

Designing a Model to Detect Diabetes using Machine Learning

Ms. Komal Patil

Department of Computer Engineering
Datta Meghe College of Engineering,
Airoli, Navi Mumbai, INDIA

Dr. S. D. Sawarkar

Department of Computer Engineering
Datta Meghe College of Engineering,
Airoli, Navi Mumbai, INDIA

Mrs. Swati Narwane

Department of Computer Engineering
Datta Meghe College of Engineering,
Airoli, Navi Mumbai, INDIA

Abstract—Many of the interesting and important applications of machine learning are seen in a medical organization. The notion of machine learning has swiftly become very appealing to healthcare industries. The predictions and analysis made by the research community for medical dataset support the people by taking proper care and precautions by preventing diseases. Through a set of medical datasets, different methods are used extensively in developing the decision support systems for disease prediction. This paper explains various aspects of machine learning, the types of algorithm which can help in decision making and prediction. We also discuss various applications of machine learning in the field of medicine focusing on the prediction of diabetes through machine learning. Diabetes is one of the most increasing diseases in the world and it requires continuous monitoring. To check this we explore various machine learning algorithms which will help in early prediction of this disease.

Keywords—*Diabetes; health care; decision tree; machine learning; application; classification; approach; algorithm.*

I. INTRODUCTION

Multiple opportunities for healthcare are created because machine learning models have potential for advanced predictive analytics. There are already existing models in machine learning which can predict the chronic illness like heart disorder, infections and intestinal diseases. There are also few upcoming models of machine learning to predict non-communicable diseases, which is adding more and more benefit to the field of healthcare. Researchers are working on machine learning models that will offer very early prediction of specific disease in a patient which will produce effective methods for the prevention of the diseases. This will also reduce the hospitalization of patients. This transformation will be very much beneficial to the healthcare organisations. [1]

The most explored area is the healthcare system which uses modern computing techniques is in healthcare research. As mentioned above the researchers in the related fields are already working with the healthcare organisation to come up with more technology ready systems. Diabetes is a disease which reduces the body's capability to produce insulin. In other words the body can not retaliate to the hormone insulin production. This results in anomalous metabolism of carbohydrates and increased blood glucose levels. Early detection of diabetes becomes very important because of the reasons mentioned above. Many people in the

world are getting affected by diabetes and this number is increasing day by day. This disease can damage many vital organs hence the early detection will help the medical organisation in treatment of it. As the number of diabetic patients is more there is an excessive important medical information which has to be maintained. With the support of increasing technology the researchers have to build a structure that store, maintain and examine these diabetic information and further see feasible dangers. [4]

The blood glucose levels become too high in the body when there is diabetes. Glucose is created in the body after eating food. The hormone insulin produced in the body helps balance the glucose levels and regulate blood sugar levels, deficiency of insulin causes Diabetes. Type 1 diabetes is a scenario where the body does not produce insulin at all to balance the sugar levels in blood. Type 2 is a diabetes type where the body produces insulin but does not utilize this hormone completely to balance blood sugar levels. The Type 2 diabetes is most common one. There is something called as prediabetes, this is a situation where the person can have high glucose level but not that high that he/she can be said to have diabetes. But the people who have prediabetes are prone to get type 2 diabetes. This disease can cause serious damage to many vital organs in the body like kidneys, heart, nerves and eyes. If a woman gets this disease during pregnancy then it is known as gestational diabetes. By managing our weight, meal plan and exercise we can control diabetes. One should always keep a check on its blood sugar levels.

II. METHODOLOGY

In this section we shall learn about the various classifiers used in machine learning to predict diabetes. We shall also explain our proposed methodology to improve the accuracy. In section A we shall explain various classifiers and in section B we shall explain our proposed system.

A. Machine learning classifiers used in diagnosis of diabetes

The variation in glucose levels is cause of diabetes. Insulin balances the blood glucose level in the body, deficiency of which cause diabetes. For the prediction of diabetes machine learning is used, these have many steps like image pre-processing/data preprocessing followed by a feature extraction and then classification. We can use any of the mentioned machine learning classifiers to predict this disease. In the above section we have learning about many classification algorithms, we can either use any one of these to predict the disease or we can explore the techniques to use the hybrid methodology to improve the accuracy over using a single one. Currently, the researches have used the a single classification algorithm and have come up to accuracy of 70 to 80% for detection of the diabetes disease. [7][9]

Depending on the application and nature of the dataset used we can use any classification algorithms mentioned below. As there are different applications, we can not differentiate which of the algorithms are superior or not. Each of classifiers have its own way of working and classification. Let us discuss each of them in details.[5]

- **Naive Bayes Classifier:** This classifier can also be known as a Generative Learning Model. The classification here is based on Baye's Theorem, it assumes independent predictors. In simple words, this classifier will assume that the existence of specific features in a class is not related to the existence of any other feature. If there is dependency among the features of each other or on the presence of other features, all of these will be considered as an independent contribution to the probability of the output. This classification algorithm is very much useful to large datasets and is very easy to use. [14]
- **Logistic Regression:** Logic regression is used for Predictive Learning Model. To determine output in this classifier, we use a statistical method to analyse the dataset. These data set can have one or more than one independent values. The output is calculated with a data in which there could be two outputs. The aim of this classification algorithm is to find the relationship between the dichotomous category and predictor variables.[6][14]
- **Decision Trees:** This classification algorithm builds the regression models. These models are builded in form of structure which is similar to tree - a tree like structure is created by this classifier. It keeps on dividing the data set into subsets and smaller subsets which develops an associated tree, incrementally. The decision tree is finally created which has decision nodes and leaf nodes. In this tree the leaf node will have details about the classification or the decision taken for classification whereas the decision will have branches. The highest decision node which will be at the top of the tree will correspond to the root node. This will be the best predictor. [3][14]
- **Random Forest:** This classification algorithm are similar to ensemble learning method of classification. The regression and other tasks, work by building a group of decision trees at training data level and during the output of the class, which could be the mode of classification or prediction regression for individual trees. This classifier accuracy for decision trees practice of overfitting the training data set.[8][14]
- **Neural Network:** As the name suggests this classifier has units known as neurons, which are arranged in layers that convert the input vector to relevant output. Each single neuron takes an input, this is most often a non-linear input, this is given to a function which is then passed to next layer to get the output. The input given to the first layer will act as an output for the next layer and so on, thus this classification algorithm follows a feed-forward method. But in this method there is no feedback to the previous layer, so weighting are also given to the signals passing through the neurons and the layers, these signal then are turned into a training phase this eventually then become a network to handle any particular problem.[2][14]
- **Nearest Neighbor:** As the name suggests the nearest neighbour algorithm is based on the nearest neighbour and this classification algorithm is supervised. It is also called as k-nearest neighbour classification algorithm. A cluster of labeled points are used to understand how the other points should be labelled. For labelling a new point it checks the already labelled points which could be closest to the point to be labelled, i.e closest to the neighbour. In this way depending on the votes of the neighbour the new point is labelled the same label which most of neighbours have. In in algorithm 'k' is the number of neighbours which are checked.[5][14]
- **Support vector machine (SVM):** This is also one of the classification algorithm which is supervised and is easy to use. It can used for both classification and regression applications, but it is more famous to be used in classification applications. In this algorithm each point which is a data item is plotted in a dimensional space, this space is also known as n dimensional plane, where the 'n' represents the number of features of the data. The classification is done based on the differentiation in the classes, these classes are data set points present in different planes.
- **XGBoost:** Recently, the researches have come across an algorithm "XGBoost" and its usage is very useful for machine learning classification. It is very much fast and its performance is better as it is an execution of a boosted decision tree. This classification model is used to improve the performance of the model and also to improve the speed. [21]

We have already learned about all the machine learning classification algorithms and approaches used to predict the disease. After doing this survey we would be proposing to use more than one classification algorithm along with any of the learning approaches which will improve the prediction accuracy of the disease by more than 80%.

It is good to use the combination of more than 2 classifiers to get the desired accuracy. We shall be using Decision tree along with other classifiers, we shall design a model to evaluate the training data. We shall evaluate each of the classifier and either use XGBoost along with Decision tree/ RF/ SVM / Naive Bayes or we can use Decision Tree / RF along with the Naive Bayes. by using the combination mention in this section we shall improve the accuracy by more than 80%. [3][7]

B. Proposed System

The proposed system predicts the disease of diabetes in patients with maximum accuracy. We shall talk about various machine learning, the algorithm which can help in decision making and prediction. We shall use more than one algorithm to get better accuracy of prediction.

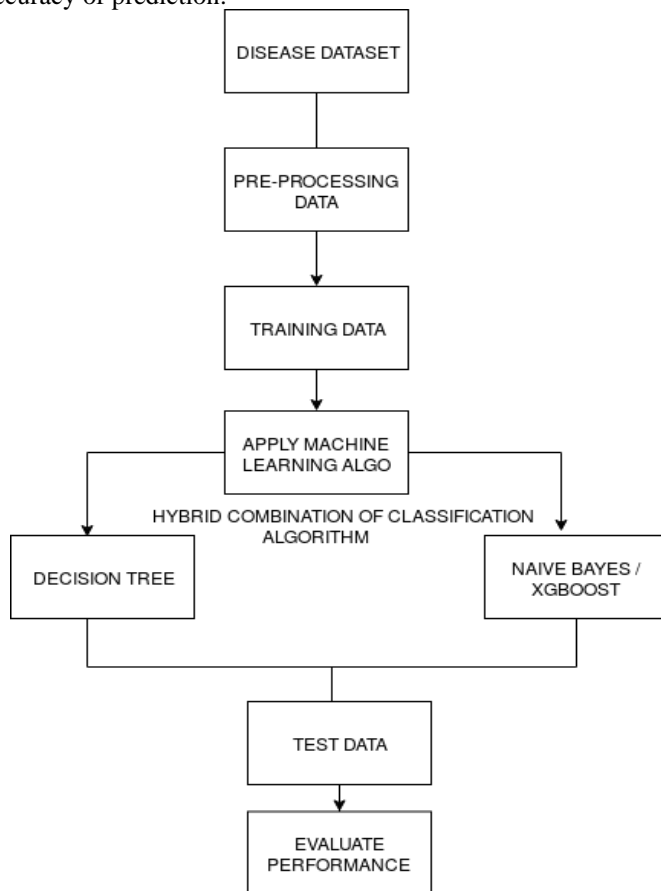


Fig 1. Proposed System Block diagram

The figure above **Fig. 1** explains the proposed work. The disease dataset is given to the system which is then pre-processed so that the data is in a useable format for analysis. If the dataset is not structured or if the dataset is huge or it has irrelevant features, we shall use feature extraction to extract the data. After this the data is trained and we apply a relevant machine learning algorithm to the dataset. The machine learning algorithms are already explained in Chapter 1. After this we use a combination of the classifier to get our desired result. This is also called a hybrid approach to test the data, in this method we propose to use the combination of two classifiers namely, Decision Tree and Support Vector Machine (SVM) or a combination of Decision Tree with XGBoost. We shall then test the data and evaluate the desired results. We shall now see the different

classifiers and discuss the hybrid combination used for our proposed system.

There are different types of classifiers, a classifier is an algorithm that maps the input data to a specific category. We have already listed and explained the different classifiers which can be used to achieve good accuracy.

After we train the model the most important aspect is to evaluate the classifier to verify its relevance. After understanding and studying each classifier in detail we propose to combine more than one classifier to get our accurate results. We shall evaluate each of the classifier and use XGBoost, Decision tree, RF, SVM, Naive Bayes and more by using the combination we shall improve the accuracy by more than 70%.

III. SYSTEM OVERVIEW

System design is used for understanding the construction of system. We have explained the flow of our system and the software used in the system in this section.

A. Flow of the system

The **Fig. 2** explains the flow chart of the system design, we shall explain each of the components of the flow chart in each section below. In Preprocessing we have done feature selection: Forward feature selection and Backward feature selection. We have provided the processed data to algorithm, we have used 5 techniques like ADA Boost, Decision Tree, XG Boost, Voting classifiers, and stacking classifier for predicting diabetes.

- **Dataset:** PIMA, Indian Diabetes dataset containing 768 cases. The objective is to predict based on the measures to predict if the patient is diabetic or not. The other dataset which we shall use will be data of all female patients to check if diabetic or not. Pima Indians Diabetes (PID) dataset of National Institute of Diabetes and Digestive and Kidney Diseases. PID is composed of 768 instances as shown in **Table 1**. Eight numerical attributes are represent each patient in data set.[22]

Table 1. Numerical attributes patient datasets

Sr. No	Attribute	Values
1	Number of times pregnant	[0-17]
2	Plasma Glucose Level	[0-199]
3	Diastolic Blood Pressure	[0-122]
4	Triceps skin-fold thickness	[0-99]
5	2 hour serum insulin	[0-846]
6	Body Mass Index	[0-67]
7	Diabetes Pedigree Function	[0-2.45]
8	Age	[21-81]
9	Class (Positive or Negative)	[0,1]

The proposed system has two main stages that will work together to get the desired results. At the first stage, the data is prepared, and at the second stage there is classification. However, the input into the system is the PID dataset and the output will be one class that represents the healthy or the diabetic. In the proposed system the input data is processed through a number of steps in order to improve the system performance. First of all, data reduction is applied on the input dataset to eliminate the noisy and inconsistent

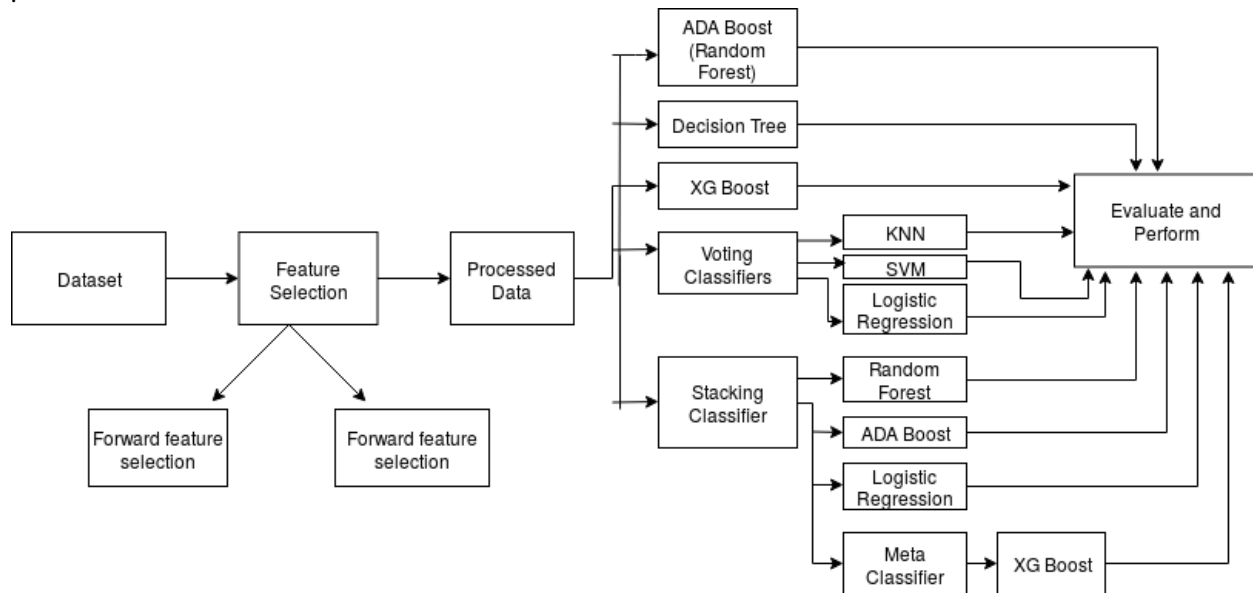


Fig 2. Flow diagram of the system

The main aim is to classify the data as diabetic or non-diabetic and improve classification accuracy. For many classification problems, the higher number of samples chosen but it doesn't leads to higher classification accuracy. In many cases, the performance of algorithm is high in the context of speed but the accuracy of data classification is low. The main objective of our model is to achieve high accuracy. Classification accuracy can be increased if we use much of the data set for training and few data sets for testing. This survey has analyzed various classification techniques for classification of diabetic and non-diabetic data. In this proposed system we use techniques like ADABOOST, Decision Tree classifier, XGBoost, voting classifier and stacking for implementing the Diabetes prediction system.[21]

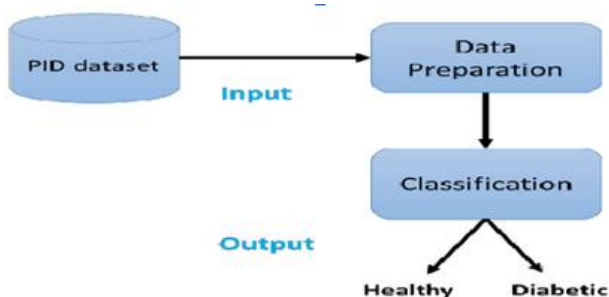


Fig 3. Main stages of the proposed system

We have already discussed the different classifiers in the above sections, we shall now discuss about stacking and voting classifier.

- **Stacking:** Stacking is an ensemble learning method that combines multiple base classification models

predictions into a new data set. This new data are taken as the input data for another classifier. This classifier employed to solve this problem. Stacking is often referred to as blending.[21]

On the basis of the arrangement of base learners, ensemble methods can be divided into two groups: In parallel ensemble methods, base learners are generated in parallel for example. Random Forest. In sequential ensemble methods, base learners are generated sequentially for example AdaBoost. On the basis of the type of base learners, ensemble methods can be divided into two groups: homogeneous ensemble method uses the same type of base learner in each iteration. heterogeneous ensemble method uses the different type of base learner in each iteration.

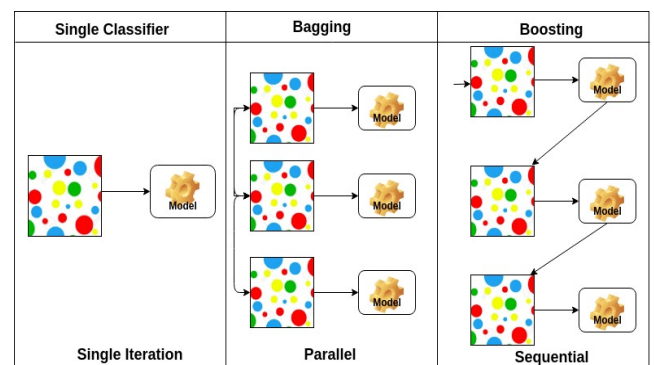


Fig 4. Stacking: Single , Parallel and Sequential learning Methods [21]

- **Voting Classifier:** The Ensemble Vote Classifier implements "hard" and "soft" voting. In hard voting, we predict the final class label as the class label that has been predicted most frequently by the classification models. In soft voting, we predict the class labels by averaging the class-probabilities (only recommended if the classifiers are well-calibrated).[23]

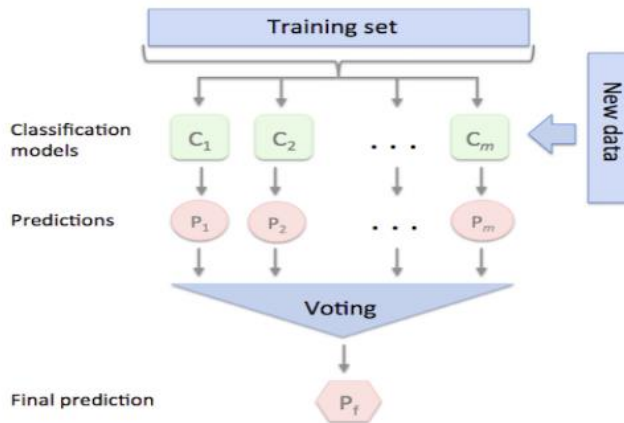


Fig 5. Voting Classifier [23]

IV. IMPLEMENTATION

This section provides knowledge about the implementation environment and throws light on the actual steps for the implementation of dataset to get better accuracy to predict diabetes by using different classifiers combination.

A. Hardware requirements

The following hardware was used for the implementation of the system:

- 4 GB RAM
- 10GB HDD
- Intel 1.66 GHz Processor Pentium 4

B. Software requirements

The following software was used for the implementation of the system:

- Windows 7
- Python 3.6.0
- Visual Studio Code

C. Implementation steps

In this section we shall discuss about the actual steps which were implemented while doing the experiment. We shall explain the stepwise procedure used to analyse the data and to predict the data accuracy for prediction of diabetes. The system consists of the following main steps:

We have selected a diabetic dataset named PIMA Indian Diabetes Dataset which consists of 768 instances classified into two classes : diabetic and non-diabetic with eight different risk factors: number of times pregnant , plasma glucose

concentration of two hours in an oral glucose tolerance test , diastolic blood pressure, triceps skin fold thickness, two – hour serum insulin , body mass index , diabetes pedigree function and age.

Feature Selection is the procedure where we automatically or manually select those features which contribute most to your prediction variable or output you are interested in. If there is irrelevant features in our data then it can decrease the accuracy of the models.

1. We are taking a diabetic dataset which is PIMA Indian dataset.
2. For pre-processing step, the system uses Feature selection method : Forward feature selection and Backward Feature selection. We train five different classifiers and decide which classifier provides high accuracy. We have used these classifiers which are ADABOOST , Decision Tree , XGBoost, Voting Classifier , Stacking Classifier.
3. Stacking Classifier uses Random Forest , ADABOOST and Logistic Regression as its base classifiers and XGBoost as its meta classifier.
4. We found Adaboost and Stacking Classifier to be the best out of all the five classifiers in the aspects of accuracy, since they give better accuracy.
5. Below are the screenshots to better understand the flow of our implementation steps and the desired results graphs. We shall show the stepwise image for the ADABOOST classifier. We have done similar steps for Decision Tree, XG Boost, Voting and stacking classifiers.

The Fig no. 6 explains the confusion matrix using ADABOOST.

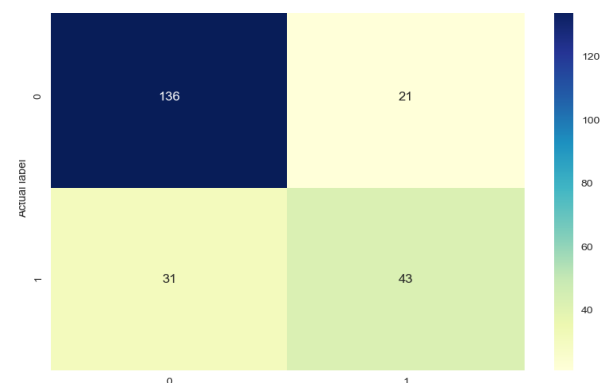


Fig 6. Confusion matrix: ADABOOST

We get the ROC curve after implementing the ADABOOST classifier. Please see Fig. 7 for the reference.

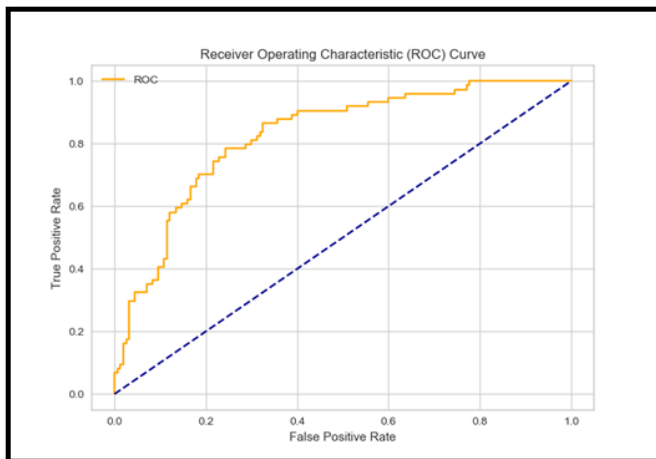


Fig 7: ROC curve: ADABOOST

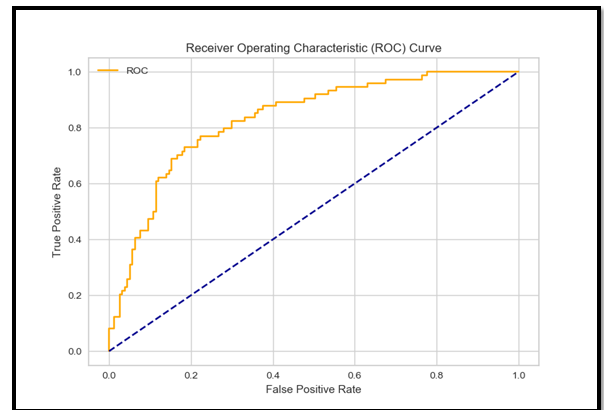


Fig10: ROC curve: XGBoost

We then explain the classification report which we get after implementation of the ADABOOST classifier.

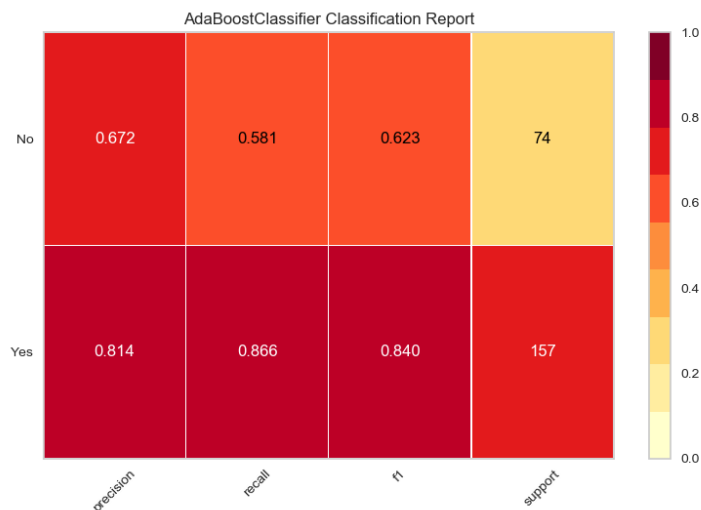


Fig 8: ADABOOST Classification Report

We have used similar steps and got ROC curves for Decision Tree, XGBoost, Voting and stacking classifiers. See the below screenshots for the same.

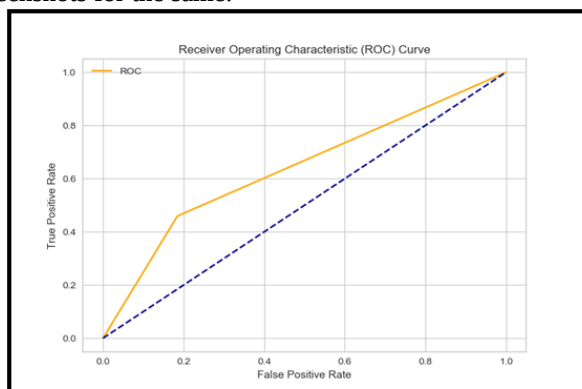


Fig9: ROC curve: Decision Tree

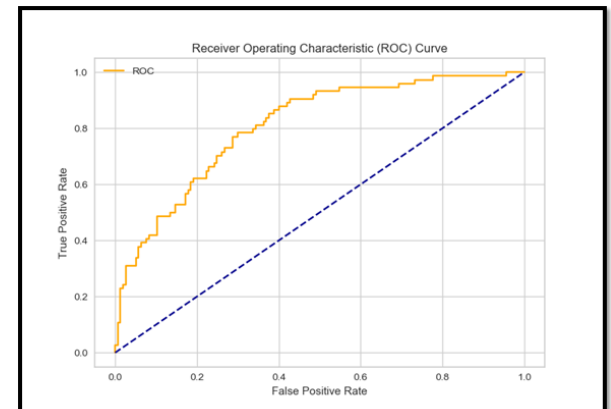


Fig11: ROC curve: Voting Classifier

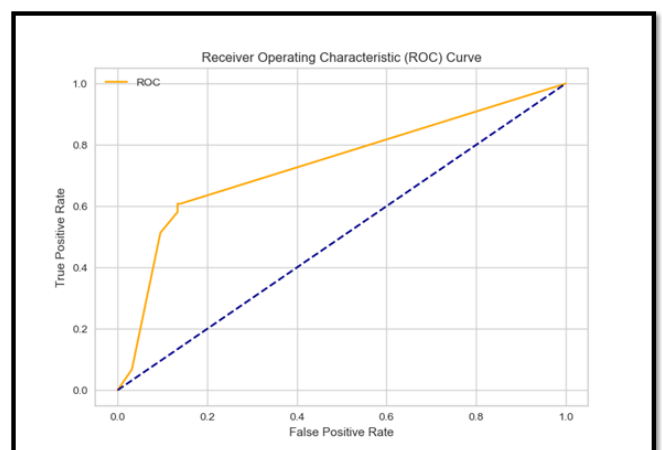


Fig12: ROC curve: Stacking Classifier

We then test the patients for diabetes. The below screen shows the test result of a diabetic patient where it is detected to be Diabetic.

Fig13: Test screen Diabetes Detected

V. OBSERVATIONS AND RESULTS

A. Observation

As discussed in the earlier sections, we have used five different classifiers to predict and improve the accuracy of the disease diabetes, Comparisons of these classifiers have been shown below in the accuracy table.

Table 2: Observation Table

Classifier	Precision	Recall	AUC	F1	Accuracy
ADA BOOST	0.82	0.90	0.83	0.86	80.08
Decision Tree	0.80	0.84	0.70	0.82	75.32
XGBOOST	0.80	0.90	0.83	0.84	77.48
Voting (KNN,SVM, Logistic Regression)	0.77	0.91	0.83	0.84	75.75
Stacking (Random Forest , ADABOOST , Logistic Regression)	0.82	0.90	0.75	0.86	80.08

B. Results

We have got the desired results of more than 80% accuracy for prediction of diabetes by using five different classifiers. Refer the graphs in Fig.14 and Fig.15 for the results. In the graph, shown in Fig. 14 shows the AUC, precision, recall and the F1 score obtained by using different classifiers. The graph shown in Fig 15. explains about the accuracy obtained by using different classifiers in the histogram graphical representation.

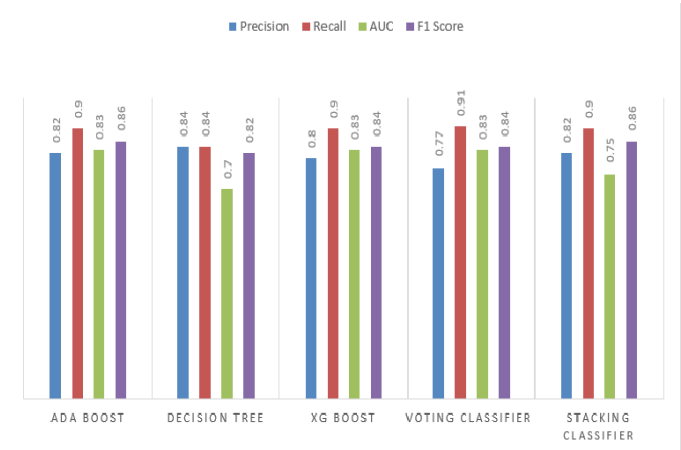


Fig.14: Graph of AUC, Precision, Recall and F1score

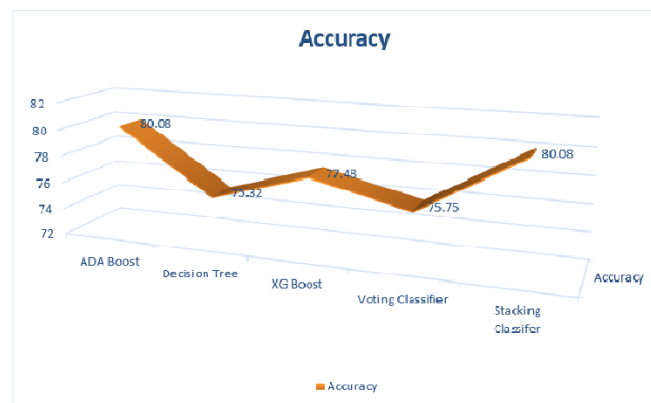


Fig 15. Result

VI. CONCLUSION AND FUTURE SCOPE

A. Conclusion

The machine learning methods can support the doctors to identify and cure diabetic diseases. We shall conclude that the improvement in classification accuracy helps to make the machine learning models get better results. The performance analysis is in terms of accuracy rate among all the classification techniques such as decision tree, logistic regression, k-nearest neighbors, naive bayes, and SVM , random forest , adaboost , xgboost. We have also seen that the accuracy of the existing system is less than 70% hence we proposed to use a combination of classifiers known as Hybrid Approach. Hybrid approach takes advantage by aggregating the merits of two or more techniques. We have found that our system provides us with 75.32 % of accuracy for Decision Tree Classifier, 77.48% accuracy for XGBoost Classifier, 75.75 % accuracy for Voting Classifier and finally 80 percentage of accuracy when using Stacking Classifier and ADA Boost. We have therefore found that the best among all the above classifiers is Stacking Classifier and Adaboost.

B. Future scope

In future, if we get a large set of diabetic dataset we can perform comparative analysis for analyzing the performance of each algorithm as well as the Hybrid algorithm so that the best one can be applied for predictive analysis. A particular method to identify diabetes is not very sophisticated way for initial diabetes detection and it is not fully accurate for predicting diseases. That's why we need a smart hybrid predictive analytics diabetes diagnostic system that can effectively work with accuracy and efficiency. We can use data mining, neural network for exploring and utilizing to support medical decision, which improves in diagnosing the risk for pregnant diabetes. Due to the dataset we have till date are not upto the mark, we cannot predict the type of diabetes, so in future we aim to predicting type of diabetes and explore it, which may improve the accuracy of predicting diabetes. We can also study the causes of diabetes and how to avoid having diabetes.

REFERENCES

- [1] Kaur, H., & Kumari, V. (2018). Predictive modelling and analytics for diabetes using a machine learning approach. *Applied Computing and Informatics*.
- [2] Carter, J. A., Long, C. S., Smith, B. P., Smith, T. L., & Donati, G. L. (2019). Combining elemental analysis of toenails and machine learning techniques as a non-invasive diagnostic tool for the robust classification of type-2 diabetes. *Expert Systems with Applications*, 115, 245-255.
- [3] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, 15, 104-116.
- [4] Mahmud, S. M., Hossin, M. A., Ahmed, M. R., Noori, S. R. H., & Sarkar, M. N. I. (2018, August). Machine Learning Based Unified Framework for Diabetes Prediction. In *Proceedings of the 2018 International Conference on Big Data Engineering and Technology* (pp. 46-50). ACM.
- [5] Patil, R., & Tamane, S. (2018). A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Diabetes. *International Journal of Electrical and Computer Engineering*, 8(5), 3966.
- [6] Dagliati, A., Marini, S., Sacchi, L., Cogni, G., Teliti, M., Tibollo, V., ... & Bellazzi, R. (2018). Machine learning methods to predict diabetes complications. *Journal of diabetes science and technology*, 12(2), 295-302.
- [7] Barik, R. K., Priyadarshini, R., Dubey, H., Kumar, V., & Yadav, S. (2018). Leveraging machine learning in mist computing telemonitoring system for diabetes prediction. In *Advances in Data and Information Sciences* (pp. 95-104). Springer, Singapore.
- [8] Choudhury, A., & Gupta, D. (2019). A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques. In *Recent Developments in Machine Learning and Data Analytics* (pp. 67-78). Springer, Singapore.
- [9] Samant, P., & Agarwal, R. (2017). Diagnosis of diabetes using computer methods: soft computing methods for diabetes detection using iris. *Threshold*, 8, 9.
- [10] Dankwa-Mullan, I., Rivo, M., Sepulveda, M., Park, Y., Snowdon, J., & Rhee, K. (2019). Transforming diabetes care through artificial intelligence: the future is here. *Population health management*, 22(3), 229-242.
- [11] Joshi, T. N., & Chawan, P. M. Diabetes Prediction Using Machine Learning Techniques.
- [12] Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *Jama*, 319(13), 1317-1318.
- [13] Nnamoko, N., Hussain, A., & England, D. (2018, July). Predicting Diabetes Onset: an Ensemble Supervised Learning Approach. In *2018 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1-7). IEEE.
- [14] Yadav, B., Sharma, S., & Kalra, A. (2018). Supervised Learning Technique for Prediction of Diseases. In *Intelligent Communication, Control and Devices* (pp. 357-369). Springer, Singapore.
- [15] Joshi, R., & Alehegn, M. (2017). Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach. *International Research Journal of Engineering and Technology*, 4(10).
- [16] Singh, D. A. A. G., Leavline, E. J., & Baig, B. S. (2017). Diabetes Prediction Using Medical Data. *Journal of Computational Intelligence in Bioinformatics*, 10(1), 1-8.
- [17] Gujral, S. (2017). Early diabetes detection using machine learning: a review. *Int. J. Innov. Res. Sci. Technol*, 3(10), 57-62.
- [18] Zia, U. A., & Khan, N. (2017). Predicting Diabetes in Medical Datasets Using Machine Learning Techniques. *International Journal of Scientific & Engineering Research Volume*, 8.
- [19] Naqvi, B., Ali, A., Hashmi, M. A., & Atif, M. (2018). Prediction Techniques for Diagnosis of Diabetic Disease: A Comparative Study. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND NETWORK SECURITY*, 18(8), 118-124.
- [20] Chen, J. C. H., Kang, H. Y., & Wang, M. C. (2018). Integrating Feature Ranking with Ensemble Learning and Logistic Model Trees for the Prediction of Postprandial Blood Glucose Elevation. *J. UCS*, 24(6), 797-812.
- [21] <https://www.greycampus.com/opencampus/machine-learning/different-types-of-classi>
- [22] <https://medium.com/@Mandysidana/machine-learning-types-of-classification-9497bd4>
- [23] <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623>
- [24] <https://en.wikipedia.org/wiki/Machinelearning>
- [25] <https://www.analyticsvidhya.com/blog/2016/02/7-important-model-evaluation-error-metric/>
- [26] <https://www.sciencedirect.com/science/article/pii/S2405959518304624>
- [27] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6232260/>
- [28] https://link.springer.com/chapter/10.1007/978-981-13-3140-4_42
- [29] <https://towardsdatascience.com/machine-learning-for-diabetes-562dd7df4d42>
- [30] <https://becominghuman.ai/diabetes-prediction-with-deep-learning-studio-a-differe>
- [31] <https://pdfs.semanticscholar.org/2c3a/6609a76762e3d40bdd90d8c07fa714d611fa.pdf>
- [32] <http://gjet.gyancity.com/Vol4No2/machine.pdf>
- [33] <https://intersog.com/blog/how-machine-learning-will-change-healthcare/>
- [34] <http://www.diabetes.org/diabetes-basics/gestational/what-is-gestational-diabetes.html>
- [35] <https://www.healthcatalyst.com/clinical-applications-of-machine-learning-in-healthcare>
- [36] <https://www.oodestechnologies.com/blogs/Medical-And-Healthcare-Applications-of-Machine-Learning/>