TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 4867-4870

Diagnosing Diabetes using Artificial Intelligence

[1]Keerthana HJ, [2]K Mohammadi Pasha, [3]R Meghana, [4]Sai Aravind Reddy K, [5]Lithin Kumble

[1,2,3,4]School of C&IT, REVA University, Bangalore, India
[5]Assistant professor, School of C&IT, REVA University, Bangalore, India
[1]Keerthanahj1@gmail.com, [2]Kmohammadipasha@gmail.com, [3]writetomeghanar@gmail.com
[4]saiaravindreddyk@gmail.com, [5]lithink@reva.edu.in

Article Info
Volume 83
Page Number: 4867-4870
Publication Issue:
May-June 2020

Abstract

Diabetes is a condition that causes high blood sugar level in body. If sugar levels aren't maintained properly, then diabetes is a disease where it causes damage to the other parts of body. If diagnosing and treatment of diabetic patients are postponed, it will lead to some major issues such as heart attacks, renal failure etc. Most important problem within the medical world is timely and precise diagnosis of disease. Issues of patience can be reduced if the disease is diagnosed at the proper time and special medical aid is provided to the patients. In this study, PIDD data set is used and different models are implemented. Different model includes Decision tree, Support vector machine, Logistic regression, K neighbor classifier, XGboost, Ensemble Hybrid [Naïve baye, KNN, Logistic regression], Multilayer perceptron, Naïve baye, Random forest. The accuracies of those models range between 67% to 78%. Proposed model is implemented using recurrent neural network [RNN] and it provides accuracy of 90.4%. The results shows that proposed RNN model provides higher accuracy level i.e. 90.4% than the prevailing models and other models.

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *16 May 2020*

Keywords: *Decision tree, Ensemble hybrid model, K neighbor classifier, RNN, XGboost.*

1. Introduction

Now a day, diabetes is significant problem with in the world. Eating unhealthy food leads to major diseases. Diagnosing this disease at earliest as possible is very important in medical world. If diabetes isn't on the top of things, it's going to cause the damage to other parts of body. Three sorts of diabetes are: Type-I, Type-II and gestational diabetes. Type-I diabetes is caused when pancreas beta cell is destroyed and it is no longer in position to generate insulin. Gestational diabetes may bea condition that is diagnosed in pregnant women's. Gestational diabetes causes high blood sugar level that affects the pregnancy. Type-II diabetes is called as adult-onset diabetes. Type-II diabetes is caused when body becomes resistant to insulin or when pancreas is unable to produce enough insulin. According to the Diabetes research centreit has been shown that diagnosing the diabetes at the early stage can prevent 80 percent complication of Type-IIdiabetes patients. Since diabetes is long lasting disease and it causes major damage to the other parts of body, Machine learning and other neural network algorithms are used to enhance detection method and to control disease which will be great help to medical world.

In the present study machine learning and neural network algorithms are used in order to build a model which predicts diabetes. There are some steps which are more important than the other steps while building the model i.e. data preprocessing and applying suitable learning algorithm which are used to achieve required results. Proposed Recurrent neural network model will be able to predict the diabetes type-II with higher accuracy than the other implemented models.

Published by: The Mattingley Publishing Co., Inc.

May-June 2020
ISSN: 0193-4120 Page No. 4867-4870

2. Literature Survey

Ambika Rani Subhash, et al. [1]. Proposed "Accuracy of classification algorithm for diabetes prediction". Aim of this paper is to implement the model with atmost accuracy. Training models are designed using WEKA and four classification algorithms namely Naïve bayes, J48, SVM and Neural network have been used on data set. Data is collected. The performance of each algorithm is analyzed on the basis of accuracy. Accuracy of Naïve bayes, J48, SVM and Neural network algorithm are 75.39%, 74.34%, 64.58% and 76.69% respectively.

Jatin N Bagrecha, et al. [2].Proposed "Diabetes Disease prediction using Neural Network". 768 data samples are collected and multilayer perceptron neural network is the algorithm employed to predict the diabetes. Result obtained has an accuracy of 84%.

K. Sai prasanna, et al. [3].Proposed "An Efficient Intelligence Diabetes Disease Prediction using AI Techniques". 768 patient records are collected. Convolution neural network is employed for predicting diabetes. The Model implemented using CNNprovides84.4%accuracy.

Nesreen Samer E1_Jerjawi, et al. [4].Proposed" Diabetes prediction using artificial neural network". They used ANN to diagnose diabetes. Their criteria is to attenuate error function in model. After training and analyzing, the model was able to provide accuracy of 87.3% and error rate of 00.1%.

Minyechil Alehegn et al. [5]. Proposed "Analysis and prediction of diabetes disease using machine learning algorithm: ensemble approach". WEKA and java tools are used to predict diabetes. They combined four algorithms i.e. KNN, Naïve bayes, Random forest, and J48 into one ensemble hybrid model. This model provides higher accuracy then the individual algorithms.

J Beschi Raja, et al. [6].Proposed "Diabetics prediction using Gradient Boosted classifier". They used UCI Pima Indian diabetic data set and the model is implemented using gradient boosted classifier and that model is compared with two machine learning model i.e. neural network and random forest. Performance evaluation of model is analyzed by calculating accuracy and recall.

Sai poojitha Nimmagadda, et al. [7] proposed "Improved Diabetes Prediction Model for predicting Type-II Diabetes". Pima Indian data set is considered. Proposed IDPA model uses two algorithm i.e. hierarchical clustering algorithm and naïve byes algorithm. IDPA model is analyzed with different models. IDPA model provide more accuracy in comparison to the other models.

3. Objective

Main objectives are:

- Implementing the model to diagnose diabetes which gives high accuracy when compared to prevailing one.
- Comparing the proposed model to other implemented model and to prove that the proposed model has higher accuracy rate than other models.

4. Proposed System

Website is the part of proposed work. Website forepart is implemented with HTML and backend with python language. Forepart of this User interface accepts the user input data and sends that data to backend for process.

Different models are implemented using machine learning algorithms and RNN. In this neural network functions like Rectified linear unit [ReLU] and sigmoid are used in order to enhance accuracy of proposed model. Proposed model is compared with other machine learning model on the basis of accuracy. Implementation process is explained below.

Implementation

Data Set

Raw data is collected [Pima Indian Dataset]. 768 samples are collected consistent with requirements. Samples have 9 attribute. Among these attributes one attribute is label. Every row has information of individual patient. Fig. 1 indicates attributes in the data set.

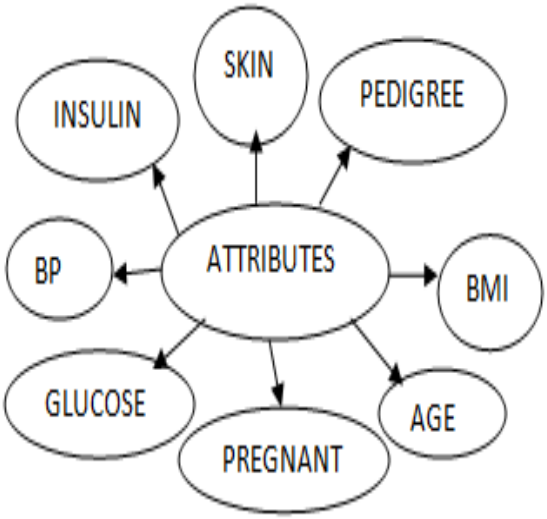

Figure 1: various features in data set

Data Preprocessing

Pre processing process is used to analyze the lost values in dataset. There are some attributes which shows lost data. These lost data should be analyzed. So, these lost values are replaced. In order to calculate the correlation between values, data set is evaluated using Pearson correlation method. Fig. 2 indicates correlation between the attributes.

Published by: The Mattingley Publishing Co., Inc.

May-June 2020
ISSN: 0193-4120 Page No. 4867-4870

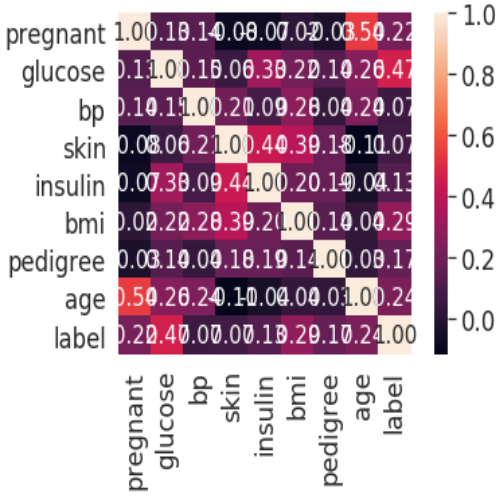

Figure 2: Correlation between attributes

Modelling

To build model, first we need to import libraries and we need toupload dataset. Data set is split into two types X and Y. X has attributes which are used to predict output and Y is foreseen results.

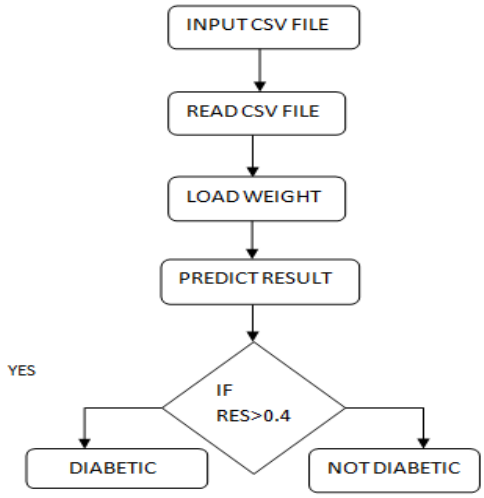

Figure 3: Block diagram of proposed system

CODING TOOL:Google colab is used as integrated development environment. It helps us to improve python programming coding skills and also to develop applications using popular libraries such as keras, pandas, Tensor flow etc.

Model consists of eight neurons in input layer and a pair of neurons in hidden layer. First hidden layer has 64 neurons and second hidden layer has 12 neurons. Then model is fitwith 150 batch size and 500 epochs.

Neural network accommodates of 8-64-12-1 neurons. ReLU is function in hidden layer. This function gives zero for all negative input.

Sigmoid is the function used in output layer. This function is non- linear in nature and 'S' formed graph.

Performance Analysis

After model is built performance of all algorithms is calculated. Accuracy of all algorithms are calculated and compared.

Comparison with Other Model

Different models were built and comparison is completed among other models and RNN model on the ideaof accuracy. Among all the models, RNN model provides higher accuracy rate of 90.4%.
Different model names and accuracy of every model is given Table .1.

Table 1: Accuracy of various models

MODEL NAME	ACCURACY OF MODEL
Logistic regression model	78.7%
Naïve bayes model	66.5%
Randomforest model	77.1%
Support vector machine model	74.4%
Decision tree model	68.8%
KNeighbors classifier model	73.2%
GBOOST model	76.7%
Multilayer perceptron model	61.8%
Ensemble Hybrid model[KNN, Naïve bayes, Logistic regression]	76.7%

Accuracy plot of various models is given in Fig.4.

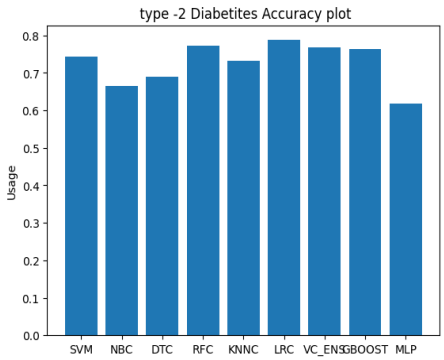

Figure 4: Accuracy plot of various models

5. Methodology

Step1: Initially Load the Data set for testing patient data. Values are taken from the user through the input form and on submitting .csv file gets generated for the further process. Fig.5 indicates the input form to accept single patient data.

May-June 2020
ISSN: 0193-4120 Page No. 4867-4870

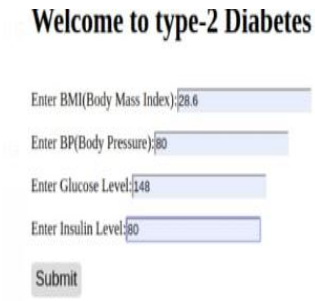

Figure 5: Input form to accept data

Step2: In this step, parsing the input data is completed. The generated file i.e. csv is read and sent to model for testing.

Step3: During training phase, models learn the importance of each and every feature. Hence, additional importance is given to the affecting attributes.

Step4: Data from user interface and .csv file is fed to model. On those data algorithms runs and results are displayed in screen.

6. Results

Main objective of this study is to built model that diagnose diabetes type-II which gives higher accuracy than prevailing one. Machine learning algorithms are used on data and comparison of models on the idea of accuracy is done. Data set contains 768 samples with 8 attributes. After training, validating and testing the dataset, among all the models, model implemented using RNN is in the position to provide90.4% accuracy.

Outputfalls into following division. Diabetic [positive] or not diabetic [negative]

Fig.6 indicates output screen.

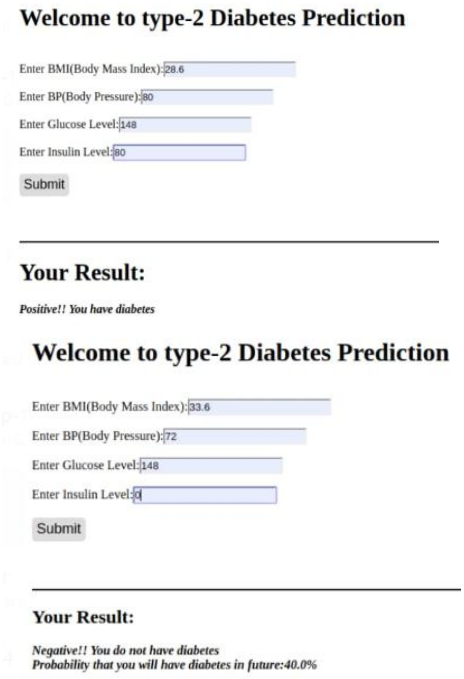

Figure 6: Result Page

7. Conclusion

Now a day's Diabetes is one among the common disease. In the present work RNN is used to predict the diabetes type-II. RNN model features an accuracy of 90.4% which is relatively above the previously implemented techniques. The designed user interface is a user understandable. ReLU and sigmoid functions helped to increase the performance of model. The model is often enhanced by implementing it to accept multiple patients' records and to implement mobile application which predicts the diabetes in patients.

References

[1] Ambika Rani subhash "Accuracy of classification Algorithms for Diabetes prediction" 2019 International Journal of Engineering and Advanced Technology (IJEAT) Volume:8, Issue :5.

[2] Jatin N Bgrecha "Diabetes Disease Prediction Using Neural Network", 2019 International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume:7, Issue:4.

[3] K. Sai Prasanna Kumar Reddy "An Efficient Intelligent Diabetes Disease Prediction using AI Techniques" 2019 International Journal of Recent Technology and Engineering (IJRTE) Volume:8, Issue:4.

[4] Nesreen sameer E1_Jerjawai "Diabetic Prediction Using Artificial Neural Network" 2018 International Journal of Advanced Science and Technology Volume:124.

[5] Minyechil Alehegn "Analysis and prediction of diabetes disease using machine learning algorithm: Ensemble approach"2017 International Research Journal of Engineering and Technology (IRJET) Volume:4, Issue:10.

[6] J. Beschi Raja " Diabetics Prediction using Gradient Boosted Classifier" 2019 International Journal of Engineering and Advanced Technology(IJEAT) Volume:9, Issue:1.

[7] Sai Poojitha Nimmagadda "Improved Diabetes Prediction Model For Predicting Type-II Diabetes" 2019 Internatioanl Journal of Innovative Technology and Exploring Engineering (IJITEE) Volume:8, Issue:12.