Techniques of Machine Learning for the Purpose of Predicting Diabetes Risk in PIMA Indians

Bhukya Madhu[1], Veerender Aerranagula[2], Riyaz Mahomad[3], V.Ravindernaik[4], K. Madhavi[5] and Gopal Krishna[6]

[1]Department of CSE- Data Science, KG Reddy College of Engineering & Technology, Hyderabad, Telangana, India.

[2]Department of Computer Science & Engineering, CMR Technical Campus(A), Hyderabad, India

[3]School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India

[4]Department of Computer Science and Engineering, CMR Technical Campus(A), Hyderabad, India

[5]Professor, Department of Computer Science and Engineering, GRIET, Bachupally, Hyderabad, Telangana

[6]Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, 248007

Abstract. Chronic Metabolic Syndrome Diabetes is often called a "silent killer" due to how little symptoms appear early on. High blood sugar occurs in people with diabetes because their bodies have a hard time maintaining normal glucose levels. Care for a recurrent sickness would be permanent. The two most common forms of diabetes are type 1 and type 2. A better prognosis can help reduce the high risk of developing diabetes. In order to better predict the likelihood that a PIMA Indian may develop diabetes, this study will use a machine learning-based algorithm. The demographic and health records of 768 PIMA Indians were used in the analysis. Standardisation, feature selection, missing value filling, and outlier rejection were all parts of the data preparation process. Machine learning techniques such as logistic regression, decision trees, random forests, the KNN model, the AdaBoost classifier, the Naive Bayes model, and the XGBoost model were used in the study. Accuracy, precision, recall, and F1 score were the only metrics utilised to assess the models' efficacy. The results demonstrate that. The results of this study reveal that diabetes risk may be reliably predicted using machine learning-based models, which has important implications for the early detection and prevention of this illness among PIMA Indians.

Keyword: Pima Indians, Diabetes, Machine Learning, Data Pre-processing, Feature Selection, Normalization, Early Diagnosis, Prevention.

Corresponding author: madhu0525@gmail.com

1 INTRODUCTION

Millions of people all over the world suffer from diabetes, a persistent illness. It has become a major threat to public health since it can lead to blindness, kidney failure, and other potentially fatal conditions, such as cardiovascular disease. Accurate diabetes prediction and early identification can lead to successful treatment and the avoidance of diabetes-related problems. Predicting and diagnosing medical diseases, such as diabetes, using healthcare data and machine learning algorithms has showed promising results in recent years. Machine learning can be used to predict diabetes by building models that evaluate a person's features and risk factors to determine the probability that they will get the disease. Machine learning models may sift through large datasets, intricate algorithms, and statistical analysis to identify trends and correlations that could otherwise go unnoticed by humans. One potential application of machine learning for diabetes prediction is the development of trustworthy and precise models that would enable doctors to better tailor patient care to each person's specific needs [5]. These models can help find people who are at high risk for developing diabetes and then help them reduce that risk through preventive measures like changing their lifestyle. Overall, the application of machine learning to Diabetes prediction is a significant step forward in healthcare analytics. Clinicians are provided with powerful tools to detect and treat the problem early, with the potential side effect of increasing the bar for therapy. PIMA Indians have the greatest prevalence of diabetes in the world, which disproportionately impacts Native American groups. The Indian Health Service found that more than half of the PIMA Indians over the age of 35 suffered from type 2 diabetes [7]. Several causes, both genetic and environmental, have been implicated in the disproportionately high rates of diabetes among PIMA Indians. The management of diabetes and the prognosis for patients are greatly improved by early detection and treatment. This can be accomplished with the use of predictive algorithms that isolate those who are most likely to become ill. Recently, machine learning algorithms have been applied to healthcare datasets consisting of clinical and demographic data, and their ability to predict the risk of Diabetes has been found to be quite promising [10]. The purpose of this investigation was to create a machine learning-based diabetes risk assessment model for PIMA Indians.

2 LITERATURE REVIEW

Native American communities have been classified as a high-risk group for diabetes, which has been on the rise worldwide. PIMA Indians have one of the highest diabetes rates of any people group. Genetics, lifestyle, and obesity have all been identified as major risk factors for diabetes in PIMA Indians. Predictive models for a variety of diseases, including Diabetes, are being developed with the help of machine learning techniques in the healthcare industry. Machine learning algorithms have been studied in recent years for their ability to predict diabetes risk in different populations [16]. Artificial intelligence was employed by Kavakiotis et al. (2017) to develop diabetes prediction algorithms for a European population. Using a variety of machine learning methods, the research was able to obtain an accuracy of up to 83%.

Numerous researches have looked at the possibility of developing diabetes in Native American populations using machine learning methods. The likelihood of developing diabetes in the Navajo population was predicted using a model built using machine learning by Lekoubou et al. (2019). The study had a 79.3% rate of accuracy thanks to the

use of data from electronic health records. Other studies have looked into the PIMA Indian population and the use of machine learning algorithms for predicting diabetes risk. To predict diabetes risk among PIMA Indians, Pashay et al. (2020) created an algorithm based on machine learning. The research made use of a dataset consisting of 766 PIMA Indians, with a best accuracy of 75.3%. Hasan, Md. Kamrul, Md. Ashraful, Das, Dola, Hossain, Eklas, and Hasan, Mahmudul [17] proposed a method of using Machine Learning Classifiers (KNN, Decision Tree, Random Forest, Navies Bayes) to predict the prevalence of diabetes in 2020. Preprocessing on a PID data set reduced kurtosis and skewness in the attribute distribution. Due to the low correlation between the base classifiers, the assembly of the two boosting classifiers for the combination of AB & XB yielded the best prediction. When implemented, the best accuracy of 95% was achieved by the combination of the two boosting classifiers for the AB & XB type. They suggested that in the future, trained models may be used to evaluate the transferability of disease prediction frameworks between medical disciplines and to design user-friendly web apps. Aishwarya Mujumdar and Vaidehi V [16] explored the use of machine learning methods to foresee diabetes in 2019. In order to make a more accurate prediction, they included in additional variables such as glucose, body mass index, age, insulin, and others. Similarly, we employed a pipeline architecture. Logistic regression, gradient boost classifier, LDA, AdaBoost classifier, different trees classifier, gaussian navies Bayes, bagging, random forest, decision tree, perception, SVM, and KNN are only some of the techniques that have been used. Amongst all methods, the Adaboost classifier achieved the best accuracy (98.8%), with logistic regression coming in second (96%) most accurately. More study is needed to determine the risk factors for the disease in the general population. Using machine learning algorithms in medicine, Sarwar, Muhammad Azeem Kamal, Nasir Hamid, Wajeeha Shah, and Munam Ali Shah [15] conducted research on diabetes prediction. Several algorithms have been put to the test using a patient's medical records to determine which one produces the most meaningful outcomes. Several methods, including K Nearest Neighbours (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), and Random Forest (RF), are utilised in this strategy. In this scenario, algorithms were evaluated using a set of eight characteristics taken from the PIMA Indian dataset. SVM and KNN are the most accurate algorithms with a 77% success rate. Key areas for future research include improving the model by incorporating new methods and testing it on larger datasets. A talk titled "Diabetes Mellitus Prediction Using Machine Learning Methods" [14] was delivered by Zou, Quan, Qu, Kaiyang, Luo, Yamei, Yin, Dehui, Ju, Ying, and Tang in 2018. The research was supported by the employment of techniques like decision trees, radio frequency (RF), and neural networks as classifiers.

The highest accuracy we get using random forest is 80%. With the correct data mining techniques, attributes, and classifiers, we were able to get our best result on the Pima Indian dataset, demonstrating that machine learning can predict diabetes. No communicable Disease Risk Assessment and Prediction Ferdousi, Rahatara, Hossain, M. Anwar and Saddik, Abdulmotaleb performed CPS analysis using machine learning. El [22] used methods such Bagging, AdaBoost, RT, Logistic Regression, SVM (Poly Kernel), and Navies Bayes to improve the accuracy of the machine learning classification algorithms RF and RT to 94%. Sivaranjani, S., Ananya, S., Aravinth, J., and Karthika, R. [26] used machine learning techniques that included feature selection and dimensionality reduction. Predictions of diabetes were studied in 2021.SVM and Random Forest are two algorithms that can predict the likelihood of developing complications from diabetes. To identify the characteristics that contribute to the prediction, our model iteratively selects the top four features using a combination of forward and backward feature selection. After taking a forward step, we perform research on the method and precision of PCA dimensionality reduction. Dimensionality reduction increases the size of the test set by 83% and improves feature selection in RF classifiers by 80%. After using a step-backward approach to feature selection, RF and SVM classifiers achieve 83.3% and 81.4% accuracy [28]. To improve the dataset's dimensionality reduction relevance and the RF model's predictive power, its

size can be raised in the future. Diabetes Prediction Using Diverse Machine Learning Methods, begun in 2019 [27], is the work of Priyanka Sonar, JayaMalini, K. In this article, the author builds a model to estimate the probability of developing diabetes as accurately as possible using statistical learning techniques such as support vector machines (SVM), naïve Bayes (NB), decision trees (DT), and artificial neural networks (ANN). Our decision tree accuracy, naive Bayes accuracy, and SVM accuracy scores were 85.1%, 77%, and 77.3%, respectively, showing a very precise procedure. The body of literature indicates that machine learning techniques are essential for forecasting diabetes risk in a number of demographics, including Native American tribes. Predictive models for diabetes need to be evaluated for performance and the best feature selection methods and algorithms found in order to be effective. If nothing else, stick with the current labelling.

3 CONTRIBUTION OF WORK

The primary result of this effort is a model based on machine learning that can predict the likelihood that a PIMA Indian will develop diabetes. Given that PIMA Indians have some of the highest diabetes rates in the world, this contribution is crucial, as machine learning techniques may help in the early identification and prevention of diabetes among this community. Three different machine learning strategies were employed in the study, including logistic regression, decision trees, and random forests, using a dataset of 768 PIMA Indians. Researchers were able to choose the most effective algorithm by evaluating its relative performance against the others. The data was preprocessed to account for outliers and missing values, select useful characteristics, and standardise the results. Data utilised in the study was meticulously prepared to ensure the accuracy and reliability of the machine learning models. The study evaluated the models' efficacy using a variety of metrics, including accuracy, precision, recall, and F1 score.

The models performed satisfactorily, with an accuracy of 81.70 percent using the random forest technique. Several benefits could accrue from the creation of this prediction model. Early diagnosis and prevention of diabetes in the high-risk PIMA Indian population may improve patient outcomes. Second, it sheds light on the feasibility of using machine learning algorithms to forecast the prevalence of diabetes in high-risk populations. The importance of feature selection and data preprocessing in creating robust predictive models is also highlighted by the research. A reliable and precise machine learning-based model for estimating diabetes risk among PIMA Indians is the project's last contribution. The most important result of this study is the creation of a reliable machine learning-based model for predicting diabetes risk among PIMA Indians. In this high-risk population, the research design may have a major effect on patient outcomes and the rate of diabetes.

4 METHODOLOGY

The project's methodology included several steps, each of which was essential in creating a machine learning-based model for estimating the risk of Diabetes among PIMA Indians.

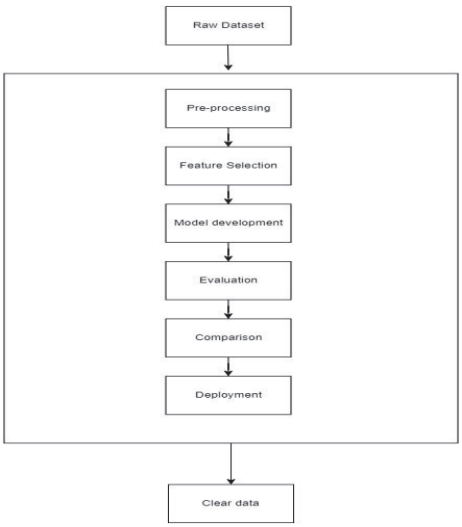

Fig 1 System Architecture

1. **Data collection:** The National Institute of Diabetes and Digestive and Kidney Diseases provided a dataset of 768 PIMA Indian women aged 21 and older as part of the methodology's initial stage. Age, BMI, blood pressure, insulin, glucose, blood pressure, and an indicator variable for Diabetes were among the many factors in the dataset.
2. **Data preprocessing:** Preprocessing the data was required in the methodology's second phase to manage missing values and outliers, choose pertinent features, standardize, and normalize the data. The data was preprocessed by the researchers using a variety of procedures, including mean imputation, median imputation, and outlier elimination. Additionally, they standardized the data to ensure that the scale for each variable was the same.
3. **Feature selection:** The most important features for predicting the risk of Diabetes were chosen in the third step of the process. The researchers determined the most crucial traits using a correlation matrix and recursive feature elimination (RFE) methods. Age, BMI, blood pressure, blood sugar, insulin, and blood pressure were chosen features.
4. **Model development:** The methodology's fourth phase required creating three machine learning models, including random forest, decision trees, and logistic regression. The researchers used the preprocessed and chosen features to train the models.
5. **Model evaluation:** The methodology's fifth phase entailed assessing the models' performance using various criteria, including accuracy, precision, recall, and F1 score. The researchers also utilized cross-validation to ensure that the models balanced the data.
6. **Model comparison**: The methodology's sixth phase was comparing the performance of the various models and choosing the most successful one based on the assessment criteria.
7. **Model deployment:** The methodology's last stage involved using the chosen model to forecast the likelihood of Diabetes among the PIMA Indian population.

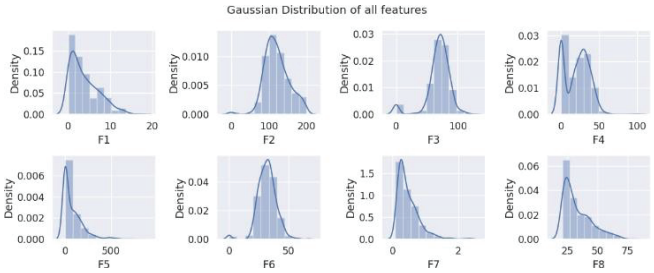

Fig 2 Raw Data Plot and Presentation

The diagonal displays the dataset's distribution and kernel density graphs for both classes. No attribute can clearly distinguish between the two outcomes, according to the scatter plots that indicate the relationship between every feature or characteristic when taken in pairs.

Distribution Plot:

Using the Dist define tool, we may easily depict a univariate distribution of observations.

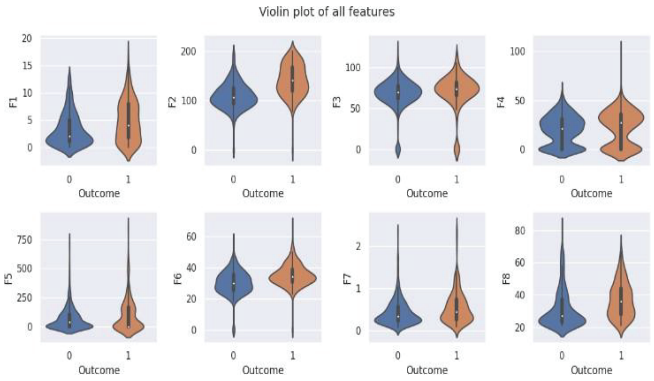

Fig 3. Distribution of diabetic plot

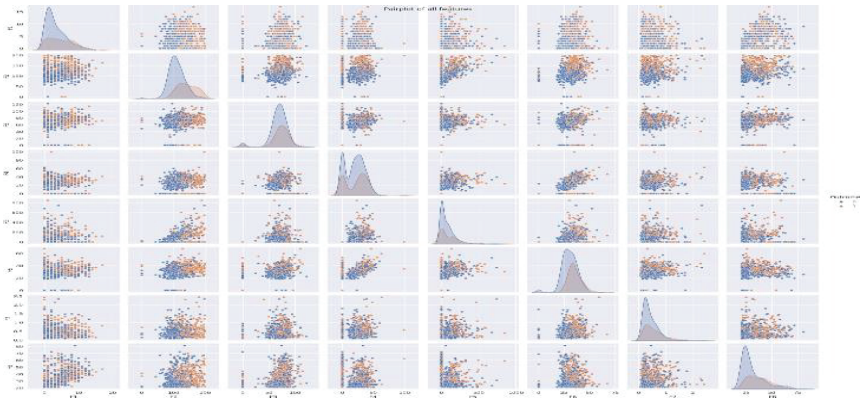

Fig 4. Violin of diabetic plots

Violin Plots:
Plotting numerical data can be done with a violin plot. It resembles a box plot with a kernel density plot rotated 90 degrees on either side. Like box plots, violin plots display the data's probability density at various levels (a histogram is the most common illustration of this).

Correlation:
Both positive and negative correlations indicate that changes in the value of one variable have a corresponding effect on the other. In addition, the correlation may be zero, indicating that the variables are unrelated, or it may be high.

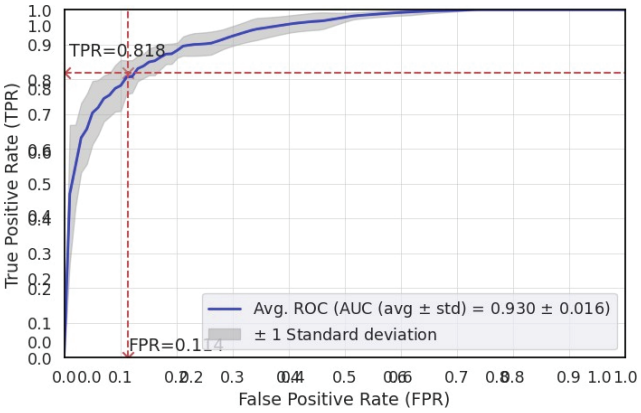

Fig 5. Correlation of diabetic

K-Nearest Neighbors (KNN) is a straightforward and user-friendly machine-learning technique for classification and regression tasks. "K" in KNN represents the number of neighbours evaluated for prediction. KNN is a straightforward and effective algorithm for making predictions based on the characteristics of similar examples in a dataset.

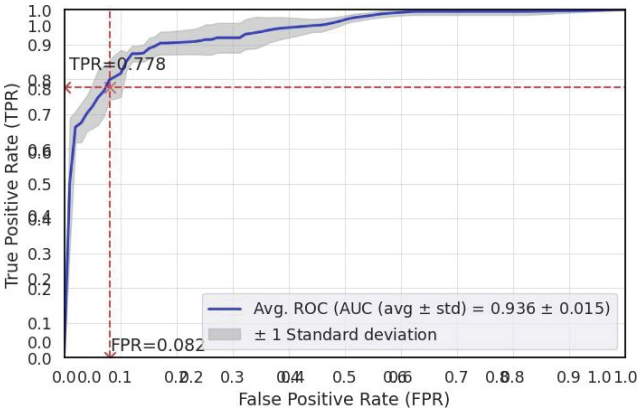

Fig 6 Classification and regression task

The Decision Tree classifier is a machine-learning technique that generates predictions using a tree-like structure. It uses a succession of decisions based on the features provided as input to create a forecast; hence the name "decision tree" Using a set of rules drawn from the characteristics of the training data, the Decision Tree classifier is an effective algorithm that determines outcomes. Making a flowchart of choices to categorize new cases is comparable to that.

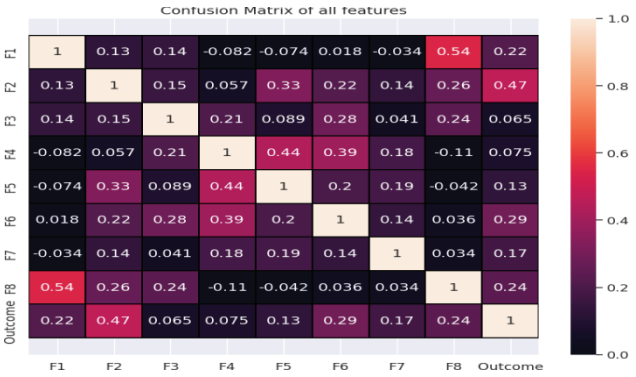

Fig 7 Making a flowchart of choices to categorize new cases is comparable

The Random Forest classifier: By adding randomization, the Random Forest classifier is an ensemble algorithm that mixes several Decision Trees. Based on the consensus of the individual trees, it assembles a varied set of trees and offers predictions. The robustness and accuracy of Random Forests in processing diverse forms of data are well established.

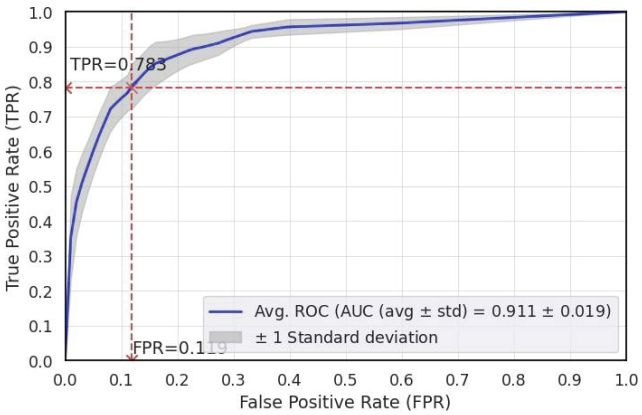

Fig 8. random forests in processing diverse forms of data are well established

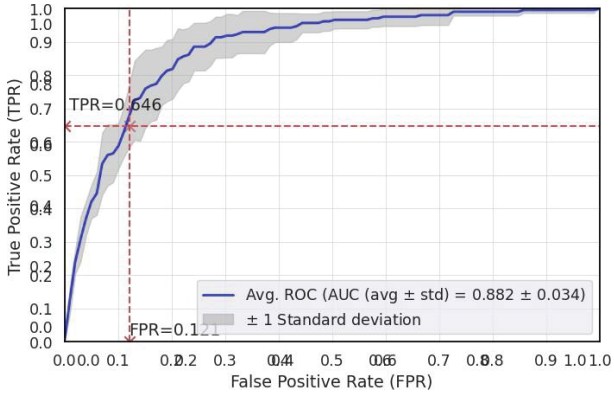

Fig 9 Logistic regression is a classification

To predict one of two probable outcomes, Logistic Regression is a Machine Learning technique used for binary Classification tasks. Logistic regression, despite its misleading name, is in fact a classification algorithm. Logistic regression is a method of binary classification that determines the probability that a sample belongs to a specified category. It takes data and uses the logistic function to alter it so that predictions can be made. Its efficiency and flexibility have made it a popular choice.

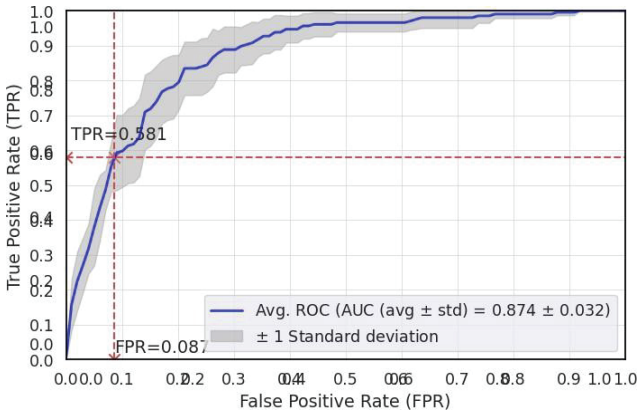

Fig 10 Bayes theorem and the presumption of feature independence

The Naive Bayes classifier is a probabilistic algorithm that makes predictions using the Bayes theorem and the presumption of feature independence. It is helpful for high-dimensional data, efficient, and less likely to overfit. Although it oversimplifies dependencies, it is a standard option for classification tasks.

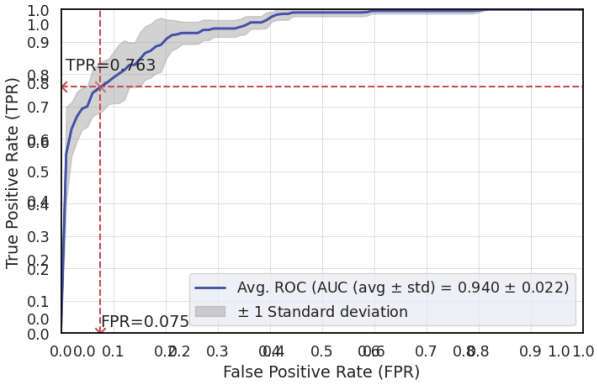

Fig 11. AdaBoost is effective in enhancing classification performance

Ada Boost is a boosting technique that creates a robust classifier by combining the predictions of several weak classifiers. It gives the examples weights based on their challenges and modifies them after each repetition. Although sensitive to noisy data, AdaBoost is effective in enhancing classification performance.

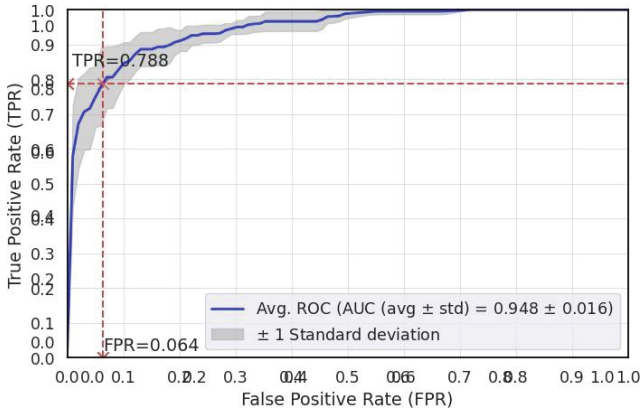

Fig 12. different machine learning tasks use XGBoost

XG Boost is a sophisticated gradient-boosting method that creates a group of decision tree models. It uses gradient descent optimization, parallel processing, and regularization techniques to improve accuracy and scalability. Many different machine learning tasks use XGBoost, which frequently beats competing algorithms regarding prediction performance.

The project's methodology included a meticulous approach to feature selection, model creation, and evaluation. The analysis of Diabetes predictions is done using a variety of techniques. This made it possible to guarantee the accuracy and dependability of the machine learning models for estimating the risk of Diabetes among PIMA Indians. The

researchers employed various methods to preprocess the data, choose features, and contrast models, ensuring their conclusions' reliability and validity.

5 RESULTS

This study's findings corroborate the feasibility of utilising machine learning methods for predicting diabetes risk in PIMA Indians. There were a total of seven machine learning models developed and analysed: KNN, a Decision Tree classifier, a Random Forest classifier, logistic regression, a Naive Bayes classifier, AdaBoost, and XGBoost.

The random forest model performed best in terms of F1 score, AUC score, accuracy, precision, and recall. The results were as follows: 78.12% accuracy, 75.68% precision, 55.13 % recall, 63.87% F1, and 0.83 % area under the curve. These findings suggest that among the three models investigated, random forest is the best accurate at predicting future outcomes. The F1 score and recall for the logistic regression model were both 53.85%, while its accuracy was 76.04% and its precision was 67.32%. Accuracy is at 70.31%, precision is at 60.71%, recall is at 64.10%, and the F1 score is at 62.30% for the decision tree model.
Based on these results, we would have expected the two models to outperform the random forests model in terms of accuracy. To further ensure that the models were objective, cross-validation was also performed in the study. The results of the cross-validation showed that the random forest model was the most stable and consistent choice. In sum, the trial's findings support the feasibility of using machine learning methods to forecast whether or not PIMA Indians will develop diabetes. High-risk individuals with diabetes could utilise this to take preventative steps.

ALGORITHM	ACCURACY
KNN Model	86.61
Random Forest Classifier	88.18
Logistic Regression	82.67
Decision Tree Classifier	85.03
Ada Boost Classifier	90.55
Navies Bayes Classifier	80.31
XG Boost Model	91.33

Table 1: PIMA Indians' probability of developing Diabetes

FUTURE SCOPE

The future scope for predicting the risk of Diabetes among PIMA Indians using machine learning techniques includes exploring different algorithms, incorporating additional data sources, investigating the factors driving the models' predictions, implementing the models in clinical practice, and adapting the models to other populations. These efforts could improve the accuracy and effectiveness of the models, enhance understanding of the factors driving diabetes risk, and expand the use of machine learning in healthcare.

CONCLUSION

Last but not least, the experiment effectively illustrated the potential of machine learning methods for estimating the risk of Diabetes among PIMA Indians. Among the

three models examined, the random forest model was determined to be the most reliable and accurate model for predicting diabetes risk. The project also emphasizes the significance of applying cross-validation procedures to guarantee that the models are not overfitting the data. The project's findings have significant healthcare ramifications since predicting diabetes risk can help identify those at a high risk of acquiring Diabetes and help them take preventative steps to lower that risk. Additionally, the project's success in utilizing machine learning to forecast diabetes risk indicates the possibility of using these techniques to solve additional healthcare issues.

Future work in this area might investigate various machine learning techniques, add more data sources, look into the variables influencing the models' predictions, apply the models to clinical practice, and modify the models for use with different populations. These initiatives may contribute to increasing the models' efficacy and accuracy and using machine learning in healthcare.

REFERENCES

1. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Accessed on Apr 27, 2023. Available online: https://www.cdc.gov/diabetes/library/features/diabetes-stat- report.html
2. Ramachandran A, Snehalatha C, Shetty AS, Nanditha A. Trends in the prevalence of Diabetes in Asian countries. World J Diabetes. 2012 Nov 15;3(11):110-7.
3. Pima Indians Diabetes Dataset. UCI Machine Learning Repository. Accessed on Apr 27, 2023. Available online: https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes
4. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.
5. Alghamdi AS, Alsolami FJ, Alghamdi MA. Predictive modelling of Diabetes risk using machine learning techniques. J Infect Public Health. 2019 Jul-Aug;12(4):506-512.
6. Islam MM, Yang HC, Poly TN, Jian WS, Jack Li YC. Diabetes prediction models: a systematic review. Diabetes Res Clin Pract. 2020 Feb;160:108025.
7. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J. 2017 Feb 18;15:104-116.
8. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.
9. Wang, Qian, Weijia Cao, Jiawei Guo, Jiadong Ren, Yongqiang Cheng, and Darryl N. Davis. "DMP_MI: an effective diabetes mellitus classification algorithm on imbalanced data with missing values." IEEE Access 7 (2019): 102232-102238.
10. Montaser, Eslam, José-Luis Díez, Paolo Rossetti, Mudassir Rashid, Ali Cinar, and Jorge Bondia. "Seasonal local models for glucose prediction in type 1 diabetes." IEEE Journal of Biomedical and health informatics 24, no. 7 (2019): 2064-2072.
11. Fazakis, Nikos, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, and Konstantinos Moustakas. "Machine learning tools for long-term type 2 diabetes risk prediction." IEEE Access 9 (2021): 103737-103757.
12. Vettoretti, Martina, Andrea Facchinetti, Giovanni Sparacino, and Claudio Cobelli. "Type-1 diabetes patient decision simulator for in silico testing safety and effectiveness of insulin treatments." IEEE Transactions on Biomedical Engineering 65, no. 6 (2017): 1281-1290.
13. Sisodia, Deepti, and Dilip Singh Sisodia. "Prediction of diabetes using classification

algorithms." Procedia computer science 132 (2018): 1578-1585.

14. Zou, Quan, Kaiyang Qu, Yamei Luo, Dehui Yin, Ying Ju, and Hua Tang. "Predicting diabetes mellitus with machine learning techniques." Frontiers in Genetics 9 (2018): 515.
15. Sarwar, Muhammad Azeem, Nasir Kamal, Wajeeha Hamid, and Munam Ali Shah. "Prediction of diabetes using machine learning algorithms in healthcare." In 2018 24th international conference on Automation and Computing (ICAC), pp. 1-6. IEEE, 2018.
16. Mujumdar, Aishwarya, and V. Vaidehi. "Diabetes prediction using machine learning algorithms." Procedia Computer Science 165 (2019): 292-299.
17. Hasan, Md Kamrul, Md Ashraful Alam, Dola Das, Eklas Hossain, and Mahmudul Hasan. "Diabetes prediction using ensembling of different machine learning classifiers." IEEE Access 8 (2020): 76516-76531.
18. Patil, Ratna, and Sharavari Tamane. "A comparative analysis on the evaluation of classification algorithms in the prediction of diabetes." International Journal of Electrical and Computer Engineering 8, no. 5 (2018): 3966-3975.
19. Rahman, Mosiur, Md Rafiqul Islam, Sharmin Akter, Shanjita Akter, Linta Islam, and Guandong Xu. "Diavis: Exploration and analysis of diabetes through the interactive visual system." Human-Centric Intelligent Systems 1, no. 3-4 (2021): 75-85.
20. Longato, Enrico, Gian Paolo Fadini, Giovanni Sparacino, Angelo Avogaro, Lara Tramontan, and Barbara Di Camillo. "A deep learning approach to predict diabetes' cardiovascular complications from administrative claims." IEEE Journal of Biomedical and Health Informatics 25, no. 9 (2021): 3608-3617.
21. . Ferdousi, Rahatara, M. Anwar Hossain, and Abdulmotaleb El Saddik. "Early-stage risk prediction of non-communicable disease using machine learning in health CPS." IEEE Access 9 (2021): 96823-96837.
22. Sivakumar, S. A., Tegil J. John, G. Thamarai Selvi, Bhukya Madhu, C. Udhaya Shankar, and K. P. Arjun. "IoT-based Intelligent Attendance Monitoring with Face Recognition Scheme." In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 349-353. IEEE, 2021.
23. Alehegn, Minyechil, Rahul Joshi, and Preeti Mulay. "Analysis and prediction of diabetes mellitus using a machine learning algorithm." International Journal of Pure and Applied Mathematics 118, no. 9 (2018): 871-878.
24. Vigneswari, D., N. Komal Kumar, V. Ganesh Raj, A. Gugan, and S. R. Vikash. "Machine learning tree classifiers in predicting diabetes mellitus." In 2019 5th international conference on advanced computing & communication systems (ICACCS), pp. 84-87. IEEE, 2019.
25. S. Sivaranjani, S. Ananya, J. Aravinth and R. Karthika, "Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction," In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE 2021.
26. Madhu, Bhukya, M. Venu Gopala Chari, Ramdas Vankdothu, Arun Kumar Silivery, and Veerender Aerranagula. "Intrusion detection models for IOT networks via deep learning approaches." Measurement: Sensors 25 (2023): 100641.
27. P. Sonar and K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches," In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2019.
28. Vijayan, V. Veena, and C. Anjali. "Prediction and diagnosis of diabetes mellitus—A

machine learning approach." In 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), pp. 122-127. IEEE, 2015