VOL. 11, NO. 3, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A PREDICTION MODEL FOR TYPE 2 DIABETES RISK AMONG INDIAN WOMEN

Alby S.[1] and B. L. Shivakumar[2]
[1]Research and Development Centre, Bharathiar University, Coimbatore, India
[2]Sri Ramakrishna Engineering College, Coimbatore, India
E-Mail: alby.siby@gmail.com

ABSTRACT

In today's world, one of the major public health challenges is Diabetic Mellitus. The report of WHO says 347 million people worldwide have diabetes. Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Researchers are working to prevent this disease at early stage by predicting the symptoms of diabetes using several methods. Early identification of populations at high risk for diabetes is therefore important for targeted prevention strategies and is necessary to enable proper efforts to be taken for prevention in the large number of individuals at high risk, while avoiding the burden of prevention and treatment for the even larger number of individuals at low risk, both for the individual and for society. In India Diabetes is a huge problem and about one million people died of diabetes in 2012. The main aim of this study is to apply General Regression Neural Networks (GRNN) as a prediction model for Prediction of Type-2 diabetes in Indian Women.

Keywords: prediction model, public health, diabetes.

1. INTRODUCTION

Diabetes Mellitus is one of the fatal diseases growing at a rapid rate in developing countries like India. Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Insulin is a hormone that regulates blood sugar. Hyperglycaemia, 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. There are three types of diabetes. Type 1 diabetes, Type 2 diabetes and Gestational diabetes.

Type 1 diabetes (previously known as insulin-dependent, juvenile or childhood-onset) 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. Type 2 diabetes (formerly called non-insulin-dependent or adult-onset) results from the body's ineffective use of insulin. Type 2 diabetes comprises 90% of people with diabetes around the world, and is largely the result of excess body weight and physical inactivity. Symptoms may be similar to those of Type 1 diabetes, but are often less marked. As a result, the disease may be diagnosed several years after onset, once complications have already arisen. Until recently, this type of diabetes was seen only in adults but it is now also occurring in children [1].

Gestational diabetes is hyperglycaemia with blood glucose values above normal but below those diagnostic of diabetes, occurring during pregnancy. Women with gestational diabetes are at an increased risk of complications during pregnancy and at delivery. They are also at increased risk of type 2 diabetes in the future.

According to WHO report, 347 million people worldwide have diabetes . WHO projects that diabetes will be the 7th leading cause of death in 2030. In 2012, an estimated 1.5 million deaths were directly caused by diabetes. More than 80% of diabetes deaths occur in low- and middle-income countries [2].

In 2000, India (31.7 million) topped the world with the highest number of people with diabetes mellitus followed by China (20.8 million) with the United States (17.7 million) in second and third place respectively. It is predicted that by 2030 diabetes mellitus may afflict up to 79.4 million individuals in India. It is estimated that 61.3 million people aged 20-79 years live with diabetes in India (2011 estimates). This number is expected to increase to 101.2 million by 2030.

Nowadays, Bioinformatics is a challenging research area. Bioinformatics is the science of storing, extracting, organizing, analyzing, interpreting and utilizing information from biological sequences and molecules. Many researches have been conducted in this area and especially in the field of diabetes. Data mining (DM) techniques are wildly being applied by many researchers in bioinformatics. Data Mining is the science of finding new interesting patterns and relationships in huge amount of data. In the area of diabetes also many Data Mining techniques are used as a classifier as well as prediction model for diabetes. The DM methods being used today are taken from diverse fields as statistics, machine learning and Artificial Intelligence. Most popular methods include regression, classification and clustering. Regression is a statistical method that makes prediction of a certain dependent variable according to the values of other independent variables. The aim of this research was to apply General Regression Neural Networks (GRNN) as a prediction model for Prediction of Type-2 diabetes.

Many researchers have been conducted in the field of Prediction of Type-2 diabetes. They used different methodologies for prediction like AMMLP, Bayes Network, Genetic Programming, Neural Network and Fuzzy K-Nearest, GRNN, etc. Figure-1 shows a comparison of accuracy level of different methodologies in the field of Prediction of Diabetes.

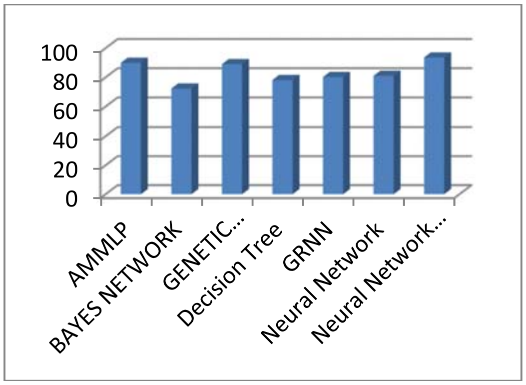

Figure-1. Comparison of accuracy level of different methodologies.

2. REVIEW OF LITERATURE

In the present scenario, it is very important if we can predict diabetes in the early stages or even before it surfaces. In this regard lot of researches are being carried out. Many researchers use various anthropometric measures for prediction of diabetes. In these cases it is observed that this cannot be fully dependent for prediction of diabetes. Here comes the importance of DM techniques for better prediction of diabetes.

In the paper "Data mining for the diagnosis of type 2 diabetes", Alexis Marcano-Cedeno *et al* proposed the artificial meta plasticity on multilayer perceptron (AMMLP) as a data mining (DM) technique for the diabetes disease diagnosis. The PIMA Indian dataset was used to test the proposed model. In this work they have included a comparison of results obtained by AMMLP with DT, BC and other algorithms. The AMMLP is based on the biological metaplasticity property of neurons and obtained an accuracy of 89.93% [3].

Yang Guo *et al* Network to build a decision make system for middle aged people to do self- prediction of type-2 diabetes at home and the accuracy was 72.3% [4]. The limitation of this study was, in this study the data collection did not consider all risk factors like family history, metabolic syndrome, and smoking, inactive lifestyles etc. In this study [5], prediction of fasting plasma glucose using Anthropometric Measures for diagnosing type-2 diabetes has been done. The drawback of this work is that it cannot establish a cause-effect relationship because of the cross – sectional nature of the data. This model may cause in correct diagnosis for high FPG status because the data in this study are based on a limited number.

In the paper "A Data Mining Approach for the Diagnosis of Diabetes Mellitus", Sonu Kumari *et al* proposed an effective methodology for the automated detection of Diabetes Mellitus. The dataset they have used includes people from different age group, gender and lifestyle. They have also considered many parameters which are linked with symptoms of diabetes like thirst increase, weight loss, etc and claim 92.8% accuracy. Another thing is that they have considered only 20 dataset for obtaining this result which is very much inadequate [6].

In [7] authors have constructed a prediction model using Decision Tree for diagnosis of diabetes. They used certain combination of pre-processing techniques to handle the missing values PIMA Indian dataset and compared the results of accuracy of the model for each technique, however the method of handling missing values presented in this paper wasn't employed in that study. After handling the missing values only 724 instances remain out of 768 with 5 attributes and obtained 78.17 % .Considering the Pima Indian diabetes dataset, there might be other risk factors that the data collections did not consider.

Authors in [8] have constructed association rules for classification of type -2 diabetic patients. They generated 10 association rules to identify whether the patient goes on to develop diabetes or not. Several of machine learning algorithms have been proposed in the context and have been successfully used in some parts. In their first stage of work the missing values were handled and applied equal interval binning with approximate values and lastly Pima Indian dataset were applied by the Apriori algorithm to generate the rules. In this study they have included only pregnant women below 21 years who are type-2 diabetic. Many other factors which influence diabetes may be considered for improving the generalization of rule.

In this paper [9], two classified techniques with principle component analysis (PCA) are implemented for the forecasting of diabetes and concluded with the best forecasting. The techniques are Neural Network and ANFIS, the dataset which they have utilized is the same Pima Indian dataset with 8 features. Using NN 72.9% and ANFIS 70.56 % with PCA 89.2% and 90.4% of accuracy respectively were being obtained. They have also designed one GUI in MATLAB using GUIDE to represent the work in simplest manner, so that any doctor who is not familiar with MATLAB can use this and it is inferred that the classification ability of all the classifiers is better for non-diabetic samples than that for diabetic ones.

M.W. Aslam *et al* in [10] uses genetic Programming (GP) and a variation of genetic programming called GP with comparative partner selection (CPS) for diabetes detection. The system produces a single individual as output from training data, that converts the available features to a single feature such that it has different values for healthy and patient(diabetes) data and in the next stage use test data for testing of that individual. The proposed system was able to achieve only 78.5% accuracy.

In 2003 K. Kayaer *et al* applied three different neural network structures namely MLP, RBF and GRNN

ARPN Journal of Engineering and Applied Sciences

to PID medical data [11]. They came up with an accuracy of 80.21% to classify a medical data. Even though in our proposed work we are also using GRNN for prediction of diabetes comparing to this work we have used a different approach for the prediction. In [12] Xiao Fang tried to make use of different DM concepts to gain knowledge about diabetes. Using several challenging applications of DM, an application for identifying diabetic patients in a small US town was presented. Even though different methods produced different accuracy levels, on an average 75% of accuracy was obtained. The limitations of this work are the population is not clustered into different risk related populations and the different variables that found critical for identifying diabetic patients are not subcategorized.

The authors in [13] propose a system which can improve the strategy to a better level where artificial metaplasticity on perceptrons is implemented on neural network. The nodes of neural network are parameters containing features like thirst increase, hungry increase, nausea, fatigue, vomiting etc. This proposed system can increase the efficiency of the system which is in existence.

Madhavi Pradhan *et al* in [14] uses a neural network implementation of the fuzzy k- nearest neighbour algorithm for designing of classifier for detection of diabetes. This method of detection of diabetes proposes a system that will be implemented in client – server architecture. The training dataset will be kept on the server, which will be used to train the neural network classifier on the mobile device. The mobile device is a feature add-on for convenience of the doctor. The accuracy of this method is calculated as 72.82%. Kalliopi. V. Dalakleidi et al present a hybrid approach based on the *combined* use of a genetic algorithm (GA) and a nearest neighbours classifier for the selection of the critical clinical features which are strongly related with the incidence of fatal and non-fatal Cardiovascular Disease (CVD) in patients with Type 2 Diabetes Mellitus (T2DM) [15]. In order to overcome the problem of unbalanced data the dual weighted k-nearest neighbour classifier was used.
The aim of this study [16] was to examine whether waist circumference (WC) or WHR improve diabetes prediction beyond body mass index in older men and women, and to define optimal cut-off points. Both overall and central adiposity indices are strong predictors of type 2 diabetes in older adults. BMI is as strong as WC in predicting type 2 diabetes in men. In women, however, WC was a significantly better measure for the identification of diabetes risk.

Studies aiming at preventing or delaying diabetes are critically dependent on the ability to accurately predict Type1 and Type 2 diabetes [17]. After a median follow up time of 8 years, Islet cell auto antibodies did not predict diabetes. BMI measured at base line was as effective as 2-h plasma glucose and fasting plasma glucose to predict diabetes in the adult population.

In [18] the objective of Meredith F Mackay et. al was to compare different anthropometric measures in terms of their ability to predict type 2 diabetes and to determine whether predictive ability was modified by ethnicity. Waist - height ratio was the most predictive measure, followed by BMI. Measures of central and overall adiposity predicted type 2 diabetes to a similar degree.

3. METHODOLOGIES

The dataset

In this research work, we analysed data of females aged at least 21 years from Pima Indian Dataset.768 instances have been considered. The following attributes have been considered:

a) Number of times pregnant
b) Plasma glucose concentration a 2 hours in an oral glucose tolerance test
c) Diastolic blood pressure
d) Triceps skin fold thickness
e) 2-hour serum insulin
f) Body mass index
g) Diabetes pedigree function
h) Age

The WHO reports identified that females after 21 years are mostly affected by diabetes when compared to males. So, we have taken females dataset into account.

GRNN algorithm

General Regression Neural Networks (GRNNs) falls into the class of probabilistic neural networks. The employment of a probabilistic neural network is particularly advantageous as a result of its ability to converge to the underlying operate of the information with solely few coaching samples obtainable. The extra information required to urge the slot in a satisfying approach is comparatively tiny and might be kept away from additional input by the user. Thus GRNN can be considered as an awfully useful gizmo for performing predictions and also for system performance comparisons [19].

Figure-2 shows the general structure of GRNN. There are four layers in all GRNN networks:

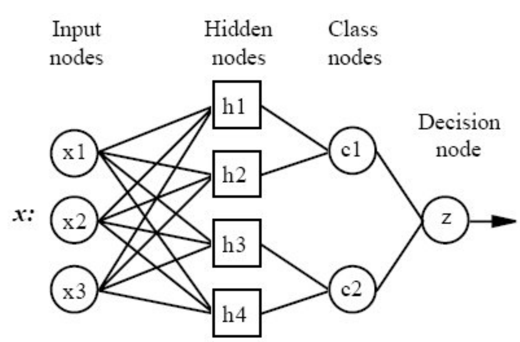

Figure-2. General structure of GRNN.

VOL. 11, NO. 3, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

1. Input layer - There is single vegetative cell for every variable within the input layer. In the case of categorical variables, N-1 neurons square measure used wherever N is that the range of classes. In process prior to the input layer, the input neurons regularizes the vary of the output by finding the difference between the medians and dividing by the interquartile vary. The input neurons then provide the values to each one of the neurons within the hidden layer.

2. Hidden layer - This layer has single vegetative cell for every case within the coaching information set. The vegetative cell contains the worth of the predictor variables for the case beside the target value. Once given with the x vector of feed values from the first layer, a hidden vegetative cell calculates the Euclidian distance of the legal action from the centre of neuron then applies the RBF kernel operate exploitation the letter value(s). The ensuing worth is introduced to the neurons within the next layer ie. Pattern layer.

3. Pattern layer/Summation layer - Succeeding layer within the network is totally dissimilar for GRNN networks and for PNN networks. For PNN networks there is one pattern vegetative cell for every class of the target variable. The particular target class of every coaching case is keep with each hidden vegetative cell; the weighted worth kicking off of a hidden vegetative cell is fed solely to the pattern neuron that resembles to the hidden neuron's class. The pattern neurons commute the values for the category they symbolize (hence, it's a weighted vote for that category).

For GRNN networks, there square measure solely 2 neurons within the pattern layer. One vegetative cell is that the divisor summation unit the opposite is that the dividend summation unit. The divisor summation unit adds up the burden values coming back from every of the hidden neurons. The dividend summation unit adds up the burden worth increased by the particular target value for every hidden vegetative cell.

4. Decision layer -The choice layer is totally different for GRNN and PNN networks. For PNN networks, the choice layer relates the weighted votes for every target class added within the pattern layer and uses the most important vote to forecast the target class.
For GRNN networks, the choice layer splits the worth added within the dividend summation unit by the worth within the divisor summation unit and uses the result because the expected target value.

Algorithm

The likelihood density perform utilized in GRNN is that the Gaussian distribution. Every coaching sample, X_j, is employed because the mean of a standard distribution.

$$Y(X) = \frac{\sum_{i=1}^{n} Y_i \, exp\left(-D_i^2 / 2\sigma^2\right)}{\sum_{i=1}^{n} exp\left(-D_i^2 / 2\sigma^2\right)}$$

$$D_i^2 = \left(X - X_i\right)^T \cdot \left(X - X_i\right)$$

The distance, D_j, between the coaching sample and therefore the purpose of prediction, is employed as a live of however well the every coaching sample will represent the position of prediction, X. If the gap, D_j, between the coaching sample and therefore the purpose of prediction is tiny, exp (- D_j two /2s two), becomes huge. For D_j=0, exp(-D_j two /2s two) becomes one and therefore the purpose of analysis is diagrammatic best by this coaching sample. The gap to all or any the opposite coaching samples is larger. a much bigger distance, D_j , causes the term exp(-D_j two /2s two) to become littler and so the contribution of the opposite coaching samples to the prediction is comparatively small. The term Y_j*exp(-D_j two /2s two) for the j^{th} coaching sample is that the largest one and contributes pretty much to the prediction. The quality deviance or the smoothness parameter is s. For a much bigger smoothness parameter, the attainable illustration of the purpose of analysis by the coaching sample is feasible for a wider vary of X. For a little worth of the smoothness parameter the illustration is proscribed to a slender vary of X, severally. With higher than equation it's attainable to

- Predict behaviour of systems considering a few training samples
- Forecast even multi-dimensional curves
- Incorporate between training samples.

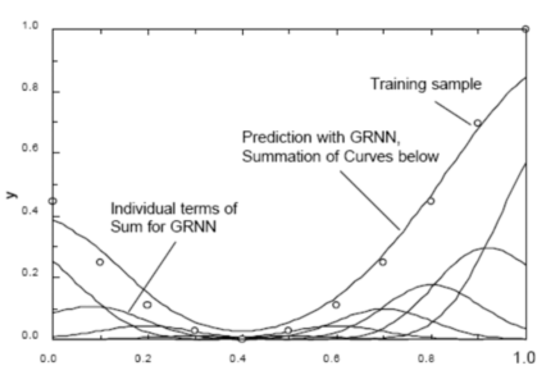

Figure-3. GRNN with individual terms contributing to prediction, s=0.1.

VOL. 11, NO. 3, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Algorithm for the proposed GRNN based prediction Model

Step 1: *Start*
Step2: *Input data.*
Step 3: *Normalization process is done for the input data.*
Step 4: *After normalization time series GRNN is generated for the input data.*
Step 5: *Then initial weights are assigned and the prediction is done.*
Step6: *Here the weights are randomly assigned to the input data.*
Step 7: *Then based on random assignment of weights and bias, the neurons in GRNN will create an equation for the input data to produce output.*
Step 8: *The output of the first input data is added to the second input data.*
Step 9: *Then output of second data is added to the third data*
Step 10: *Repeat the process till end of the data set*
Step 11: *The randomly assigned weights and bias will make the neurons to converge.*
Step 12: *Output is produced.*
Step 13: *End*

The basic plan is that an expected target price of an item is probably going to be concerning a similar as different things that is having shut values of the predictor variables.

Considering that every case within the coaching set has 2 predictor variables, x and y. The cases are planned victimisation their x,y coordinates as illustrated within the figure 3. Additionally take up that the target variable has 2 classes, positive that is represented by a sq. and negative that is represented by a splash. Now, assume we tend to try to predict the worth of a replacement case delineate by constellation with predictor values x=6, y=5.1. Ought to we tend to predict the aim as positive or negative?

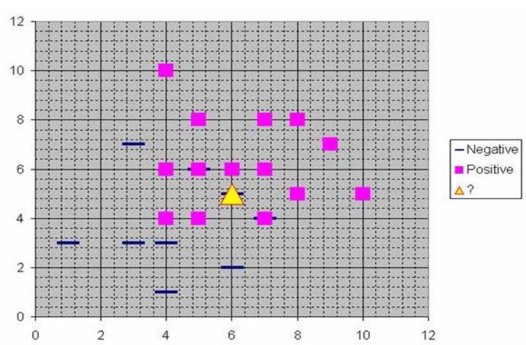

Figure-4. Working of GRNN.

Notice that constellation is position nearly precisely on prime of a splash representing a negative price. However that dash is during a fairly uncommon position associated to the opposite dashes that are grouped below the squares and left of centre. Therefore it may be that the underlying negative price is an odd case.

The nearest neighbour classification performed for this instance depends on what number neighbouring points thought-about. If 1-NN is employed and solely the nearest purpose is taken into account, then clearly the new purpose ought to be categorized as negative since it's on prime of a well-known negative purpose. On the opposite hand, if nine-NN categorization is employed and also the highest 9 points are thought-about, then the result of the encircling eight positive purposes could overbalance the shut negative point.

A probabilistic neural network constructs on this foundation and simplifies it to think about all of the opposite points. The space is added from the purpose being assessed to every of the opposite points, and a radial basis perform (RBF) (also referred to as a kernel function) is applied to the space to cipher the burden (influence) for every purpose. The radial basis performance is therefore named as a result of the radius distance is that the argument to the performance.

Weight = RBF (distance)

The more another purpose is from the new purpose, the less influence it's.

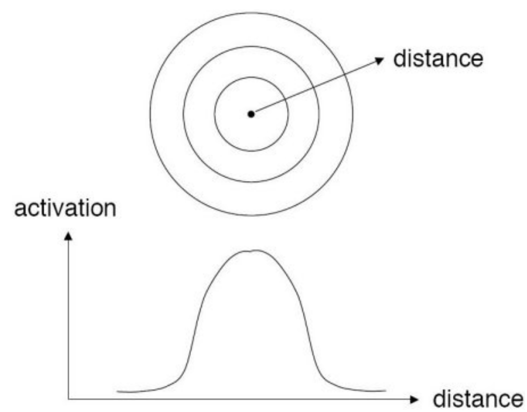

Figure-5. Radial basis function (Calculating Distance).

Radial basis operation

Different types of radial basis functions may well be used; however the foremost common is that the mathematician functions:

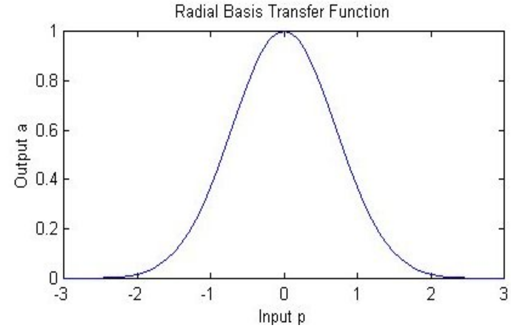

Figure-6. Radial basis using Gaussian function.

VOL. 11, NO. 3, FEBRUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

If there's over one variable quantity, then the RBF perform have as several dimensions as there are variables. Here could be a RBF perform for 2 variables:

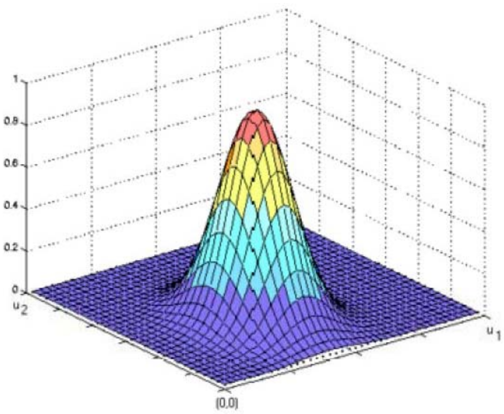

Figure-7. Radial basis function for two variables.

The best expected price for the new purpose is found by summing the values of the opposite points weighted by the RBF operate.

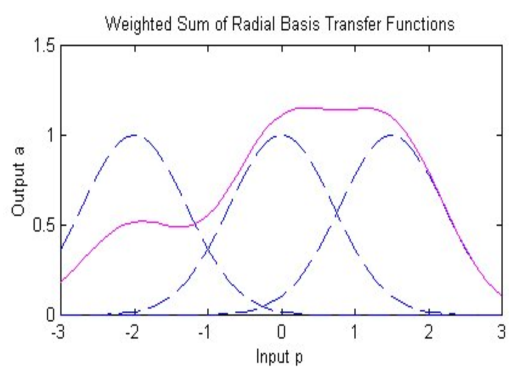

Figure-8. Spread of radial basis function.

The peak of the radial basis operate is often focused on the purpose it's coefficient. The letter worth (σ) of the operate determines the unfold of the RBF function; that's, however quickly the operate declines because the distance magnified from the purpose.

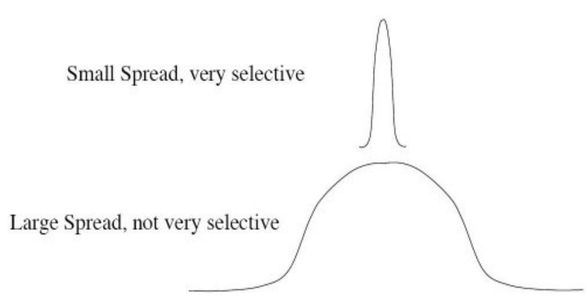

Figure-9. Spread of radial basis function.

With larger letter of the alphabet values and a lot of unfold, distant points have a larger influence. The primary work of coaching a GRNN network is choosing the optimum letter of the alphabet values to manage the unfold of the RBF functions.

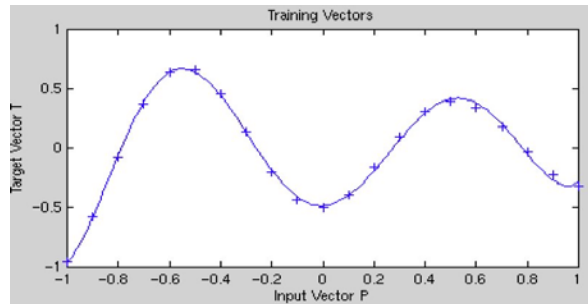

Figure-10. Selecting sigma values to control spread of RBF.

4. RESULT AND DISCUSSIONS

In ordinary GRNN, based on weights and bias, the neurons will create an equation for the input data to produce required output. The neurons will train the input data's with that equation to produce the required output. In our proposed work, first normalization process is done for the input data. After normalization time series GRNN is generated for the input data. Then initial weights are assigned and the prediction is done. Here the weights are randomly assigned to the input data. Then based on random assignment of weights and bias, the neurons in GRNN will create an equation for the input data's to produce output. The output of the first input data is added to the second input data. Then output of second data's is added to the third data. Like this the process repeats till end of the data set. The randomly assigned weights and bias will make the neurons to converge. So, convergence occurs quickly here when compared to the working of existing GRNN.In our work, we have obtained an accuracy level of 91.45% in training and 93.64% in testing.

The table shows the results obtained by GRNN

Table-1. Results obtained by GRNN.

	Training	Testing
GRNN	91.45%	93.64%

5. CONCLUSIONS

In the analysis conducted, it is evident that it is very important that to detect diabetes at a very early stage than treating it once it is surfaced. There are various methodologies used for prediction of diabetes applied by many researchers. In this work GRNN is applied as prediction model for predicting diabetes. In the future work, the plan is to use a better technique with the same dataset and features for improved accuracy levels.

ARPN Journal of Engineering and Applied Sciences

REFERENCES

[1] B. L. Shivakumar, S. Alby, "A Survey on Data-Mining Technologies for Prediction and Diagnosis of Diabetes", International Conference on Intelligent Computing Applications (ICICA) 2014, pp. 167-173

[2] http://www.ArogyaWorld_IndiaDiabetes_FactSheets_CGI2013_web.

[3] Alexis Marcano-Cedeno, Diego Andina. 2012. Data mining for the diagnosis of type 2 diabetes. World Automation Congress (WAC), Puerto Vallarta, Mexico. pp. 24-28.

[4] Yang Guo, GuohuaBhai, Yan hu. 2012. Using Bayes Network for Prediction of Type-2 Diabetes. The 7th International Conference for Internet Technology and Secured Transactions. 978-908320-08/7/$25.00©2012 IEEE. pp. 471-472.

[5] Bum Ju Lee, BonchoKu, JihoNam, Duong Duc Pham, Jong Yeol Kim. 2014. Prediction of Fasting Plasma Glucose Status Using Anthropometric Measures for Diagnosing Type-2 diabetes. IEEE Journal of Biomedical and Health Informatics. 18(2): 555-561.

[6] Sonu Kumari. Archana Singh. 2012. A Data Mining Approach for the Diagnosis of Diabetes Mellitus. 978-1-4673-4603-0112/$31.00 ©2012 IEEE. pp. 373-375.

[7] Al Jarullah A.A 2011. Decision Tree Discovery for the Diagnosis of Type II Diabetes. International Conference on Innovations in Information Technology (itt). pp. 303-307.

[8] Patil B.M., Joshi R.C., Toshniwal D. 2010. Association Rule for Classification of Type-2 Diabetic Patients. Second International Conference on Machine Learning and computing (lCMLC). pp. 330-334.

[9] Rakesh Motka, Viral Parmar, Balbindra Kumar, A. R. Verma. 2013. Diabetes Mellitus Forecast Using Different Data Mining Techniques. 4th International Conference on Computer and Communication Technology (ICCCT). pp. 99-103, DOI: 10.1109/ICCCT.2013. 6749610.

[10] M.W. Aslam, A.K.Nandi. 2010. Detection of Diabetes using Genetic Programming. 18th European Signal Processing Conference, pp. 1184-1188.

[11] K. Kayaer and T. Yildrim. 2003. Medical Diagnosis on Pima Indian Diabetes using General Regression Neural Networks. International Conference on artificial neural networks and neural information processing. pp. 181-184.

[12] Xiao Fang. 2009. Are you becoming a Diabetic? a Data Mining Approach. Sixth International Conference on Fuzzy systems and Knowledge Discovery. 5: 18-22.

[13] Poonguzhali E, SabarmathiKabilan, Sandia Kannan, Sivgami P. 2014. Diagnosis of Diabetes Mellitus Type -2 using Neural Network. International Journal of Emerging Technology and Advanced Engineering. 4(2).

[14] Madhavi Pradhan,Ketki Kohale, ParagNaikade, AjinkayaPachore, EknathPaiwe. 2012. Design of Classifier for Detection of Diabetes using Neural Network and Fuzzy k-Nearest Neighbour Algorithm. International Journal of Computational Engineering Research. 2(5): 1384-1387.

[15] Kalliopi V Dalakleidi, Konstantia Zarkogianni, Vassilios G Karamanos, Anastasia C Thanopoulou, Konstantina S Nikita. 2013. A Hybrid Genetic Algorithm for the Selection of the Critical Features for Risk Prediction of Cardiovascular Complications in Type 2 Diabetes Patients. IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE), DOI: 10.1109/BIBE.2013.6701620, pp. 1-4.

[16] S. G. Wannamethee, O. Papacosta, P. H. Whincup, C. Carson, M. C. Thomas, D. A. Lawlor, S. Ebrahim, N. Sattar. 2010. Assessing prediction of diabetes in older adultsusing different adiposity measures: a 7 year prospectivestudy in 6, 923 older men and women. Springer, Diabetologia, 53: 890-898, DOI 10.1007/s00125-010-1670-7.

[17] O. Rolandsson, E. Hagg, M. Nilsson, G. Hallmans, L. Mincheva-nilsson, A. Lernmark. 2001. Prediction of diabetes with body mass index, oral glucosetolerance test and islet cell autoantibodies in a regional population", Journal of Internal Medicine. 249: 279±288.

[18] Meredith F Mackay, Ralph B Agostino. 2009. Prediction of type-2 diabetes using Alternate Anthropometric Measures in a Multi- Ethnic Cohort. Diabetes Care. 32(5).

[19] https://minds.wisconsin.edu/bitstream/handle/1793/7779/ch2.pdf.