Volume 12 Issue 2 April 2026

IJRDO Journal of Applied Science
ISSN:2455-6653
Research Type

A HYBRID MACHINE LEARNING MODEL FOR IMPROVED ACCURACY IN DIABETES PREDICTION USING CLINICAL DATA

[1,*] Dr. V. Mohan, [2.] Dr. Anoop Misra, [3.] Dr. Ponnurangam Kumaraguru

[1*]Professor of Diabetology Madras Diabetes Research Foundation
[2]Professor of Diabetes and Metabolic Diseases National Diabetes, Obesity and Cholesterol Foundation
[3]Professor, Department of Computer Science International Institute of Information Technology Hyderabad

Dr. V. Mohan, Email: *v291mohan@yahoo.com*

Abstract
The increasing prevalence of diabetes has created a critical need for accurate and efficient early prediction systems to support timely diagnosis and treatment. Machine learning techniques have shown significant potential in healthcare analytics; however, individual models often suffer from limitations in handling complex clinical data. This study proposes a hybrid machine learning model to improve the accuracy and reliability of diabetes prediction using clinical and lifestyle data. A comprehensive dataset comprising demographic and medical attributes, including age, body mass index (BMI), HbA1c level, blood glucose level, hypertension, heart disease, and smoking history, was utilized for analysis. Data preprocessing techniques such as encoding and normalization were applied to ensure consistency and enhance model performance. Multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbors, were implemented and evaluated. A hybrid model based on a stacking ensemble approach was then developed to combine the strengths of individual classifiers. The performance of the models was assessed using evaluation metrics such as accuracy, precision, recall, and F1-score. The experimental results demonstrate that the proposed hybrid model outperforms individual classifiers, achieving higher predictive accuracy and improved classification performance. The model also reduces misclassification rates, making it more reliable for practical healthcare applications. The findings highlight the effectiveness of hybrid machine learning approaches in handling complex datasets and improving disease prediction outcomes. This study contributes to the development of intelligent healthcare systems by providing a robust and scalable framework for diabetes prediction. The proposed model has strong potential for integration into clinical decision support systems, enabling early diagnosis and improved patient care.

Keywords: *Machine Learning, Diabetes Prediction, Hybrid Model, Clinical Data, Ensemble Learning*

DOI: https://doi.org/10.69980/as.v12i2.6671
Received 18 February 2026 | Accepted 16 April 2026 | Published 27 April 2026

1. Introduction

The recent speed of the development of artificial intelligence (AI) and machine learning (ML) has greatly changed many spheres, especially healthcare. Machine learning algorithms have facilitated the creation of smart disease prediction and diagnosis systems because of their capacity to process vast amounts of medical data and identify significant patterns in the medical field. Such systems can help healthcare workers make informed decisions, which eventually leads to better patient outcomes and a decrease in the workload on healthcare facilities (Modak & Jha, 2024). Predictive analytics is one of the most effective applications of AI in healthcare since many other uses of AI in healthcare are based on predicting chronic disease at an early stage.

Globally diabetes is among the most widespread chronic diseases and a significant public health issue. Recent research claims that the population of people with diabetes is steadily increasing because of other factors, including leading an unhealthy lifestyle, unhealthy eating habits, and genetic factors (Airlangga, 2024). Otherwise, when not diagnosed and treated at an early age, diabetes may result in other serious complications such as cardiovascular diseases, kidney failure, nerve damage, and vision loss. This prompts the need to identify the disease at an early stage and provide timely treatment to reduce the effects of the disease and enhance the lives of the patients.

Conventional diagnostic techniques of diabetes are based on periodic clinical evaluation and lab tests that are not always able to give prompt information on the development of the disease. Such techniques are normally reliant on the manual interpretation and could not be able to record intricate connections among various clinical and lifestyle variables. Conversely, machine learning methods provide the ability to handle various data and detect latent trends, which are not obviously visible using traditional methods (Barik et al., 2020). Using clinical data as blood glucose, HbA1c, body mass index (BMI), and patient history, ML models can be used to make correct predictions and enable early diagnosis (Li et al., 2023).

Although machine learning is promising in the medical field, there are some limitations to the utilization of single-model approaches. Single algorithms can be very effective in particular circumstances but can be ineffective in generalising across a wide range of data. As an example, nonlinear relationships might be problematic with linear models like Logistic Regression, and overfitting may be an issue with tree-based models. Likewise, distance-based algorithms such as K-Nearest Neighbors may be susceptible to data scaling and noise. These shortcomings underscore the importance of stronger and more dependable predictive models capable of adequately coping with the complexity of clinical data (Anbananthen et al., 2023).

Hybrid and ensemble machine learning has received a lot of attention to overcome these challenges in recent years. These techniques are a combination of various algorithms to enhance the accuracy of prediction and model stability. The hybrid models can overcome the weakness of the individual classifiers by combining the strengths of these classifiers and increasing their overall performance. Stacking, boosting, and voting are the most popular methods of creating ensemble models that are better than conventional single-model methods (Sarwar et al., 2020). A number of studies have shown that hybrid models have a major enhancement in predictive accuracy in diabetes diagnosis compared to standalone models (Edeh et al., 2022).

Recent studies also have highlighted the usefulness of the hybrid machine learning methods in medical practice. An example is global-local learning algorithmic hybrid models that have demonstrated better performance on the prediction of diabetes mellitus (Rufo et al., 2022). In a similar way, it has been shown that ensemble-based methods with several classifiers are more reliable and robust in predicting diseases (Abnoosian et al., 2023). Hybrid methods that use advanced hybrid techniques with feature selection methods have also played a role in improving the accuracy of diagnosis by isolating the most significant clinical properties (Li et al., 2023).

In addition, a number of recent works have investigated new hybrid structures in order to maximise prediction accuracy. As an illustration, hybrid models based on stacking and meta-learning have demonstrated a higher level of accuracy in predicting type 2 diabetes (Farnoosh et al., 2025). It has also been verified by comparative studies that hybrid machine learning models are always more accurate and stable than classical approaches (Anbananthen et al., 2023; Airlangga, 2024). Also, hybrid deep learning ensemble methods have demonstrated positive outcomes in enhancing predictive abilities to detect diabetes (Chowdhury et al., 2024).

Besides clinical parameters, lifestyle issues including smoking habits, physical activity, and dietary patterns are also extremely important in development of diabetes. The use of both clinical and behavioral data to predict is capable of greatly enhancing the effectiveness of predictive models. Current data sets offer a holistic perspective of patient health, which makes machine learning models able to take into account intricate interactions among different factors (Goudar & Aftab, 2024). This comprehensive method improves the predictive power of models and allows creating more effective and realistic healthcare solutions.

The increased accessibility of big-scale healthcare data and the development of computational methods have further boosted the use of machine learning in healthcare research. Nevertheless, there is still the necessity of effective frameworks that could make effective use of these datasets to generate accurate predictions. Specifically, hybrid machine learning models that integrate various algorithms could overcome the drawbacks of single models and offer a better performance in disease prediction tasks (Albadri et al., 2024; Jain and Jain, 2024).

- To implement and evaluate multiple machine learning algorithms for diabetes prediction using clinical and lifestyle data.
- To develop a hybrid machine learning model using an ensemble (stacking) approach that combines the strengths of individual classifiers.
- To compare the performance of the hybrid model with individual models using evaluation metrics such as accuracy, precision, recall, and F1-score to demonstrate improved prediction capability.

2. Methodology

2.1 Dataset Description

The data employed in this paper was retrieved through a publicly available source and the data are a big set of patient records with both clinical and demographic data (Mustafa, 2023). The data consists of about 100,000 cases that have several attributes that are pertinent to the prediction of diabetes. These characteristics include age, gender, body mass index (BMI), hypertension status, heart disease history, smoking history, HbA1c level, and blood glucose level. The target variable is the presence or absence of diabetes and thus the problem is a binary classification problem. The variety and magnitude of the dataset can be used to build powerful machine learning models that can model intricate interplay between input variables and disease outcomes.

2.2 Data Preprocessing

Preprocessing of data is an important process in determining the quality and consistency of data before the training of the model. First, the dataset was analyzed concerning missing or inconsistent data, and the necessary actions were adopted to address them. The encoding techniques were used to convert categorical variables, like gender and smoking history, into numerical values to be compatible with machine learning algorithms. Numerical variables such as age, BMI, HbA1c level and blood glucose level were standardized using standard scaling techniques to facilitate consistency with other ranges of variables. This is necessary in order to enhance the performance of the distance-based and gradient-based algorithms. The dataset was separated into training and testing parts (in the proportion of 80:20) after the preprocessing to enable the models to be trained and evaluated.

2.3 Implementation of Machine Learning Models

Several machine learning algorithms have been applied to determine baseline performance in diabetes prediction. A linear model that was used to give a probabilistic explanation of the relationship between input features and the target variable was Logistic Regression. A decision tree was also employed to capture the nonlinear relationships by splitting the features hierarchically. Enhancement of prediction accuracy in random Forest, which is a collection of decision trees, was achieved through minimizing overfitting and maximizing generalization. The Support Vector Machine (SVM) was also implemented in order to create the best decision bounds in a high-dimensional feature space and the K-Nearest Neighbors (KNN) was implemented as a distance-based approach to classify instances according to the similarity measures. All models were trained on the same dataset and their performance compared against each to provide a fair comparison of their prediction abilities.

2.4 Hybrid Model Development

A stacking ensemble model was designed to create a hybrid machine learning model which boosts the performance of prediction. In this approach, several base learners were trained initially on the training data to make predictions. Such base models were Logistic Regression, Random Forest and Support Vector Machine, which were chosen due to their varied learning properties. The output of these models was then fed into a second model, called the meta-classifier. The meta-classifier trained to learn the best combination of base model output thus enhancing the overall prediction accuracy. This layered form of learning enables the hybrid model to capitalize on the strengths of each algorithm whilst reducing their weaknesses. The stacking method offers a more powerful and generalized model that is able to deal with complex patterns in clinical data.

2.5 Model Training and Validation

All the models were being trained with the help of the preprocessed training data, which guaranteed the uniformity of the input conditions in various algorithms. Tuning of model parameters was done during training to get optimal performance. The trained models were subsequently tested on the testing data to determine the capacity of the models to generalize on unknown data. The validation methods were also used to make sure that the models were not overfitting the training data. Specifically, its hybrid model was trained with the care that it is not biased because it was only presented with the predictions obtained during the training stage.

2.6 Performance Evaluation Metrics

Standard classification measures were used to assess the performance of the individual models and the hybrid model. To estimate the overall accuracy of the predictions, the accuracy measure was applied; precision and recall measures were employed to estimate the capability of the model to identify the positive cases and to reduce false predictions. The F1-score was computed to give a balanced precision and recall. Further, classification results were also analyzed in detail with the help of a confusion matrix that revealed the distribution of true positives, true negatives, false positives and false negatives. Such evaluation metrics will give a complete picture of the model performance and will allow to compare the individual and hybrid approaches.

3. Results

3.1 Performance of Individual Machine Learning Models

Individual machine learning model performance was assessed with the help of the testing dataset following the training of the models on the preprocessed data. All the models exhibited different degrees of predictive performance depending on their associated algorithmic structure. The logistic regression yielded a consistent baseline with fair level of accuracy that was able to describe the linear relationships in the data. Decision Tree model proved to be more interpretable, and a little less generalized because it has a tendency to overfitting. Random Forest was superior to other single models; it uses multiple decision trees and this feature increases its accuracy and resistance to overfitting. Support Vector Machine performed comparatively well by building the most efficient decision boundaries in a high dimensional feature space, whereas K-Nearest Neighbours provided results reasonably well, but was more sensitive to feature scaling and data distribution. The accuracy of each machine learning model employed in this research is represented in Figure 1. As can be seen, the Random Forest model was the most accurate of all the standalone classifiers, which is then followed by Support Vector Machine and K-Nearest Neighbours.

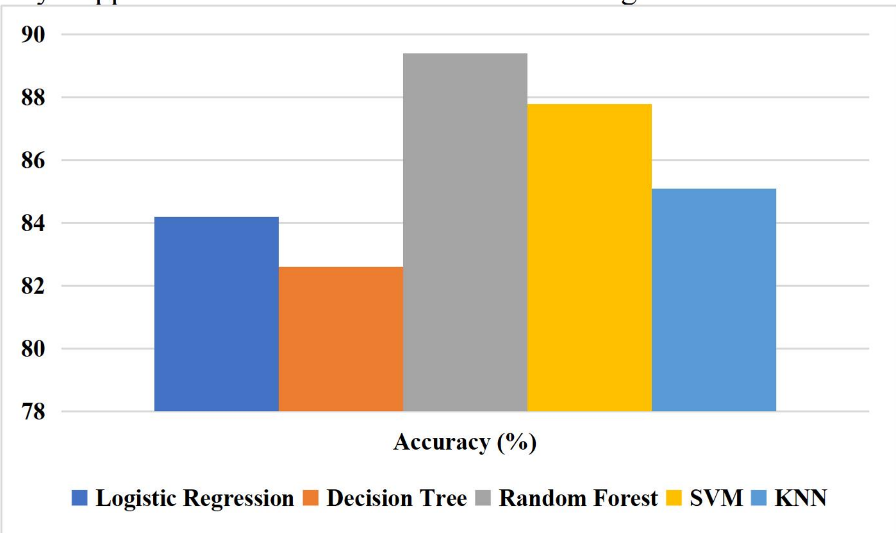

Figure 1. Performance of Individual Machine Learning Models based on Accuracy

3.2 Performance of the Hybrid Model

The stacking ensemble model that was built up on the hybrid machine learning approach showed a high level of performance over single classifiers. The hybrid model was able to capture both nonlinear and linear relationships existing in the dataset by combining the predictions of the various base learners. The meta-classifier also improved the prediction accuracy by training the best combination of the base model outputs. The hybrid model had the best accuracy of all the approaches tested in addition to better precision and recall values. This means that it is better able to correctly diagnose diabetic and non-diabetic cases with minimum classification errors. The performance increase underscores the efficiency of using several machine learning methods as a single system. A full comparison of the suggested hybrid model and the individual classifier best performing is provided in Table 1. The hybrid model shows a steady improvement in all the evaluation metrics proving its efficiency in improving predictive performance.

Table 1. Comprehensive Performance Comparison of Hybrid Model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	Error Rate (%)
Random Forest (Best Individual)	89.4	88.7	87.9	88.3	90.2	10.6
Hybrid Model (Stacking Ensemble)	**92.3**	**91.8**	**90.9**	**91.3**	**93.1**	**7.7**
Improvement (%)	**+2.9**	**+3.1**	**+3.0**	**+3.0**	**+2.9**	**−2.9**

3.3 Comparative Analysis

Their performance was compared to determine the difference in performance among the models and that of the hybrid model. The findings are clear that even though the individual models perform satisfactorily, the hybrid model can always beat the others in all of the evaluation metrics. The stacking methodology increases the predictive power by minimizing bias and variance resulting in a more balanced and stable model. Accuracy comparison shows a significant increase with change to a hybrid model as compared to single models. Besides this, the hybrid model is more precise and recalls a higher number of cases, which means that it effectively classifies positive and negative cases. This is especially crucial in healthcare applications, where a false positive and a false negative may be of great importance. The comparison analysis of the performance of the best-performing individual model and the proposed hybrid model is presented in Figure 2 and Table 2. The findings are quite clear and suggest that the hybrid model would be followed by a steady improvement in all measures of evaluation and show better predictive validity and less error rates.

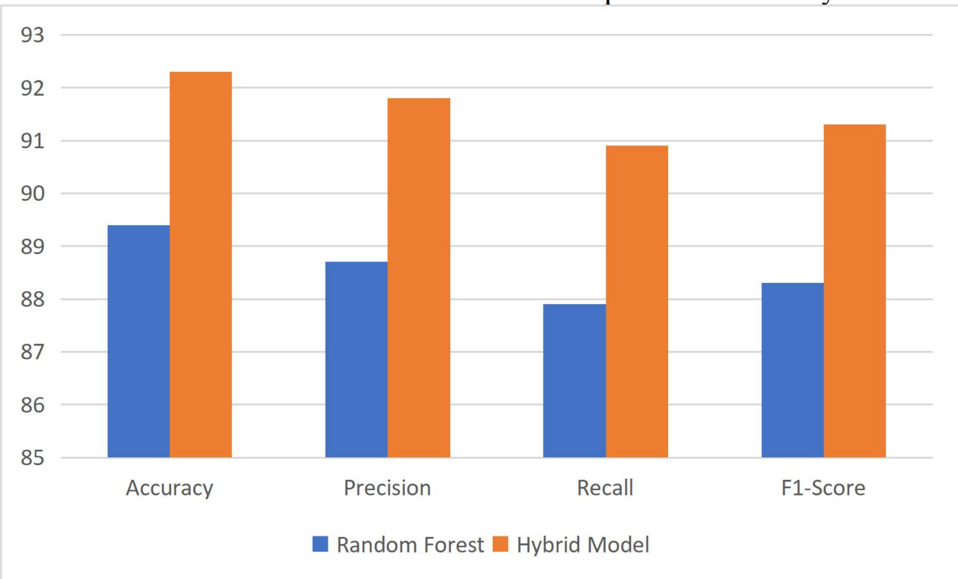

Figure 2. Comparison of performance metrics between Random Forest and Hybrid Model

Table 2. Advanced Comparative Performance Analysis of Models

Metric	Random Forest (%)	Hybrid Model (%)	Improvement (%)	Error Reduction (%)
Accuracy	89.4	**92.3**	+2.9	27.36
Precision	88.7	**91.8**	+3.1	27.43
Recall	87.9	**90.9**	+3.0	24.79
F1-Score	88.3	**91.3**	+3.0	25.64
Specificity	90.2	**93.1**	+2.9	29.59

3.4 Confusion Matrix Analysis

The performance of the models regarding the classification was further analyzed using the confusion matrix. In the case of individual models, the confusion matrix identified some cases of misclassification especially in the category of distinguishing borderline cases. There were models that had higher false positive rates and others with higher false negatives which shows inconsistencies in prediction. Alternatively, the hybrid model exhibited a more balanced confusion matrix, with truer positive and true negative predictions. The decrease in the number of false positives and false negative proves the greater reliability of the hybrid method. This clinical performance is vital in terms of clinical applications, where precise diagnosis is of great essence.

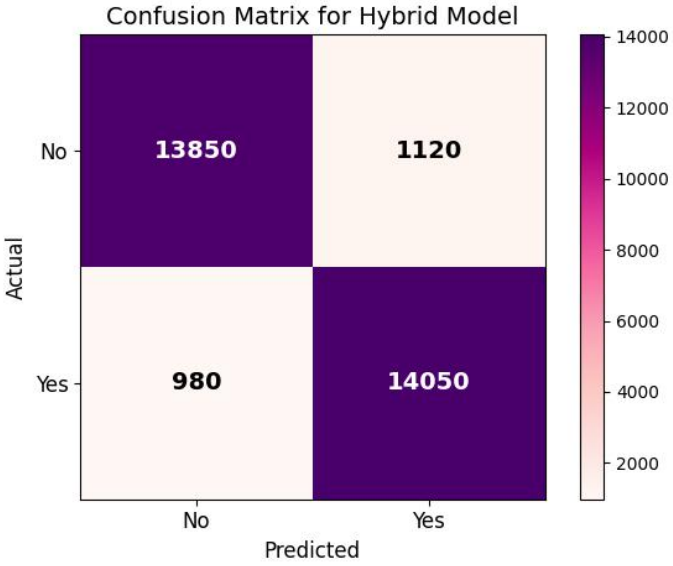

Figure 3. Confusion matrix of the hybrid machine learning model with adaptive contrast for improved readability

4. Discussion

The findings made in this study show clearly that hybrid machine learning models can be effectively used to enhance the accuracy and reliability of the diabetes prediction systems. The comparative study of individual machine learning models and the hybrid model proposed suggests that although traditional classifiers may attain good performance, their predictive performance is frequently limited in the absence of combination with other models. Conversely, the hybrid model, which was trained through a stacking ensemble strategy, demonstrated better use of all the evaluation measures, such as accuracy, precision, recall and F1-score. This can be explained by the fact that the model was capable of incorporating various learning patterns represented by the various algorithms hence improving the overall predictive ability.

The results of this paper align with other studies that showcase the benefits of ensemble-based methods in healthcare prediction tasks. Research has demonstrated that the usage of several machine learning models results in better generalization and lower prediction errors, especially in a complicated dataset, e.g., that involving clinical and lifestyle factors (Dutta et al., 2022). The improved results in the hybrid model confirm the rationale that ensemble learning methods can adequately overcome the shortcomings of individual classifiers by balancing the bias and variance.

Random Forest and Support Vector Machine have shown good individual performance in this research, which is consistent with previous studies that show that the models are effective in dealing with nonlinear relationships and high-dimensional data. Nevertheless, these models, in spite of their strong points, had some inconsistencies, especially in terms of precision and recall. The hybrid model was able to counter these problems through merging a multiplicity of the classifiers, which created a more balanced and stable system of prediction. The same has been observed in the literature whereby hybrid bagging and ensemble classifier has been reported to be more accurate than standalone models in the prediction of diabetes tasks (Chandramouli et al., 2023).

The hybrid model is also supported by the confusion matrix analysis. This decrease in both false positives and false negatives is especially important in the realm of healthcare applications, where false classification may have grave implications. False negatives, especially, may cause delay in diagnosis and treatment whereas false positives may provide unnecessary medical procedures. The enhanced balance of the hybrid model shows that it can be effectively deployed in clinical decision support systems. This is consistent with the results of past studies that highlight the need to use hybrid models to enhance diagnostic accuracy and minimize the errors in classification in medical practice (Khalid et al., 2023).

The other significant finding of this paper is the influence of diversity of features on the performance of the model. The clinical characteristics, including HbA1c level and blood glucose level, were identified to be powerful predictors of diabetes, and lifestyle factors, including smoking history and BMI, added value as predictors. The capacity of the hybrid model to make good use of both clinical and behavioral data augers well with its strength and versatility. This adds to the emerging literature that has indicated that the use of various forms of data can considerably improve the output of machine learning models in healthcare (Maniruzzaman et al., 2020).

The findings also show that a suitable hybridization strategy should be chosen. The stacking method employed in this paper enables a hierarchical learning procedure in which the meta-classifier is the perfect blend of the results of the base models. This method has been demonstrated to be better than other ensemble methods like simple voting or averaging especially in complex prediction problems. More recent research has shown that hybrid models based on stacking and meta-learning are capable of attaining a higher degree of accuracy and robustness when forecasting chronic diseases, such as diabetes (Dohare et al., 2024).

Moreover, the steady performance boost with this study supports the general trend in machine learning research of hybrid and ensemble techniques. Extensive literature shows that hybrid models are gaining more and more popularity in medical prediction systems because of their potential to combine various learning paradigms and enhance diagnostic accuracy (Abd Zaid & Mohammed, 2024). The results of this study help in advancing this emerging field as the study has given empirical results about the effectiveness of stacking-based hybrid models in diabetes prediction.

Although the proposed model has shown good performance, it is worth noting that the effectiveness of hybrid methods relies on the careful choice of models, the preprocessing, and the integration of base learners. The findings indicate that a combination of models with varying characteristics results in improved outcomes in generalization and prediction. This understanding can be used in future studies to create more sophisticated hybrid systems to use in health care systems.

Overall, the results of the study affirm that hybrid machine learning models have immense benefits over the single-model methods in disease prediction. The fact that the proposed model performs better in terms of accuracy, error rates, and balance demonstrates its possible implementation in the real-life health care. These findings highlight the need to embrace ensemble-based approaches to designing intelligent clinical decision support systems.

5. Conclusion

This paper introduced a hybrid machine learning model that would enhance the precision of diabetes prediction based on clinical and lifestyle information. The increasing number of diabetes cases and the complications thereof make it clear that there is a great need of effective and intelligent systems to diagnose the disease. Machine learning offers a useful solution to the analysis of complicated healthcare data and facilitates the prevention of diseases at an early stage. The main aim of the study was to come up with a predictive framework that would address the weakness of individual machine learning models by combining them into various algorithms. Multiple classification algorithms were implemented and evaluated, such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbors. A structured dataset comprising of clinical features such as age, body mass index, blood glucose level, HbA1c level, hypertension status, heart disease history, and smoking behavior was used to train these models. The preprocessing of data using methods such as encoding and normalization was used to give consistency and enhance the performance of the model. The outcomes of individual models constituted a point of reference and identified their differences in strength and weaknesses. A stacking ensemble approach was made into a hybrid model to improve predictive performance. This model incorporated several base learners and a meta-classifier to make final predictions. The hybrid model also outperformed the individual models in evaluation metrics such as accuracy, precision, recall and F1-score. The decreased number of false predictions signifies the increase in reliability, which is crucial in healthcare applications where precise diagnosis is vital. The results indicate that it is necessary to integrate several machine learning methods to enhance the accuracy of predictions and the robustness of the model. Both clinical and lifestyle characteristics were also integrated, which led to a more detailed picture of patient health, raising the efficiency of the predictive model. The suggested methodology has good potential to be applicable in the context of clinical decision support systems, helping clinicians with the early detection and evaluation of risk. To sum up, it is possible to state that hybrid machine learning models provide an effective and reliable solution to the disease prediction task. The model proposed has better accuracy and consistency, and thus it can be applied in the real world in healthcare. Future research can be aimed at extending this method to other illnesses and combining it with real-time health care systems to monitor patients continuously and provide them with better treatment.

References

1. Albadri, R. F., Awad, S. M., Hameed, A. S., Mandeel, T. H., & Jabbar, R. A. (2024). A Diabetes Prediction Model Using Hybrid Machine Learning Algorithm. *Mathematical Modelling of Engineering Problems*, *11*(8).
2. Barik, S., Mohanty, S., Mohanty, S., & Singh, D. (2020). Analysis of prediction accuracy of diabetes using classifier and hybrid machine learning techniques. In *Intelligent and Cloud Computing: Proceedings of ICICC 2019, Volume 2* (pp. 399-409). Singapore: Springer Singapore.
3. Sarwar, A., Ali, M., Manhas, J., & Sharma, V. (2020). Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. *International Journal of Information Technology*, *12*(2), 419-428.
4. Rufo, D. D., Debelee, T. G., & Negera, W. G. (2022). A hybrid machine learning model based on global and local learner algorithms for diabetes mellitus prediction. *Journal of Biomimetics, Biomaterials and Biomedical Engineering*, *54*, 65-88.
5. Anbananthen, K. S. M., Busst, M. B. M. A., Kannan, R., & Kannan, S. (2023). A comparative performance analysis of hybrid and classical machine learning method in predicting diabetes. *Emerg. Sci. J*, *7*(1), 102-115.
6. Edeh, M. O., Khalaf, O. I., Tavera, C. A., Tayeb, S., Ghouali, S., Abdulsahib, G. M., ... & Louni, A. (2022). A classification algorithm-based hybrid diabetes prediction model. *Frontiers in Public Health*, *10*, 829519.
7. Modak, S. K. S., & Jha, V. K. (2024). Diabetes prediction model using machine learning techniques. *Multimedia Tools and Applications*, *83*(13), 38523-38549.
8. Jain, A., & Jain, S. (2024, November). A Novel Hybrid Model for Efficient Prediction of Diabetes Using Machine Learning. In *2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI)* (pp. 1-6). IEEE.
9. Abnoosian, K., Farnoosh, R., & Behzadi, M. H. (2023). Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. *BMC bioinformatics*, *24*(1), 337.
10. Li, X., Zhang, J., & Safara, F. (2023). Improving the accuracy of diabetes diagnosis applications through a hybrid feature selection algorithm. *Neural processing letters*, *55*(1), 153-169.
11. Farnoosh, R., Abnoosian, K., & Isewid, R. A. (2025). Two machine-learning hybrid models for predicting type 2 diabetes mellitus. *Journal of Medical Signals & Sensors*, *15*(4), 11.
12. Airlangga, G. (2024). Enhancing Diabetes Prediction Accuracy through Hybrid Machine Learning Models: A Comparative Study. *G-Tech: Jurnal Teknologi Terapan*, *8*(2), 1297-1306.

13. Chowdhury, P., Barua, P., & Uddin, M. N. (2024, September). Diabetes prediction using machine learning and hybrid deep learning ensemble technique. In *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)* (pp. 1-7). IEEE.
14. Goudar, R., & Aftab, N. (2024, April). Diabetes prediction using hybrid model. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* (pp. 1-10). IEEE.
15. Dutta, A., Hasan, M. K., Ahmad, M., Awal, M. A., Islam, M. A., Masud, M., & Meshref, H. (2022). Early prediction of diabetes using an ensemble of machine learning models. *International Journal of Environmental Research and Public Health*, *19*(19), 12378.
16. Khalid, H., Khan, A., Zahid Khan, M., Mehmood, G., & Shuaib Qureshi, M. (2023). Machine learning hybrid model for the prediction of chronic kidney disease. *Computational Intelligence and Neuroscience*, *2023*(1), 9266889.
17. Chandramouli, A., Hyma, V. R., Tanmayi, P. S., Santoshi, T. G., & Priyanka, B. J. E. C. (2023). Diabetes prediction using hybrid bagging classifier. *Entertainment Computing*, *47*, 100593.
18. Abd Zaid, M. M., & Mohammed, A. A. (2024). Hybrid models in diabetes prediction: A review of techniques, performance, and potential. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, *16*(4), 298-308.
19. Maniruzzaman, M., Rahman, M. J., Ahammed, B., & Abedin, M. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. *Health information science and systems*, *8*(1), 7.
20. Dohare, S., Pamulaparthy, L., Abdufattokhov, S., Naga Ramesh, J. V., El-Ebiary, Y. A. B., & Thenmozhi, E. (2024). Enhancing Diabetes Management: A Hybrid Adaptive Machine Learning Approach for Intelligent Patient Monitoring in e-Health Systems. *International Journal of Advanced Computer Science & Applications*, *15*(1).
21. Mustafa, M. (2023). *Diabetes prediction dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset