DIAGNOSIS OF DIABETES MELLITUS USING ARTIFICIAL NEURAL NETWORKS

Ms.Isha Vyas[1] ,Akash Patil[1], Ruchita Pame[1] ,Shruti Pingale[1]

Dept of Electronics and Telecommunication Engineering
Pimpri Chinchwad College Of Engineering , Pune, India

***Abstract*:- Diabetes is a disease that occurs when the body's glucose levels become too high. Diabetes should not be ignored; if addressed, it can lead to major problems such as damage to the eyes, blood pressure, kidneys, heart, and other bodily organs. Diabetes can be controlled if detected early. We will do early diabetes prediction in a human body or patient using a range of machine learning techniques to improve accuracy. methods for machine learning You can forecast occurrences more accurately by constructing models utilizing patient datasets. We will use ensemble approaches and machine learning classification on a dataset to predict diabetes in this work. These methods include Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF). The accuracy of each model differs when compared to other models. According to the research study, the model can reliably predict diabetes with a 95% confidence level or above.**

Keywords*: *Diabetes, Machine, Learning, Prediction, Dataset, Ensemble

I. INTRODUCTION

Diabetes mellitus is a condition defined by a metabolic abnormality and an abnormal rise in blood sugar levels brought on by a lack of insulin. Diabetes is one of the worst diseases there is. Obesity, a high blood glucose level, and other factors can cause diabetes. It changes the way insulin hormone functions, resulting in an abnormal metabolism in individuals with diabetes and increased levels of blood sugar. The World Health Organization estimates that 422 million people worldwide have diabetes, mostly in low- or middle-income nations. And up until the year 2030, this might be increased to 490 billion. As a result, data analytics might be used to provide an early diabetes prediction. Analyzing massive amounts of data to find hidden patterns and make conclusions is known as data analytics. Machine learning algorithms are used in the healthcare industry to analyze medical data and create machine learning models that can perform medical diagnosis. Diabetes prediction using machine learning classification and ensemble techniques. Machine learning is a technique used to intentionally train computers or other devices. By creating multiple categorization and ensemble models from the obtained information, various machine learning techniques efficiently capture knowledge. The information gathered may be helpful in predicting diabetes. Several machine learning approaches are capable of making predictions, but selecting the right method may be challenging. We thus use common classification and ensemble methods on a dataset for prediction in order to achieve this goal.

II. LITERATURE REVIEW

T. Jayalakshmi, Dr. A. Santhakumaran proposed A novel classification method for diagnosis of diabetes mellitus using artificial neural networks.This paper approached the aim of diagnoses by using ANNs and demonstrated the need for preprocessing and replacing missing values in the dataset being considered. Through the Modified training set, a better accuracy was achieved with lesser time required for training the set.[1]

J. Pradeep Kandhasamy, S. Balamurali proposed Performance Analysis of Classifier Models to Predict Diabetes Mellitus.This research study compares the performance of algorithms. Those are used to predict diabetes using data mining techniques. Authors compared four prediction models for predicting diabetes mellitus under two different situations. One is before pre-processing the dataset. Here the studies conclude that the decision tree J48 classifier achieves higher accuracy of 73.82 % than other three classifiers. After pre-processing, the dataset gave a more accurate result when compared to the previous studies. From this we can come to know that after removing the noisy data from our dataset it will provide good results for our problem.[2]

K. Pradeep, N. Naveen proposed Predictive analysis of diabetes using J48 algorithm of classification techniques.the performance of machine learning techniques was compared and measured based on their accuracy. The accuracy of the technique is varied from before pre-processing and after preprocessing as they identified in this study. This indicates that in the prediction of diseases the pre- processing of the data set has its own impact on the performance and accuracy of

the prediction. The algorithms used by this paper were five different algorithms GMM, ANN, SVM, EM, and Logistic regression. Finally, the researchers concluded that ANN (Artificial Neural Network) was providing High accuracy for prediction of Diabetes.[3]

III. PROPOSED METHODOLOGY

The following strategy can be implemented in order to develop an AIML-based deep learning-based diabetes prediction system:

The gathering of a sizable dataset consisting of medical information from diabetic patients is the initial step in the process of developing a Diabetes Prediction System by utilizing AIML deep learning. It is important that the dataset contains pertinent information such as age, gender, body mass index (BMI), blood pressure, cholesterol levels, and any other characteristics that can influence the onset of diabetes.

After the dataset has been acquired, it needs to be preprocessed, which involves deleting any data that is either missing or unnecessary, normalizing the data, and encoding categorical variables using either one-hot encoding or label encoding. The next thing that needs to be done is figuring out which characteristics have the strongest link to diabetes and eliminating any characteristics that have very little to no bearing on the outcome of the prediction.

Following that, the dataset needs to be partitioned into a training set and a testing set. The AIML model will be trained using the training set, and then its performance will be tested using the testing set. The training set will be used.

It is necessary to choose the proper AIML deep learning algorithm that is suited for diabetes prediction. Depending on the specifics of the issue at hand, the algorithms could include anything like K-Nearest Neighbors (KNN), Support Vector Machine (SVM), or Random Forest, amongst others.

The next stage, following the selection of the right algorithm, is to train the model using the dataset that will be used for training. Adjustments need to be made to the model's hyperparameters in order to get the greatest possible performance out of it. Some of these hyperparameters include the number of layers, nodes, and activation functions.

It is necessary to do an analysis of the performance of the model utilizing the testing dataset. The performance of the model can be evaluated based on a number of different metrics, including accuracy, sensitivity, specificity, and F1-score.

After the model has been educated and assessed, it needs to be implemented into a real-time setting so that it can make predictions based on newly collected data.

It is possible to increase the performance of the model by utilizing optimization strategies such as hyperparameter tweaking, regularization, and dropout. These strategies can be used. It is also vital to do routine monitoring and maintenance on the model in order to guarantee that it will continue to produce correct forecasts. It is essential to perform frequent reassessments of the model, and any necessary modifications should be made to the model.

By adhering to this methodology, it will be possible to construct an AIML-based deep learning Diabetes Prediction System, which will be capable of reliably predicting the onset of diabetes in patients.

IV. MODEL BUILDING

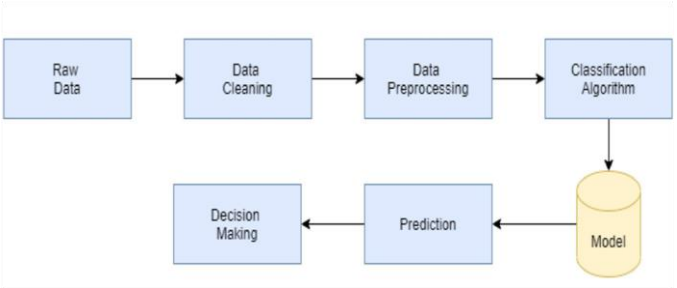


FIG IV.a: Flow Diagram For Data Processing And Model Creation

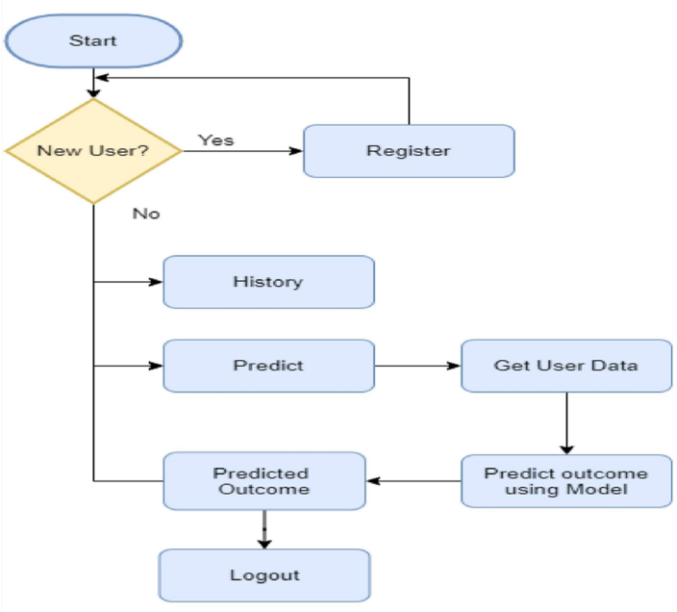
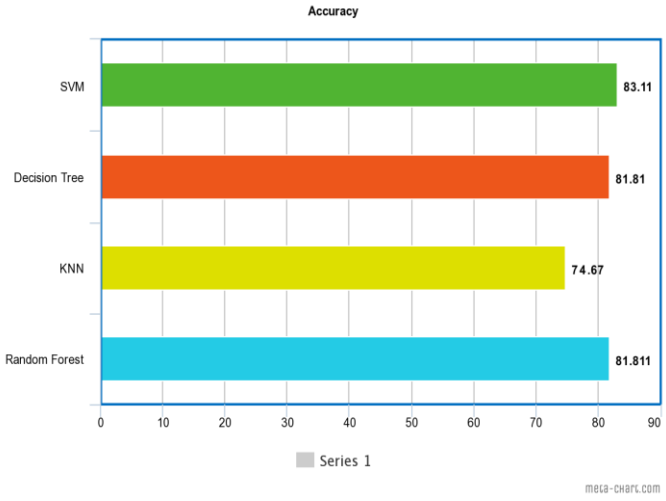
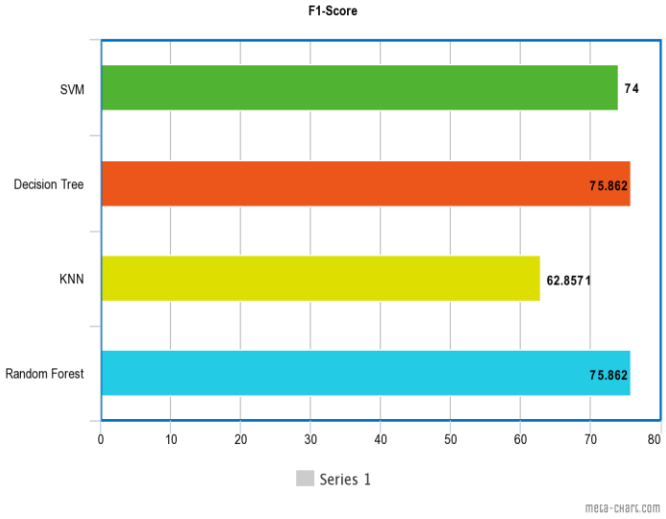


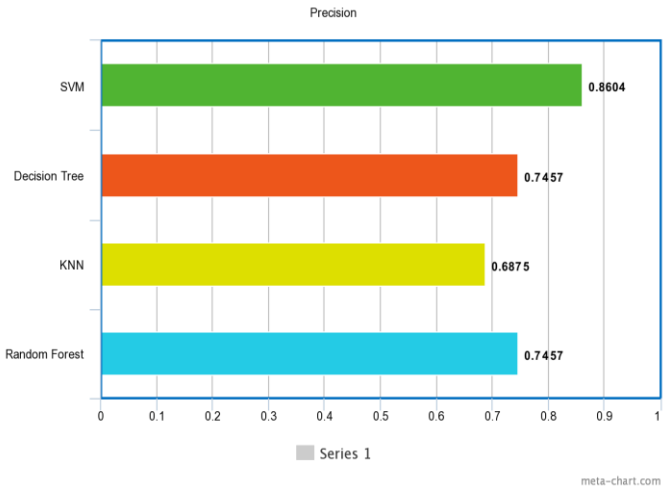
FIG IV.b:Flow Diagram for our system when a user Visits our Webapp

V. EXPERIMENTAL RESULTS

Dataset statistics		Variable types	
Number of variables	9	Numeric	8
Number of observations	768	Categorical	1
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	54.1 KiB		
Average record size in memory	72.2 B		







In our data, there are 8 independent variables and 1 dependent variable. There are a total of 768 observations with 500 as healthy and 268 as diabetic. From the data representation graph is it clear that the dataset has lots of zeros as missing values. To counter this, we can replace these values with the medians of their respective columns.

Secondly, the correlation matrix states that skin thickness and blood pressure contribute the least to the outcome. So, we can get rid of them to improve the accuracy of the model. Through Box Plots we can get to see a lot of mild and extreme outliers present in the dataset. This can be handled by eliminating these extreme outliers and replacing them with an upper bound or 0.75 quantile for that column.

Algorithms like KNN use distance for their prediction. So such models require uniform data which is inside a normalized range. An uniformed dataset may led to poor predictions. So we can use minmaxscaler or standardscaler to normalize our dataset. To prevent overfitting of the model, we will split our data into training and testing sets. This lowers the chances of overfitting in our model. We can also use cross validation or remove unwanted features to avoid overfitting.

For selecting a suitable model for the webapp, we would compare the accuracies of the model against the same input data.

Algorithm	Precision	F1-Score	Accuracy
Support Vector Machine	0.8604	74.0	83.11
Decision Tree	0.7457	75.8620	81.81
KNN	0.6875	62.8571	74.67
Random Forest	0.7457	75.8620	81.811

According to the graph above, Support Vector Machine, followed by Decision Tree, provides the best predictions for the processed data. Based on this graph, we can use the Support vector machine for our project. Support vector machine is an ensemble classifier constructed from numerous decision trees, where ensemble refers to the use of various machine learning algorithms to achieve predictive performance.

VI. CONCLUSION

This research's primary goal was to build and implement techniques for predicting diabetes using machine learning, and to assess the effectiveness of such approaches. The suggested method makes use of a variety of classifiers, including SVM, KNN, Random Forest, and Decision Tree. Moreover, a classification accuracy of 83% was attained. The experimental findings can help healthcare providers make early predictions and decisions to treat diabetes and save lives.

Early diabetes detection is essential for effective treatment. With the use of cutting-edge computational techniques and the availability of a sizable number of epidemiological and genetic diabetes risk datasets, machine learning has the potential to completely transform the capacity to forecast the risk of developing diabetes. The implementation of a software system that would take into account the different aspects that impact diabetes has been the project's main goal.

The structured dataset used in this study is a restriction, however in the future, we may also use the models with unstructured data. Because of the data, we are also unable to forecast the type of diabetes, therefore in the future, we will try to do so in order to make a better prognosis.

VII. REFERENCES

[1]A Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks | IEEE Conference Publication | IEEE Xplore

[2]Performance Analysis of Classifier Models to Predict Diabetes Mellitus —ScienceDirect

[3]Predictive analysis of diabetes using J48 algorithm of classification techniques |IEEE Conference Publication | IEEE Xplore

[4] Priyanka Indoria, Yogesh Kumar Rathore, "A Survey: Detection and Prediction of Diabetes Using Machine Learning Techniques" ,Vol. 7 Issue 03, March-2020, International Journal of Engineering Research & Technology (IJERT).

[5] L. Beqiri; A. Velinov; B. Fetaji; L. Loku; A. Buçuku; Z. Zdravev,"Analysis of Diabetes Dataset", 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO),IEEE.

[6] B. SRIDHARA MURTHY; J. SRILATHA, Comparative Analysis on Diabetes Dataset Using Machine Learning Algorithm, 02 August 2021,IEEE

[7] B. Nithya and Dr. V. Ilango," Predictive Analytics in Healthcare Using Machine Learning Tools and Techniques", International Conference on Intelligent Computing and Control Systems, 978-1-5386-2745-7, 2021