, and get the most influential parameters as follow:

I.Age
II.Medication of High Blood Pressure
III.Physical Activity for 30 minutes Daily
IV.Increased Urination, Hunger, and thirst

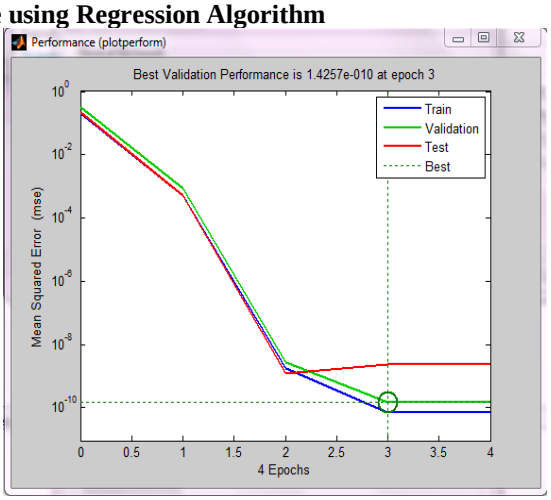

Figure 10: Training performance of Back Propagation Algorithm

As shown in figure 10, the graph simulates the performance of back propagation, the training, testing, validation and the best for both male and female population X against the number of parameters X(i). The graph above shows the rate of change of training and testing with the data feed on to the ANN.

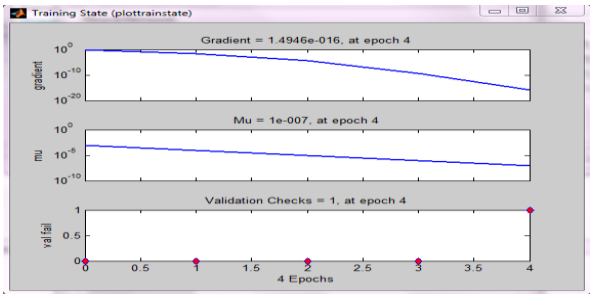

Figure11:Training State for Training

Figure 11 shows the training state stages which the ANN pass through.

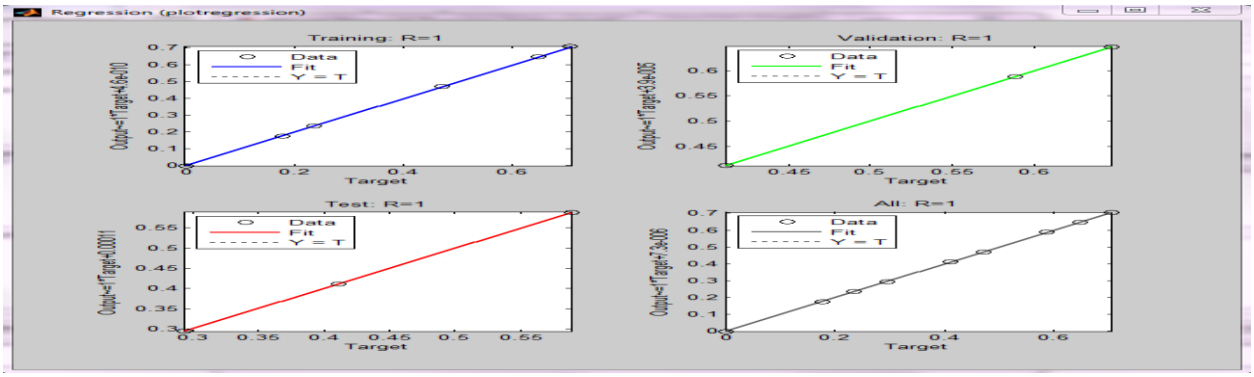

Figure 12: Training Evaluating the Performance using Regression Algorithm

Regression graphs in figure 12 shows the best fit which correlates between the training and testing using thick line, while the dotted lines shows perfect fit produced by back propagation algorithm.

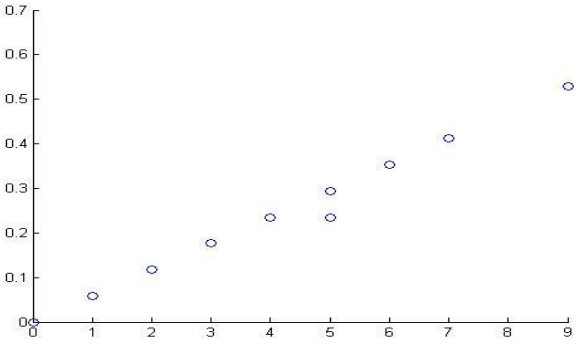

Figure 13: Scatter Graph for Prediction during Testing

By concluding data derived from the scatter plots graphs in figure 13 so as to show the best fit point, the most influential factors were determined, these factors include:

I. Age
II. Medication of High Blood Pressure
III. Physical Activity for 30 minutes Daily
IV. Increased Urination, Hunger, and thirst

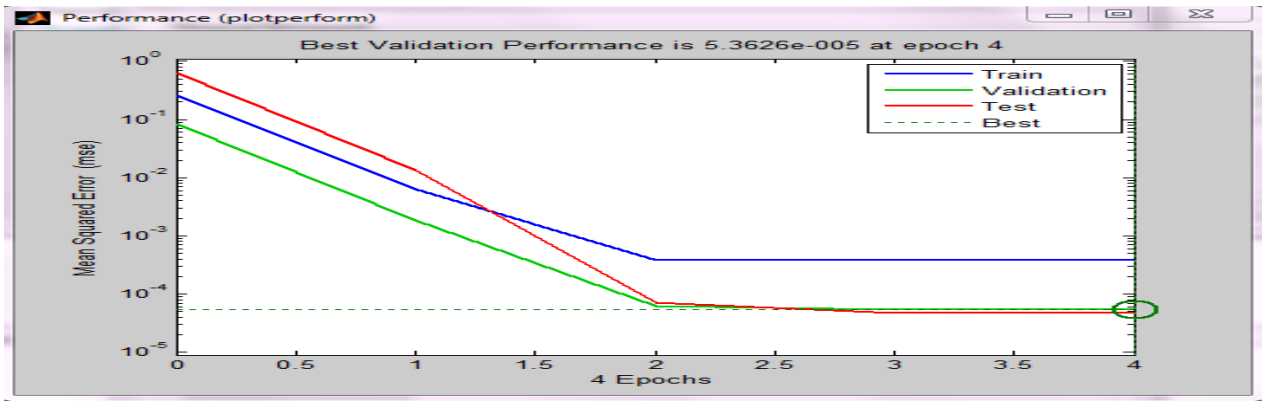

Figure 14 Testing Performance of Back Propagation Algorithm

Figure 14 the graph simulates the performance of back propagation, the training, testing, validation and the best for both male and female population X against the number of parameters X(i).

The graph above shows the rate of change of training and testing with the data feed on to the ANN.

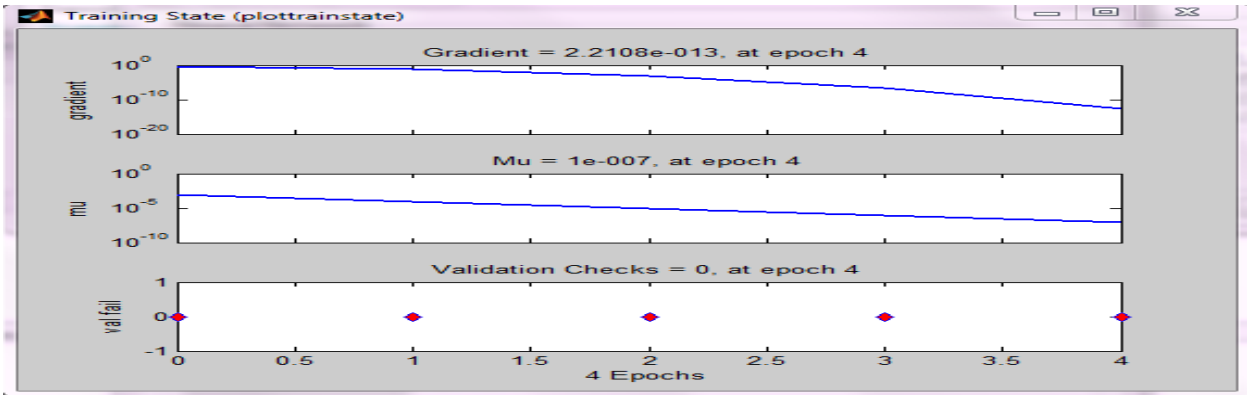

Figure15: Training State for Training

Figure 15 shows the training state stages which the ANN pass through.

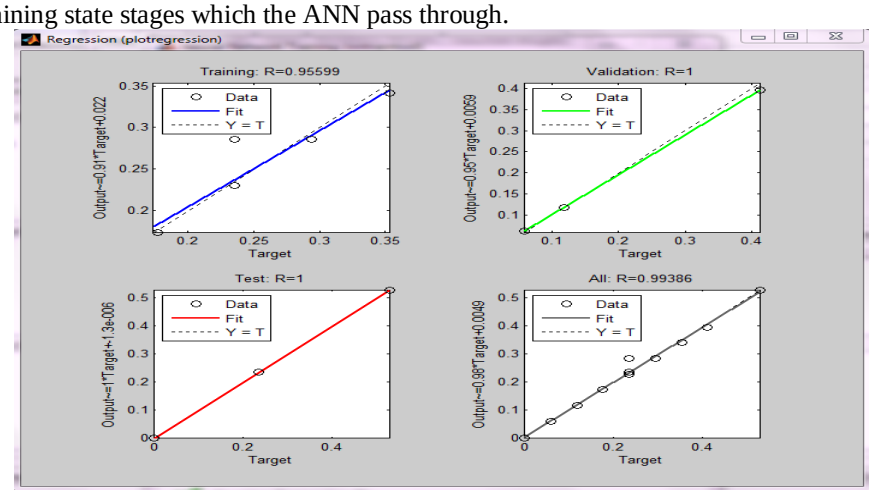

Figure 16: Training Evaluating the Performance using Regression Algorithm

Regression graphs in figure 16 above shows the best fit which correlates between the training and testing using thick line, while the dotted lines show perfect fit produced by back propagation algorithm

5. DISCUSSION OF RESULT, RECOMMENDATION AND CONCLUSION

Recommendation

In developed nations rational measures have been adopted so as to effectively curb the trend of diabetes. However, worldwide this trend has been on the rise as most developing nations (including Nigeria) have been pleased at just accepting the growing international recommendations, while little or no effort has been put in place to address the problem. This behaviour could be traced to the lack of political will and inadequate financial allocation to the healthcare sector.

Illustrations derived from this model affirms that diagnosis and recommendations made by specialist and experts in the field of metabolic diseases such as diabetes and health megamenu as a whole should be strictly adhered to curb such trend.

Individuals should participate actively by developing healthy practices which include avoiding tobacco, inactivity, alcohol, obesity and controlling cholesterol levels.

Conclusion

In this study, a proposed model technique which could be to achieve better accuracy of prediction of diabetes.

An AAN model was applied to the most common factors used in predicting diabetes using back propagation to the data groups.

This research work is to improve the accuracy of early detection of diabetes with will help the patient take serious percussion in advance.

6. REFERENCE

[1] Derouich M, Boutayeb A. (2002) the effect of physical exercise on the dynamics of glucose and Insulin.Journal of Biomechanics.

[2] Geneva, The world health report today's challenges. W orld Health Organization.http://www.who.int/whr/2014 /en.

[3] Lippmann, R.P. (1987). An introduction to computing with neural networks. IEEE Accost. SpeechSignal Process. Mag.

[4] LiszkaHackzell, J. J. (April 1999) Prediction of blood glucose levels in diabetic patients using ahybrid AI technique. Computers and Biomedical Research London.

[5] Nahla H. Barakat, Andrew P. Bradley, Senior Member, IEEE, and Mohamed Nabil H. Barakat(2010) Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitusl, IEEETransactions on Information Technology in Biomedicine, VOL. 14.

[6] T.Jayalakshmi and A.Santhakumaran, (2010) A novel classification method for classificationof diabetes

mellitus using artificial neural networks. International Conference on DataStorage and Data Engineering.
[7] Uloko, A.E., Musa, B.M., Ramalan, M.A. et al. Prevalence and Risk Factors for Diabetes Mellitus in Nigeria: A Systematic Review and Meta-Analysis. Diabetes Ther 9, 1307–1316 (2018). https://doi.org/10.1007/s13300-018-0441-1
[8] Uloko, A.E., Musa, B.M., Ramalan, M.A. et al. Prevalence and Risk Factors for Diabetes Mellitus in Nigeria: A SystematicReview and Meta-Analysis. Diabetes Ther 9, 1307–1316 (2018). https://doi.org/10.1007/s13300-018-0441-1
[9] Wilding P, Morgan M, Grygotis A, Shoffner M, Rosato E. (1994) Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. Cancer Lett.
[10] McCulloch, W.S., Pitts, W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5, 115–133 (1943). https://doi.org/10.1007/BF02478259

