

Diabetes Prediction using Neural Network

Namrata Nerkar¹, Vaishnavi Inamdar², Likhith Kajrolkar³, Prof. Rohit Barve⁴

^{1,2,3}Department of Information Technology, Vidyalkar Institute of Technology, Wadala, Mumbai

⁴Prof. Rohit Barve, Department of Information Technology, Vidyalkar Institute of Technology, Wadala, Mumbai

Abstract - Diabetes is a stage where the measure of insulin levels of blood lessens. Indications of high blood glucose brings about frequent urination, feeling thirsty, and increased appetite. If not treated, it will prompt various difficulties, may lead to death. Serious troubles lead to cardiovascular sickness, foot injuries, and blurry visions. When there is an increase in blood glucose level, it is said to be early stages of diabetes. Various algorithms present different decision support systems for assisting health specialists. Using Artificial Intelligence in the medical field now-a-days plays an important role in prediction of diseases. If we can predict early, it can not only reduce the cost of the whole procedure and treatment, but also prevent the multiple organ dysfunction and casualties. Artificial Neural Network (ANN) is chosen for building the model to predict diabetes among several Machine learning algorithms, because Artificial Neural Networks provide promising results for Non-linear data. The objective is to build a decision support system to predict and diagnose diabetes with maximum precision given the parameters and a chatbot to assist you by taking a virtual test. The parameters are set such that the best accuracy is obtained. A user friendly GUI and chatbot is prepared to take the input values from the users and show the predicted result

Key Words: Machine Learning, Artificial Neural Networks, Classification, Prediction, Chatbot

1. INTRODUCTION

Diabetes is a fast-growing chronic disease among the individuals even among the young generation, everywhere on the globe. Rise in the degree of the sugar (glucose) in the blood causes diabetes. Diabetes can be characterized into two categories, for example, type 1 diabetes and type 2 diabetes. Type 1 diabetes is an auto-immune disease. In this situation, the body demolishes the cells that create insulin to absorb the sugar to produce energy. Also, this sort of diabetes can cause Obesity. Obesity is an increase in Body Mass Index (BMI) than the normal degree of BMI of an individual. Type 1 diabetes can happen at a very young age or in childhood. Type 2 diabetes usually affects the adults who are overweight. In this sort of diabetes, the body resists insulin or fails to create insulin. Type 2 generally happens in the aged groups or middle-aged groups. In addition, there are different foundations for diabetes, for example, bacterial or viral disease, poisonous or compound substance in food, immune system reaction,

obesity, disturbed diet routine or unhealthy diet, changed lifestyle, dietary patterns, environmental pollution, etc.

Diabetes prompts different sicknesses, for example, cardiovascular complications, Paralysis, Blindness, foot ulcers, etc. Hence, the early prediction of Diabetes can support experts and patients for required treatments. And with the help of latest technologies such as Machine Learning and various algorithms and Prediction models we can be informed about the health risks. Machine Learning is a subfield of AI, and Machine learning calculations can be characterized into three classifications, such as, Supervised Learning, Unsupervised Learning and Reinforcement Learning. In the Proposed system, we use supervised learning algorithms for testing out accuracy among some sort of popular Machine Learning (ML) algorithms. Supervised Learning algorithms learn the pattern from pre-existing data and try to predict new results based on the previous learning. ML algorithms are used to recognize existing data like probability based, function based, rule-based, tree-based, occurrence-based, etc. Various Machine Learning algorithms are presented using different data mining algorithms for helping experts in the medical fields.

2. LITERATURE SURVEY

Diabetes is a chronic and metabolic disease characterized by increased levels of blood glucose (or glucose), which leads to serious harm to the heart, veins(blood vessels), eyes, kidneys and nerves. Among the types of diabetes, Type 2 is the most common sort, normally found in middle-aged to aged people, which happens when the body gets resistant to insulin or doesn't produce enough insulin. Type 1 diabetes, when known as juvenile diabetes or insulin-dependent diabetes, is a persistent condition, where the pancreas creates very less or negligible amounts of insulin. Around 422 million people around the globe have diabetes, the majority of which are found in low to middle-income nations, and 1.6 million deaths occur of diabetes every year. Both the number of cases and the commonness of diabetes have been consistently expanding in the course of recent many years. [1].

Prior prediction of diabetes can lead to improved treatment, but due to lack of early diagnosis, the symptoms worsens in many cases. It is very essential to have a prediction device which can be utilized to determine if somebody has diabetes or not. There are a

few techniques which produce precise prediction and the Artificial neural network with the help of Back propagation neural network algorithm is one of them [2].

2.1 Artificial Neural Network

An ANN is a data processing system which consists of a large number of simple, highly interconnected processing elements in an architecture inspired by the structure of the cerebral cortex portion of the brain. [3]. The neural network performs functions as the human nervous system and the brain processes information and thus tries to replicate the way humans learn. Neurons are the fundamental cellular unit of the brain. The neurons are responsible for receiving sensory information from the external world via dendrites, processes the information and gives output via axons. Similarly an Artificial neural network consists of an input layer which consists of many neurons that takes the input and an output layer which gives the output to the external world. In most cases a hidden layer is present between the input and output layers which transforms the input into something which can be used by the output layer.

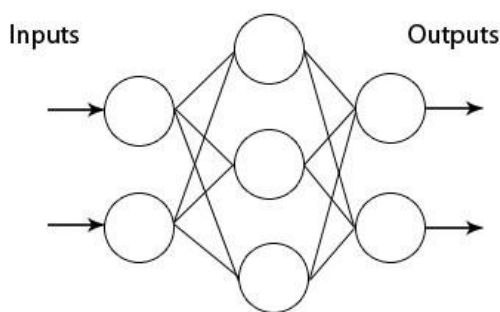


Fig. An Artificial Neural Network

In ‘architecture of multi-layer feed forward neural network’, Rajeeb Dey and Vaibhav Bajpai [4] proposed two concepts. One which is 6-10-1(single hidden layer with 10 neurons) and second 6-14-14-1(double hidden layer with 14 neurons) architecture. Very few parameters which are the same parameters used for diabetes testing in lab were used such as fasting blood sugar level, random blood sugar level, post-plasma blood sugar level, age, sex, occupation. ‘Smooth linear activation function’ is applied to input neurons and ‘non-linear log sigmoid function’ is applied to hidden neurons. Hidden and output layers were supplemented with constant bias. Very low learning rates are used for training. The specified error goal was 10-5 percent. From this paper it is clear that the error rate is attainable by adjusting the number of neurons and learning rate in the network. Hence by varying the learning rate and increasing the number of epochs less error can be set.

In ‘Diabetes Disease Prediction Using Data Mining’, Deeraj Shetty et al [5], proposed the use of algorithms like Bayesian and KNN (K-Nearest Neighbor) to apply on diabetes patient’s databases and analyze them by taking various attributes of diabetes for prediction of diabetes disease. They concluded that the large dataset yields better results. They also illustrated a recommendation system to the patient on how to control or prevent diabetes in case of minor signs of diabetes. The attributes used in the system were age, pregnancy, pg concentration, tri fold thick, serum ins,body mass index (bmi), dp function, diastolic bp i.e. the factors are majorly responsible for diabetes.

In ‘Diabetes Prediction Using Ensemble Perceptron Algorithm’, R Mirshahvalad and N. A. Zanjani [6] proposed a learning algorithm which is an ensemble boosting algorithm with perceptron algorithm to improve performance of perceptron algorithm in prediction of undiagnosed patients. These datasets used in the system have many features but only two well-known risk factors of diabetes are selected to be used in this experiment that are age and body mass index. The proposed algorithm, ensembles boosting algorithm with perceptron algorithm, which uses more than one weight vector for the test data classification. The proposed algorithm is validated on three different NHANES datasets confirming that AUC value improves from 0.72 to 0.75 by the proposed algorithm.

In ‘Random Forest Algorithm for the Prediction of Diabetes’, K.VijayaKumar [7] proposed a system which can perform early prediction of diabetes for a patient with a higher accuracy by using the Random Forest algorithm in machine learning technique.

In ‘Prediction of Diabetes using Classification Algorithms’, Deepti Sisodia and Dilip Sisodia [8] proposed a model which can prognosticate the likelihood of diabetes in patients with maximum accuracy. Therefore three machine learning classification algorithms namely Decision Tree, SVM and Naive Bayes are used in this experiment to detect diabetes at an early stage. Results obtained show Naive Bayes outperforms with the highest accuracy of 76.30% comparatively other algorithms. They suggested the work can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms.

3. PROPOSED SYSTEM

The proposed system is a user friendly web page which contains a form to take the prediction parameters as user inputs in a web form. The system is designed in flask and the backend consists of a Neural network which trains on the preprocessed data and predicts the result based on user parameters. The system also consists of a user friendly chatbot to help with the basic needs.

A. Neural Network Architecture

Neural Network architecture used consists of three layers input, hidden and output layers. The input layer takes simple inputs, the hidden layers where all the processing actually happens through a system of connections characterized by weights and biases. The output of our Neural network will either predict 1 or 0.

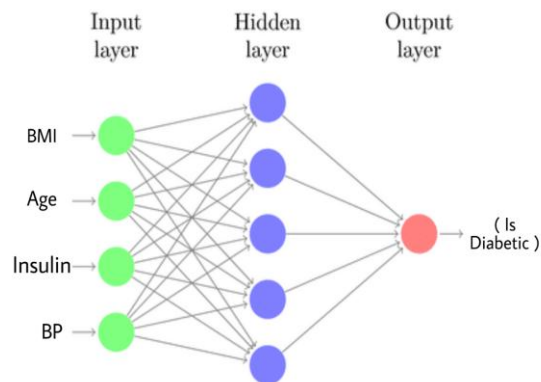


Fig. Neural Network Architecture for proposed systems

B. Backpropagation Algorithm

The Backpropagation algorithm takes the minimum value of the error function in weight space using a technique called the delta rule or gradient descent. The weights that minimize the error function is then considered to be a solution to the learning problem.

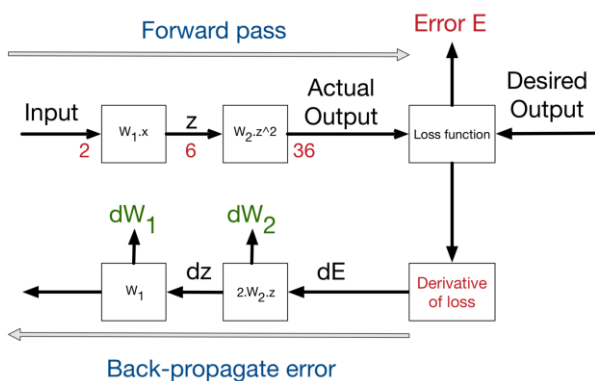


Fig. Working of Backpropagation Algorithm

C. Training Dataset

A labeled dataset that contains some vital measurements of patients (for example, age and blood insulin level), as well as a true label indicating the onset of diabetes in the patient sometime after the measurements were taken, then we can train a neural network on this data and use it to make predictions on new patients.

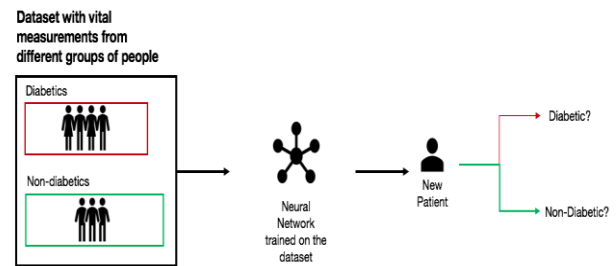


Fig. Proposed module for training dataset.

4. METHODOLOGIES

1. Data Preprocessing:

The data is from the PIMA Indian diabetes dataset [9] which is available in the csv format. The 8 parameters used are the number of times pregnant, Body mass index, plasma glucose, diastolic blood pressure, triceps skin fold thickness, diabetic pedigree function. After loading the data preprocessing is performed. Data preprocessing is the processing of a dataset in which data is transformed and encoded in a form such that the machine learning algorithm can parse it and only useful information is being extracted from the dataset. The values are then read sequentially for further training.

2. Training:

Training a Neural Network involves using an optimization algorithm to search weights to best match the input values. Inputs are fed to the input neurons and outputs are obtained by applying the initial weights. Second step is to calculate the error in the backpropagation step. The error is minimized until an optimized set of weights are obtained. Keras library in tensorflow is used for training the Neural Network. Hyperparameters are adjusted to achieve maximum accuracy. The trained weights are stored in a file for later predictions on new inputs.

3. Predictions

A GUI is developed in flask for user input values for the parameters. The input values are loaded in the backend and prediction is done using the keras library of tensorflow from the stored weight file. The result is then displayed on the Web page. Users can input the values in an interactive chatbot. The chatbot is developed in python and the responses are parsed using the Natural Language toolkit library in python. The chatbot shows the prediction using the same backend mechanism as the web page and the result is displayed in the chat window.

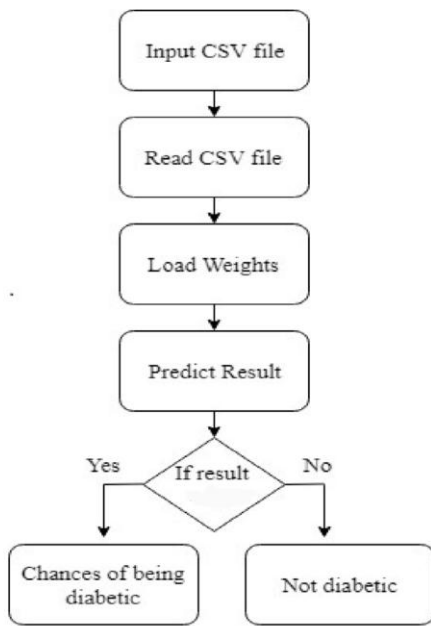


Fig. Flowchart of Diabetes prediction Model

5. RESULTS

The 8 features helped us in diabetes prediction. In order to improve the accuracy we have performed some data preprocessing steps. The hyperparameter tuning of the Neural network increases the overall accuracy Maximum accuracy over 85% is achieved on 12 hidden layer neurons and for 500 epochs based on the diagnostic nature of Indian Pima dataset. The GUI is provided to input the user values, the ANN model is able to predict whether a patient is diabetic or not. The result for the prediction is displayed on the web page. Users have been also provided an option of a user friendly and interactive chatbot. The chatbot asks questions to the users which acts as the input parameters for prediction. The result of the prediction is then displayed in the chat window.

6. CONCLUSION AND FUTURE SCOPE

Our proposed system eliminates the need to physically visit the clinic for diabetes diagnosis. As we have proposed and developed an approach for this using deep learning algorithms, it has significant potential in the field of medical science for the detection of various medical data accurately. In the future, the Diabetes prediction web application can be made more efficient in terms of feature importance and new standardized dataset and correctly predicting the outcome. Since we are using Neural Networks, research to develop new algorithms which are more efficient than currently existing algorithms. Implementing these in our system will greatly enhance the performance of the system. Diabetes prediction can be improved further if more features are taken into consideration. Finally, time efficiency can also be

increased for various applications and integrations with other applications.

REFERENCES

- [1] Diabetes. https://www.who.int/health-topics/diabetes#tab=tab_1.
- [2] S. Joshi and M. Borse, "Detection and Prediction of Diabetes Mellitus Using Back-Propagation Neural Network," 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), Ghaziabad, 2016, pp. 110-113, doi: 10.1109/ICMETE.2016.11.
- [3] R. E. Uhrig, "Introduction to artificial neural networks," Proceedings of IECON '95 - 21st Annual Conference on IEEE Industrial Electronics, Orlando, FL, USA, 1995, pp. 33-37 vol.1, doi: 10.1109/IECON.1995.483329.
- [4] R. Dey, V. Bajpai, G. Gandhi and B. Dey, "Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus," 2008 IEEE Region 10 and the Third international Conference on Industrial and Information Systems, Kharagpur, 2008, pp. 1-4, doi: 10.1109/ICIINFS.2008.4798367.
- [5] D. Shetty, K. Rit, S. Shaikh and N. Patil, "Diabetes disease prediction using data mining," 2017 International Conference on Innovations in Information, Embedded and Communication System (ICIIECS), Coimbatore, 2017, pp. 1-5, doi: 10.1109/ICIIECS.2017.8276012.
- [6] R. Mirshahvalad and N. A. Zanjani, "Diabetes prediction using ensemble perceptron algorithm," 2017 9th International Conference on Computational Intelligence and Communication Networks (CICN), Girne, 2017, pp. 190-194, doi: 10.1109/CICN.2017.8319383.
- [7] K. VijayaKumar, B. Lavanya, I. Nirmala and S. S. Caroline, "Random Forest Algorithm for the Prediction of Diabetes," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2019, pp. 1-5, doi: 10.1109/ICSCAN.2019.8878802.
- [8] Deepti Sisodia, Dilip Singh Sisodia, "Prediction of Diabetes using Classification Algorithms", International Conference on Computational Intelligence and Data Science (ICCIDS 2018).
- [9] Pima Indians Diabetes Database "https://www.kaggle.com/uciml/pima-indians-diabetes-database"