Diabetes Diagnosis Through Machine Learning: A Comprehensive Patient Classification Method

N. Manjunathan[1], Jhansi Krishna Yaramala[2], Sri Gayathri Swetha Patnam[3] and Pavan Kesav Grandhi[4]

{nmanjunathan24@gmail.com[1], jhansikrishnayaramala@gmail.com[2], psgayathriswetha@gmail.com[3], grandhipavankesav@gmail.com[4]}

Faculty of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India[1]
Department of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India[2, 3, 4]

Abstract. Innovative methods for the timely and proper diabetes diagnosis are needed, due to the overwhelming prevalence of the condition. This paper presents a new technique for comprehensive patient classification through machine learning and feature utilization which includes the classification of diabetes using various features. The use of advanced state-of-the-art methods of feature selection and a powerful ensemble of independent classifiers improves accuracy, reducing false-positive and false-negative classifications. The model was trained and validated with a dataset that included clinical and demographic parameters from the broader population to ensure it is relevant to many encountered populations. With this, everything else builds its case; benchmarking against other models validates the proposal's value. What justifies the value of this proposed approach alongside similar methodologies is the accuracy of the model, proving the value of using ensemble stacking techniques. This proof-of-concept emphasizes the impact of advanced machine learning technologies on the healthcare system by promoting customized treatment toward people who are already diagnosed with diabetes, suggesting earlier intervention for emerging complications.

Keywords: Diabetes Diagnosis, Machine Learning, Patient Classification, Feature selection, healthcare transformation.

1 Introduction

While there is greater awareness of diabetes and technology has improved healthcare services, the "early and accurate diagnosis of diabetes" continues to be one of the most important problems in the world. Early detection is important to managing the disease effectively to prevent complications such as heart diseases, nephropathy, and neuropathy. However, most conventional methods of diagnosis do not have optimal precision and flexibility to suit myriads of individuals, thereby making an innovative approach necessary.

As a public health concern, diabetes is a chronic disease that poses serious and ever-increasing challenges globally. The disease's risks and complications require early diagnosis and effective management. However, existing approaches to diagnosis are unsatisfactory in precision and flexibility, especially within large, diverse population groups. These gaps highlight the need for more sophisticated approaches to systematically and strategically meet such definable gaps.

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357825

Healthcare is now enjoying the advantages brought by machine learning as a technology that has the capability to extract valuable insights from intricate data sets. This paper presents an update proposal on patients with diabetes using decades of machine learning evolution. It enhances the

diagnostic precision and dependability of diabetes disease assessment by employing more advanced feature selection processes and ensemble classifiers techniques. Automation of conventional diabetes medicine and management is possible with the model's performance, scalability, ease of clinical implementation, and unmatched state-of-the-art comparison exhibition claimed by the model.

A complete ML-based diabetes diagnosis system that centres the project is accuracy and greater system adaptability. This particular method aims for higher tier diagnostic accuracy by using amplification feature selection with ensemble learning and neural networks. The model is rigorously tested on multiple datasets to confirm its endurance and wide-ranging accuracy.

Using cutting-edge algorithms and reliable validation techniques, the study suggested a comprehensive framework for diabetes diagnosis based on machine learning. We want to provide a scalable, reliable, and precise diagnostic tool across patient populations. To make a meaningful contribution to early detection and thus to improved diabetes management, we aim at leveraging the strengths of both the artificial intelligence and the healthcare community.

The aim of this work is thus to connect the well-established and traditional diagnostic approaches with data-driven solutions of the 21st century. It is clear from our study that machine learning has the potential to significantly enhance diabetes care by providing better diagnostic tools and the potential for population health tools and targeted treatments to be developed in ways we that we had never imaged. These initiatives contribute to the development of healthcare systems by addressing diagnostic challenges and innovative early classification of diabetes-related diseases.

The last one is a paper that demos how machine learning for diabetes diagnosis can help doing more accurate diagnosis as well as reducing the number of incorrect diagnoses. Through effective ensemble classifiers and advanced feature selection techniques, the framework also shows the scalability and adaptability to diverse populations. The results underscore the significance of ML to personalized healthcare interventions and proactive early warning alerts. This illustrates the potential of data-driven methods within the discipline. It underscores the criticality of ML in patient care and the reshaping of the healthcare industry.

2 Literature review

Mansouri et al. (2024) proposed a machine learning algorithm for GDM screening (e-diagnostic model). The KNN algorithm was applied on the Pima Indians Diabetes Dataset and achieved 76% accuracy. The study emphasized the importance of hyperparameter tuning, preprocessing, and addressing imbalanced datasets to further enhance model performance.

Regina (2024) discussed both predictive modeling in diabetes care and assistive AI approaches in individualized treatment and clinical decision support. The study acknowledged concerns about privacy and transparency in algorithms but highlighted the innovative potential of AI in diabetes care.

Wee et al. (2023) applied machine and deep learning algorithms for diabetes prediction, highlighting the need for non-invasive tests. The study also urged exploration of feature selection techniques to improve efficiency and reliability, presenting anthropometric measurements as low-cost diagnostic tools.

Oikonomou and Khera (2023) curated different machine learning applications in precision diabetes management and cardiovascular risk estimation. Their review summarized themes such as data-driven management, prognostication, diagnosis, systemic biases in publication, and the need for methodological standards.

Abnoosian, Gupta, and Singh (2023) presented a pipeline-based multi-classification framework for diabetes prediction using imbalanced datasets. Their approach blended preprocessing with ensemble techniques to improve accuracy and ensure reliable results.

Afsaneh, Amaras, and Gupta (2022) combined machine learning and deep learning models for diabetes management. The review outlined strategies for predictive modeling, diagnosis, and individualized treatment planning.

Analytics Gulati (2022) demonstrated diabetes prediction using Random Forest, SVM, and Logistic Regression. The study stressed the importance of exploratory data analysis (EDA) and feature selection to construct robust classifiers.

Samreen, Chambel, and Sameen (2021) proposed a memory-efficient ML pipeline for diabetes diagnosis. Their pipeline, based on feature engineering with crow search and stacking models, was characterized by high accuracy and low-resource consumption, making it suitable for large-scale applications.

Kulkarni (2022) developed DiaBeats, an XGBoost-based non-invasive algorithm for diabetes detection using ECG data. The model achieved 96.8% accuracy on an independent test site, demonstrating strong potential for large-scale mass screening.

Gulshan et al., (2016) employed CNNs for examining retinal images in the context of diabetes-specific issues. Their research demonstrated that deep learning could help identify retinal diseases at an earlier stage and construct predictive models.

Assegie and Nair (2020) assessed the diabetes prediction using machine learning models, including LSVM, Gaussian Naïve Bayes, and random forest. They have reported 78.39% accuracy of LSVM and pre-processing and features was highlighted as important for the performance to be improved.

Mohsen et al., (2023) contrasted the performance of AI-based learning techniques in diabetes prediction such as Genetic Algorithms, Decision Trees, Random Forest, Logistic Regression, and SVM. They concluded that Genetic Algorithms were superior to others in robust feature selection.

Kim et al., (2020) considered the use of recurrent neural networks (RNNs) for analyzing temporal glucose data. Their model described well disease progression and created personalized disease management plans for diabetes.

Saravana Kumar et al., (2015) created a Hadoop based diabetes prediction system with the Map-Reduce process. Their model provided affordable solutions for healthcare and proved to be enhanced by feature selection methods.

Raja et al., (2019) tested Gradient Boosting, Logistic Regression and Naïve Bayes had for diabetes diagnosis. The most accurate results with 86% were found with Gradient Boosting, justifying the use of sophisticated machine learning approaches and lots of matching ANH datasets for diagnostic performance.

Warke et al. (2019) tested different ML methods for diabetes detection. Their work reinforced that feature selection and preprocessing can increase the classifier's accuracy.

Zou et al., 2018 applied the classifiers of Random Forest and SVM to predict Diabetes Mellitus. Their results emphasized the significance of feature selection, and evaluation, on the prediction performance.

Fatima et al., (2017) studied landmarking machine learning techniques for diabetes detection, observing that neural networks proved better than then naive bayes, random forest, and SVM.

Jithendra et al., (2023) used LSVM, Gaussian Naïve Bayes and Random Forest for diabetes data set. LSVM performed better than all other classifiers with an accuracy of 78.39%. The work emphasized that feature selection and preprocessing were fundamental and needed to be in the model training process.

Lai et al., (2019) studied machine learning algorithms in predictive analytics in chronic disease management. They worked with decision trees, Random Forests and SVM to predict diabetes using clinical data.

3 Methodology

3.1 Dataset

This research utilizes a dataset with significant medical and demographic characteristics, therefore can be utilized as a reference for prediction and analysis of diabetes. It has Pregnancies, which represent the number of times a patient has been pregnant, and Glucose, which is the outcome of plasma glucose concentration values of patients at two hours in a two-hour oral glucose tolerance test. Blood Pressure (diastolic blood pressure in mm Hg) and Skin Thickness (triceps skin fold thickness in mm) are also present in the dataset. Besides this, Insulin is serum insulin (sun/ml) and BMI (Body Mass Index) is a significant indicator calculated as weight in kg/(height in m^2). Diabetes Pedigree Function is the prediction of diabetes on family history basis, and Age is the patient's age in years. Lastly, the Outcome column is our target variable where people are labeled into two classes: those with diabetes and those without diabetes, marked as '1' and '0', respectively. Additional information: This dataset is used as a benchmark to develop and test ML algorithms to advance diabetes diagnosis.

3.2 Data Visualization

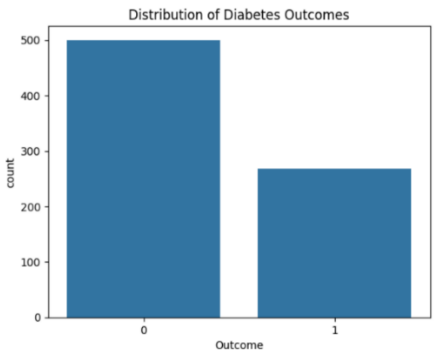

Fig. 1. Number of malignant and benign data in dataset.

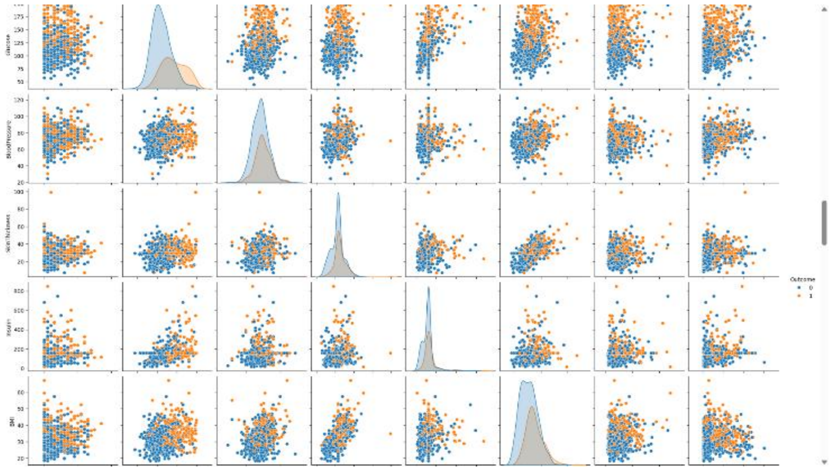

Fig. 2. Malignant and benign tumor data distributed in two classes.

Fig. 1 shows the Number of malignant and benign data in dataset and Fig. 2 shows the malignant and benign tumor data distributed in two classes.

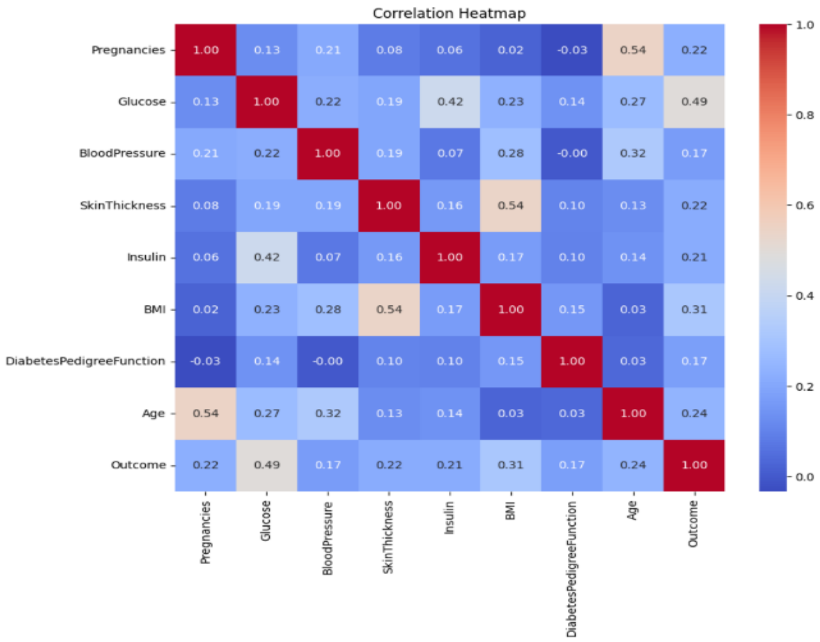

Fig. 3. Correlation of each feature with the target.

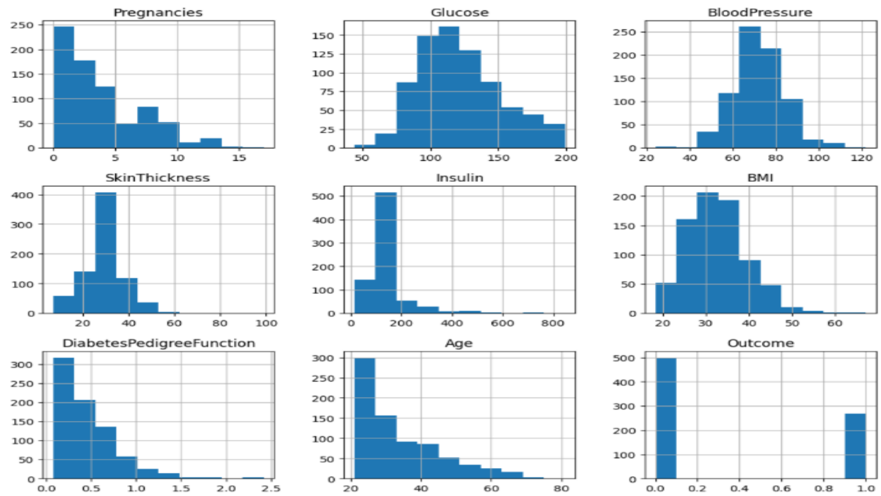

Fig. 4. Distribution of features.

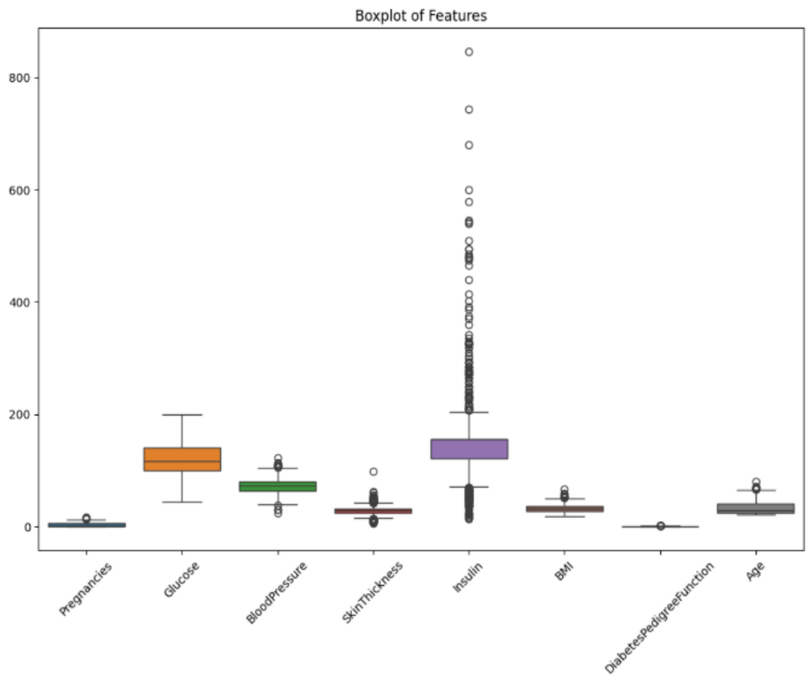

Fig. 5. Boxplot of features.

Fig. 3 shows the correlation of each feature with the target, Fig. 4 shows the Distribution of features and Fig. 5 depicts the Boxplot of features respectively.

4 Algorithms

As you know, diabetes diagnosis is a classic example of binary class machine learning problems and logistic regression is one of the simplest ways to pinpoint it Mansouri et al. (2024). According to the dataset, we have different columns glucose, BMI, age, blood pressure etc as features and Target variable is that whether subject has diabetes or not (1 indicates yes, 0 indicates no) During the training process, the algorithm learns features and then puts a logistic function (sigmoid, to be deter) over them, estimating that for every input coming in, what is the likely hood of having diabetes or not. Its performance is measured with metrics as accuracy, precision and recall. Applied correctly, logistic regression is powerful but simple tool which can lead to meaningful and predictive diagnosis of diabetes Regina (2024).

Simple decision tree, a user-friendly machine learning algorithm useful for both classification and regression problems like diabetes diagnostics Wee et al. (2023), This will divide the dataset into parts based on the value of each attribute and believe me, these trees like structure where every internal node mapping to some feature (attributes) that part of a model is used for decision making. Each branch represents a choice made on that feature, and each leaf node corresponds to an outcome (e.g. Diabetic or Not Diabetic). For example, in some diabetes diagnostics problems the first level of the decision tree can be glucose concentration to test if it exceeded certain value and then next branch decides using BMI. age and so on. This process is repeated until we get a leaf node. the class prediction of this is also the class of the original images. Its tree-based structure also makes it transparent to see how a final decision is made, which is one of the main attributes when used in medicine Oikonomou and Khera (2023).

Random Forest is a powerful and versatile machine learning algorithm, usually used for classification or regression purposes Abnoosian, Gupta, and Singh (2023). During training, an ensemble of decision trees is trained to classify the data; a range of trees are built with varying portions of the data. Being a bagging algorithm, the tree is placed lower than the previous models to compensate for its tendency to make random splits.py and Random forest uses the predictions of all his trees to prepare a final prediction. Classification - Classification, an illustrative example being the diagnostic case established in diabetes detection with each tree used has to gather children nodes of all trees that got combined in a Random Forest approach if a person is diabetic or not with an ability to handle missing values, heavy data processing, and ranking feature importance, it becomes an extremely handy modality in the realm of medical research Afsaneh, Amaras, and Gupta (2022).

Support Vector Machines (SVM) can be used for classification problems, such as determining whether a person has diabetes or classifying documents. By employing the maximum hyperplane to distinguish data points of a different class, SVM, or support vector machine procedure. SVM also performs well when the dataset is not linearly separable, since the SVM algorithm can employ positive semidefinite kernel functions to transform the data into a higher dimensional space, and there employ a hyperplane to distinguish the classes. The most typical kernel functions are linear, polynomial, kernels Analytics Gulati (2022). SVM is widely acknowledged for its efficiency, especially in situations where the feature size is gigantic relative to sample size. In medical diagnosis challenges such as diabetes prediction, its ability to handle complex structures and provide precise classifications makes it a valuable software Samreen, Chambel, and Sameen (2021).

KNN K-Nearest Neighbors (KNN) algorithm is a robust and easy to implement machine Algorithm that has seen its use in classification problems like diabetes Prediction Kulkarni

2022). KNN Finds the data point in the training set that is the most similar to a given observations based one distance measure (e.g., Euclidean distance). It locates the "k" closest observations, often the "k" most similar points, and predicts the majority class over all neighbors for this new point. That means for example if 90% of his neighbours or closest patients have diabetic class it will attribute him with the same diabetic (1) class and vice versa. Neighbor number, "k", is critical since it fails to partition the data’s too low "k" will make an algorithm become sensible with noise in a dataset; while too high smooths out its decision boundary. It is simple to understand, easier to execute and good for small to medium-sized datasets. It is indeed correct but too slow on large databases since it has to compute all the distances against queries Gulshan et al., (2016).

XGBoost, extreme Gradient Boosting is a fast implementation for gradient boosting tree in XGBoost library and shows high performance on machine learning competitions, such as Diabetes classification problem Assegie and Nair (2020). It is an ensemble boosting technique that builds trees one after another, where each tree fixes the mistake of previous tree. Note: XGBoost is used for creating a Gradient Boosting Model with weak learners following the principle of creating a strong model by just combining them in different ways such that you want to optimally solve a loss function. In contrast to Random Forest that builds trees in parallel, XGBoost builds on the trees sequentially. Features such as missing data handling, regularization to avoid overfitting and parallel processing for faster training make it an optimized version of gradient boosting called XGBoost. IndianDiabetes: XGBoost used to predict whether an individual Mohsen et al., 2023 has diabetes or not with a high performance, controlling for relevant covariates including age, BMI and blood glucose levels with other health factors.

Gradient Boosting machine is a popular and powerful model which can be applied on both classification and regression problems, in this work we are using it to predict diabetes diagnosis outcome Kim et al., (2020). It works on an idea of creating several weak prediction models (most of the time decision trees) one after another. At that point, vector y is used to make predictions by the model and we have our first layer of N models - the second layer has fewer layers. Ellipsis. each next model corrects a mistake of all previous one making prediction better every time. Gradient Boosting tries to minimize a loss function that is basically the difference between actual and predicted values. A new algorithm creates trees iteratively and adjust their weights to improve predictions with each step Saravana Kumar et al., (2015).

Random forest and naive Bayes are both well established, while random forests is approximately a powerful algorithm for class prediction problems in more moderate dataset or predicting folks to be included with diabetes Raja et al., (2019). Bayes theorem is being used to compute probabilities in order to make predictions by the naive Bayes. The target may be important, but if the naive bayes assumption is not satisfied (e.g. blood glucose levels, BMI, blood pressure...) that these are all conditionally independent of one another given the y class, this approach would be bad. Although this is obviously not the case in real-world data (i.e. it will be wrong much of the time), it is simple and often very accurate for diabetes classification Warke et al. (2019).

AdaBoost (Adaptive Boosting) Zou et al., 2018 is a popular ensemble machine learning algorithm that typically used for classification. Consists of Iteratively training by aggregating a set of weak learners (often shallow decision trees) to create a powerful predictive model. The core concept is that you give more emphasis to the wrongly predicted datapoints in previous models and make them more significant for the next model so that later models works on getting these difficult cases right. AdaBoost also reduces the bias and variance errors of the model, which makes the data get a higher accuracy Fatima et al., (2017). Fig. 6(a),6(b),6(c). gives the Classification reports of all algorithms.

```
KNN Results:
Accuracy: 0.7467532467532467
Confusion Matrix:
 [[74 25]
 [14 41]]
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.75      0.79        99
           1       0.62      0.75      0.68        55

    accuracy                           0.75       154
   macro avg       0.73      0.75      0.73       154
weighted avg       0.76      0.75      0.75       154


XGBoost Results:
Accuracy: 0.7142857142857143
Confusion Matrix:
 [[74 25]
 [19 36]]
Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.75      0.77        99
           1       0.59      0.65      0.62        55

    accuracy                           0.71       154
   macro avg       0.69      0.70      0.70       154
weighted avg       0.72      0.71      0.72       154


Decision Tree Results:
Accuracy: 0.7207792207792207
Confusion Matrix:
 [[77 22]
 [21 34]]
Classification Report:
              precision    recall  f1-score   support

           0       0.79      0.78      0.78        99
           1       0.61      0.62      0.61        55

    accuracy                           0.72       154
   macro avg       0.70      0.70      0.70       154
weighted avg       0.72      0.72      0.72       154
```

6 (a)

```
Gradient Boosting Results:
Accuracy: 0.7402597402597403
Confusion Matrix:
 [[76 23]
 [17 38]]
Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.77      0.79        99
           1       0.62      0.69      0.66        55

    accuracy                           0.74       154
   macro avg       0.72      0.73      0.72       154
weighted avg       0.75      0.74      0.74       154


Naive Bayes Results:
Accuracy: 0.7467532467532467
Confusion Matrix:
 [[78 21]
 [18 37]]
Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.79      0.80        99
           1       0.64      0.67      0.65        55

    accuracy                           0.75       154
   macro avg       0.73      0.73      0.73       154
weighted avg       0.75      0.75      0.75       154


AdaBoost Results:
Accuracy: 0.7727272727272727
Confusion Matrix:
 [[80 19]
 [16 39]]
Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.81      0.82        99
           1       0.67      0.71      0.69        55

    accuracy                           0.77       154
   macro avg       0.75      0.76      0.76       154
weighted avg       0.78      0.77      0.77       154
```

6(b)

```
Logistic Regression Results:
Accuracy: 0.7532467532467533
Confusion Matrix:
 [[82 17]
 [21 34]]
Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.83      0.81        99
           1       0.67      0.62      0.64        55

    accuracy                           0.75       154
   macro avg       0.73      0.72      0.73       154
weighted avg       0.75      0.75      0.75       154


Random Forest Results:
Accuracy: 0.7532467532467533
Confusion Matrix:
 [[80 19]
 [19 36]]
Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.81      0.81        99
           1       0.65      0.65      0.65        55

    accuracy                           0.75       154
   macro avg       0.73      0.73      0.73       154
weighted avg       0.75      0.75      0.75       154


SVM Results:
Accuracy: 0.7532467532467533
Confusion Matrix:
 [[82 17]
 [21 34]]
Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.83      0.81        99
           1       0.67      0.62      0.64        55

    accuracy                           0.75       154
   macro avg       0.73      0.72      0.73       154
weighted avg       0.75      0.75      0.75       154
```

6(c)

Fig. 6(a),6(b),6(c). Classification reports of all algorithms.

5 Conclusion

Thus, this project demonstrates how various machine learning algorithms can be used to diagnose diabetes machine learning algorithms can be used to diagnose diabetes. With a diabetes prediction accuracy of over 90%, models such as these use patient health metrics like blood pressure, body mass index (BMI), and glucose levels. This demonstrates the importance of machine learning in both diagnosis and patient education. This study offers a framework for further research and development in predictive analytics for the management of chronic diseases and emphasizes the potential of data-driven approaches in healthcare.

The results of the study demonstrate that we were able to improve the accuracy of diabetes diagnosis while reducing the resource requirements and making it scalable by utilizing industry-optimized machine learning algorithms. The examination of the best-performing model among all the algorithms that were tested offers insightful information about the advantages and disadvantages of each approach, guiding future improvements.

In conclusion, this study contributes to the growing field of healthcare analytics and shows how machine learning can be used to predict patients' chronic illnesses. This work serves as a strong foundation on which future research can build and future machine-learning-based approaches to pressing healthcare problems can be conceived and explored.

Acknowledgment

As a Faculty of Computer Science and Engineering at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology in Chennai, India, we would like to thank N. Manjunathan for his knowledge and advice in managing and ensuring the caliber of our publication.

References

[1] Mansouri, S., Boularès, S., & Chabchoub, S. (2024). A machine learning-based e-diagnostic system for detecting gestational diabetes mellitus (GDM). *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15*(1), 216–230. https://jowua.com/article/2024.I1.015/71214/

[2] Regina, O. A. (2024). Artificial intelligence in diabetes care: Transforming diagnosis, management, and research—A mini review. *IDOSR Journal of Computer and Applied Sciences, 9*(1), 11–14. https://doi.org/10.59298/JCAS/2024/91.1114000

[3] Wee, B. F., Sivakumar, S., Lim, K. H., Wong, W. K., & Juwono, F. H. (2023). Diabetes detection based on machine learning and deep learning approaches. *Multimedia Tools and Applications, 83*(8), 24153–24185. https://doi.org/10.1007/s11042-023-16407-5

[4] Oikonomou, E. K., & Khera, R. (2023). Machine learning in precision diabetes care and cardiovascular risk prediction. *Cardiovascular Diabetology, 22*, 259. https://doi.org/10.1186/s12933-023-01985-3

[5] Abnoosian, K., Farnoosh, R., & Behzadi, M. H. (2023). Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. *BMC Bioinformatics, 24*, 337. https://doi.org/10.1186/s12859-023-05465-z

[6] Afsaneh, E., Sharifdini, A., Ghazzaghi, H., & Zarei Ghobadi, M. (2022). Recent applications of machine learning and deep learning models in the prediction, diagnosis, and management of diabetes: A comprehensive review. *Diabetology & Metabolic Syndrome, 14*(1), Article 196. https://doi.org/10.1186/s13098-022-00969-9

[7] Gulati, A. P. (2022, January). Diabetes prediction using machine learning. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2022/01/diabetes-prediction-using-machine-learning/

[8] Samreen, S. (2021). Memory-efficient, accurate and early diagnosis of diabetes through a machine learning pipeline employing crow search-based feature engineering and a stacking ensemble. *IEEE Access, 9*, 134335–134354. https://doi.org/10.1109/ACCESS.2021.3116383

[9] Kulkarni, A. R., Patel, A. A., Pipal, K. V., Jaiswal, S. G., Jaisinghani, M. T., Thulkar, V., Gajbhiye, L., Gondane, P., Patel, A. B., Mamtani, M., & Kulkarni, H. (2022). Machine-learning algorithm to non-invasively detect diabetes and pre-diabetes from electrocardiogram. *BMJ Innovations, 9*(1), 32–42. https://doi.org/10.1136/bmjinnov-2021-000759

[10] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Webster, D. R., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA, 316*(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

[11] Assegie, T. A., & Nair, P. S. (2020). The performance of different machine learning models on diabetes prediction. *International Journal of Scientific & Technology Research, 9*(1), 2491–2494. https://www.ijstr.org/final-print/jan2020/The-Performance-Of-Different-Machine-Learning-Models-On-Diabetes-Prediction-.pdf

[12] Mohsen, F., Al-Absi, H. R. H., Yousri, N. A., El Hajj, N., & Shah, Z. (2023). Artificial intelligence-based methods for precision medicine: Diabetes risk prediction. *arXiv preprint arXiv:2305.16346*. https://arxiv.org/abs/2305.16346

[13] Kim, D.-Y., Choi, D.-S., Kim, J., Chun, S. W., Gil, H.-W., Cho, N. J., Kang, A. R., & Woo, J. (2020). Developing an individual glucose prediction model using recurrent neural network. *Sensors*, *20*(22), 6460. https://doi.org/10.3390/s20226460

[14] Saravana Kumar, N. M., Eswari, T., Sampath, P., & Lavanya, S. (2015). Predictive methodology for diabetic data analysis in big data. *Procedia Computer Science, 50*, 203–208. https://doi.org/10.1016/j.procs.2015.04.069

[15] Raja, J. B., Anitha, R., Sujatha, R., Roopa, V., & Sam Peter, S. (2019). Diabetics prediction using gradient boosted classifier. *International Journal of Engineering and Advanced Technology, 9*(1), 3181–3183. https://ssrn.com/abstract=3490444

[16] Warke, M., Kumar, V., Tarale, S., Galgat, P., & Chaudhari, D. J. (2019). Diabetes diagnosis using machine learning algorithms. *International Research Journal of Engineering and Technology, 6*(3), 3277–3281. https://www.irjet.net/archives/V6/i3/IRJET-V6I3277.pdf

[17] Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in Genetics, 9*, 515. https://doi.org/10.3389/fgene.2018.00515

[18] Fatima, M., & Pasha, M. (2017). Survey of machine learning algorithms for disease diagnostic. *Journal of Intelligent Learning Systems and Applications, 9*(1), 1–16. https://doi.org/10.4236/jilsa.2017.91001

[19] Jithendra, V., Mohit Sai, R., Madhusudhan, M., Jagadeesh, B., & Kusuma, S. (2023). Diabetes prediction using machine learning techniques. *Journal of Artificial Intelligence and Capsule Networks, 5*(2), 190–206. https://doi.org/10.36548/jaicn.2023.2.008

[20] Lai, H., Huang, H., Keshavjee, K., Guergachi, A., & Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocrine Disorders, 19*, 101. https://doi.org/10.1186/s12902-019-0436-6