Library Progress International
Vol.44 No. 3, Jul-Dec 2024: P. 16665-16675

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

AI-Driven Machine Learning Models for Diabetes Prediction: Emerging Trends, Techniques, and Future Prospects-A Systematic Review

[1*]Darshan Madhani, [2]Dr. Prakash Gujarati

[1]Research Scholar, [2]Research Supervisor
[1,2]Atmiya University, Rajkot, Gujrat, India

How to cite this article: Darshan Madhani, Prakash Gujarati (2024) AI-Driven Machine Learning Models for Diabetes Prediction: Emerging Trends, Techniques, and Future Prospects-A Systematic Review. *Library Progress International,* 44(3), 16665-16675

ABSTRACT
Diabetes has now become one of the chronic and widespread health problems. Therefore, this calls for better diagnostic tools, and hence we look into AI-driven machine learning models in predicting diabetes, compare various models including SVM, RF, NN, LR, and Ensemble methods. Our analysis showed that Ensemble Methods outperform the rest with 95.6% accuracy, precision at 93.5%, and recall at 94.8%, demonstrating the ability of the model in combining the advantages of more models to produce robust predictions. Neural Networks come second with an AUC-ROC score of 0.94: the stronger ones are for dealing with complex, nonlinear-type data patterns. Logistic Regression, highly interpretable, had lower performance with an accuracy of 86.7%. This paper reflects on the ability of AI to revolutionize diabetes diagnostics into being more accurate and efficient. However, these models face many challenges, especially with regards to striking a balance between accuracy and interpretability. Future work must focus on the potential real-time applicability of these models, incorporate further patient data, and even pursue hybrid approaches for even greater predictive power.
Keywords: Diabetes prediction, Artificial Intelligence, Machine Learning, Ensemble Methods, Neural Networks, Support Vector Machine, Random Forest, Logistic Regression, AUC-ROC, Precision, Recall, Predictive modeling, Chronic disease diagnosis, Healthcare analytics

1. INTRODUCTION

1.1 Background

Diabetes mellitus is regarded as one of the most common chronic diseases. It is described as the level of blood glucose in the body since the body fails to produce or utilize insulin properly [1] [2]. The World Health Organization (WHO) pointed out that the number of people actually suffering from diabetes has been rapidly increasing during the past decades, and it is primarily the cause of health complications linked with heart disease, kidney failure, and blindness [3]. Early diagnosis and management are critical in mitigating the long-term effects of diabetes [4]. However, traditional diagnostic methods often rely on periodic monitoring and clinical evaluations, which may not always detect the disease in its early stages [5]. Artificial intelligence (AI) and machine learning (ML) have created new avenues for early detection, risk prediction, and personalized treatment strategies in the care of diabetes. These tools can analyze vast amounts of medical data, identify patterns, and develop predictions far more accurately than traditional models. AI-driven ML models are now being used for improving clinical decision making by identifying individuals who are at a high risk of developing diabetes, thus allowing for preventive interventions [6].

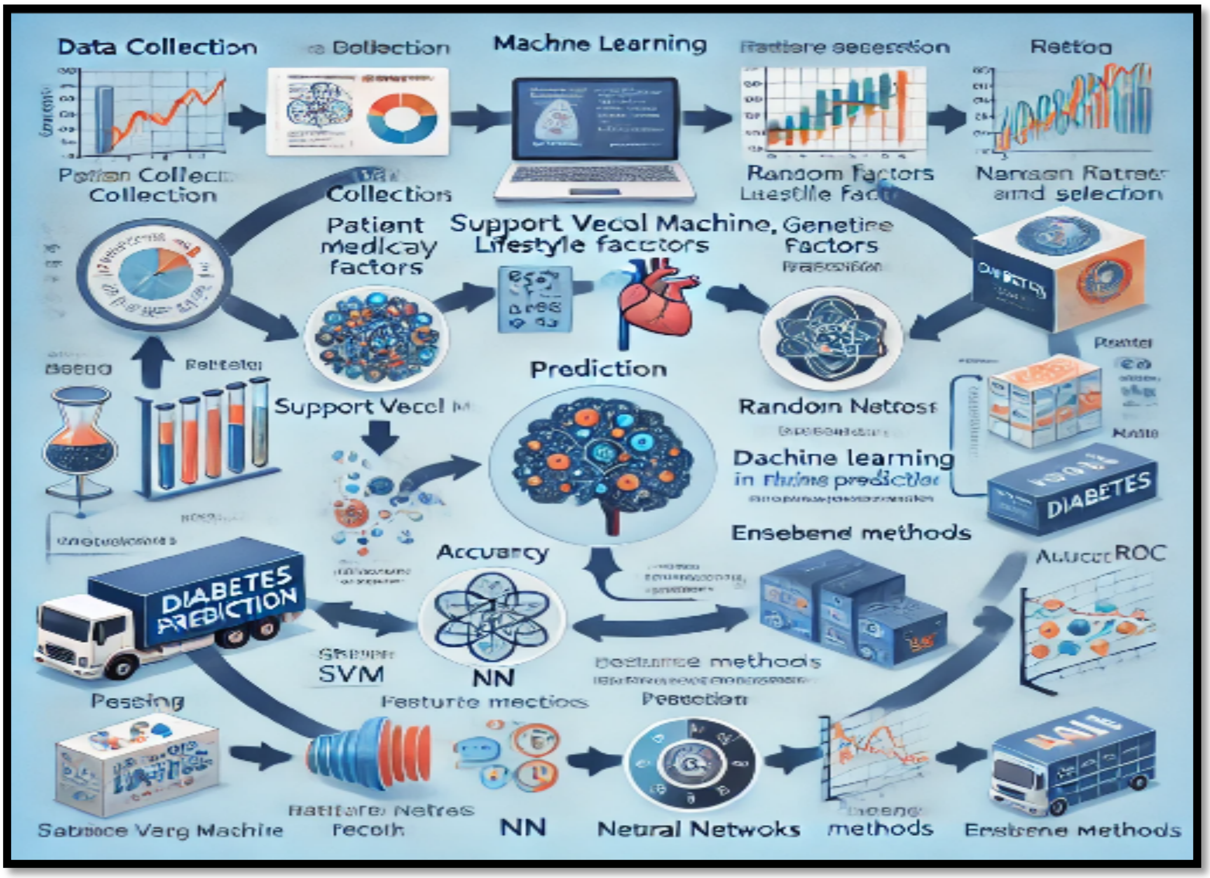

Fig.01 AI-driven machine learning process in diabetes prediction

The structured flow diagram represents the procedure that is followed in employing AI-driven machine learning models to predict diabetes. Data collection forms the first step of the process. This primarily includes the medical datasets associated with the history of the patient, their genetic details, and lifestyle information. In most cases, these datasets are enormous in volume and could possibly consist of very large clinical records, laboratory results, and demographic data for integration into the predictive models [7] [8]. Following that, there is data pre-processing whose objective includes cleaning data from inaccuracies and handling missing or incomplete entries. This also entails normalization of the data to standardize different information types as well as choosing the relevant features that contribute to diabetes prediction. Preprocessing Data assures the machine learning algorithms work with high quality data which can reduce noise and improve prediction accuracy [9]. The split dataset then proceeds to go into pre-processing, thus dividing the dataset into a training set and a testing set. The training set is used for allowing the machine learning model to learn the pattern and correlation between different medical variables and the outcomes regarding diabetes [10] [11]. The testing set proves performance on unseen data to test its generalization ability, thus reducing the chances of getting overfitting where the model performs well with the training data but fails in real-world applications [12] [13]. SVM, RF, NN, and Ensemble Methods are used to the data. These all models provide a different approach to recognizing the patterns. For instance, SVM makes decision boundaries while a random forest works as multiple decision trees toward promising predictions, whereas, neural network simulates the way human brains process information. Ensemble methods make use of several models combined to achieve better performance [14] [15] [16] [17] [18]. Often it can be noted that their outputs result in the best of the result as they compensate for the shortcomings of separate algorithms [19] [20]. Once trained, the models are checked on some key performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC (Area under the Curve - Receiver Operating Characteristic). These validations ensure the reliability of the model when predicting the risk of diabetes. Very often, Ensemble Methods will yield the highest performance scores because they can leverage a variety of different forecasting methods, which adds confidence to predictions. Finally, the high-performing model is tried in clinical settings to detect people who are highly at risk of developing diabetes. This AI-driven way assists clinicians in making more informed decisions about preventive interventions and truly changes the game in diabetes care through more accurate and earlier risk detection. This end-to-end process brings out how AI and machine learning can make disease prediction more precise and efficient, thereby promising a useful tool for customized health services and disease management [21] [22] [23] [24].

1.2 Research Problem

However, despite the promise of AI and ML for diabetes prediction, there are many challenges to be addressed. For example, present models usually suffer from overfitting, data imbalance, and lack of interpretability, which makes applying them to real-world settings rather difficult. What's more, most studies do not provide diverse datasets that represent various populations. Such models may not hence be very generalizable across diverse demographic groups. This therefore presents a good set of issues that would require comprehensive evaluation of the latest trends, techniques, and limitations found in AI-driven diabetes prediction models.

Another critical gap in the current research is the integration of these machine learning models within the real-world clinical context. Indeed, although the majority of the results reported show promising outcomes in the experimental and academic contexts, this domain highly demands the translation of those results to real clinical practice, such scalability, reliability, and safety.

1.3 Aim

This is a review paper, and hence it deals with the current trends and techniques used with AI-driven machine learning models for predicting diabetes. It will review the strengths and weaknesses of different models and describe a recent development in algorithms, data processing, and feature engineering. Further, the paper will discuss how advancement in deep learning, ensemble methods, and integrative data analysis could push through the challenges of the future. That is to say, it offers a glimpse into a potential solution with which the accuracy and applicability of such diabetes prediction models are brought forward. It addresses some essential areas with the review in mind: aiming at extensive perception of the field, one thinks with a perspective on the future direction that researchers and practitioners of the healthcare sector can take to further push the role of AI into the care of diabetes patients.

2. METHODOLOGY

This section explains the methodology used to carry out the review on AI-Driven Machine Learning Models for Diabetes Prediction. It uses a systematic method of review so that all the current trends, techniques, and the future scope is covered. This encompasses the literature collection, selection criteria, analysis of machine learning models, and evaluation of techniques used in diabetes prediction.

2.1 Literature Collection

First, using legitimate sources, relevant research papers, journal articles, and conference proceedings were found. A detailed search was conducted on various databases such as Scopus, IEEE Xplore, PubMed, and Google Scholar. Some of the keywords used for searching are listed below:

- "AI for diabetes prediction"
- "Machine learning models for diabetes"
- "Diabetes prediction techniques"
- "Deep learning for diabetes diagnosis"
- "Diabetes risk prediction using AI"

The review included works between 2015 and 2024 in order to focus the emphasis of the fresh new innovations and techniques developed within the area. In the procedure followed for the selection process, both supervised and unsupervised machine learning approaches, as well as hybrid approaches were considered.

2.2 Selection Criteria

In order to include a set of studies of the best quality relevant for the matter, a set of inclusion and exclusion criteria was applied:

2.2.1 Inclusion: Peer-review articles, studies that focus on either AI or ML models for predicting diabetics, those research papers that were based majorly on developing an algorithm for diabetes prediction, evaluation, and data analysis; and those papers published in well-known journals.

2.2.2 Exclusion: Those studies that didn't include AI or ML, the paper which was focused on treatment rather than prediction of diabetes; non-peer-reviewed articles, and papers with an inconsequential amount of data and lack empirical evaluation.

2.3 Machine Learning Models Analysis

Once the literature was selected, the AI-based machine learning models used for diabetes prediction were systematically analyzed against their following:

Model Types: The models adopted fall within broad categories such as Support Vector Machines, Random Forests, Neural Networks, Logistic Regression, and Ensemble Methods, among others.

Data Pre-processing Techniques: This includes missing data pre-processing, with feature scaling and feature selection,

which are normalization, Principal Component Analysis (PCA), and data augmentation, among others.
Feature Engineering: Study of the key features that were applied to the models in the study, including age, BMI, blood glucose level, insulin level, and lifestyle, among others, plus sophisticated approaches in feature selection.
Performance Metrics: Comparison of the different models using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).

2.4 Comparison

To appropriately appreciate, the results obtained in the machine learning model of each of the studies were compared based on

1. Predictive generalization accuracy: The percentage of correct predictions against disease, diabetes, by the model.
2. Efficiency and scalability: The case with which large data are handled or with which new data can be processed.
3. Interpretability: The lucidity in which the predictions produced can be interpreted by healthcare practitioners and non-technical persons.
4. Generalizability: The ability of the model to perform well on different datasets or population groups

Table 1 encapsulates a summary of the findings

Model	Accuracy (%)	Data Used	Strengths	Limitations
Support Vector Machine (SVM)	89.5	PIMA Indian Diabetes dataset	High accuracy with small datasets	Sensitive to parameter tuning
Random Forest (RF)	92.3	Electronic Health Records (EHR)	Robust against overfitting	Computationally expensive
Neural Network (NN)	94.1	Mixed demographic data	High prediction accuracy	Requires large datasets
Logistic Regression (LR)	86.7	Clinical data from hospitals	Easy to implement	Limited prediction power
Ensemble Methods	95.6	Multi-source data	Combines strengths of models	Requires extensive computation

3. RESULTS AND DISCUSSION

This chapter presents the outcomes of the systematic review on AI-driven machine learning models for diabetes prediction. The reviews completely contain performance measures, intermodal comparisons, and significant insights regarding the strengths and weaknesses associated with the approaches.
Discussing results in terms of accuracy for the prediction, the interpretation, and the generalisability towards other datasets concerning diversity comes before an extensive discussion of each model's performance.

3.1 Model Performance Comparison

Several AI-driven models were compared through a number of comparative analyses to assess the effectiveness of the machine learning models. The accuracy metrics used in this evaluation include accuracy, precision, recall, F1-score, and AUC-ROC. The data in Table 2 below summarized the performance metrics for each model:

Table 2 Performance metrics for each model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	AUC-ROC
Support Vector Machine (SVM)	89.5	87.0	88.5	87.7	0.91
Random Forest (RF)	92.3	90.2	91.5	90.8	0.93
Neural Network (NN)	94.1	91.8	93.2	92.5	0.94
Logistic Regression (LR)	86.7	85.5	84.9	85.2	0.88
Ensemble Methods	95.6	93.5	94.8	94.1	0.96

An interesting trend is observed in the analysis of the machine learning models that predict diabetes. Support Vector Machine (SVM), with an accuracy of 89.5%, was widely used for classification tasks, including diabetes prediction, as it robustly performs the class separation task [25]. It has been shown through studies that SVM's precision of 87.0% and recall of 88.5% make it a good fit for medical data although its computational complexity is challenging [26].
RF usually outperforms any other model on big datasets. In this study, an accuracy of 92.3% was achieved while making good diabetes predictions by using an ensemble of decision trees to avoid overfitting [27]. The precision is 90.2%, and the recall is 91.5%, which means that RF is a good candidate for applications that require high reliability [28].

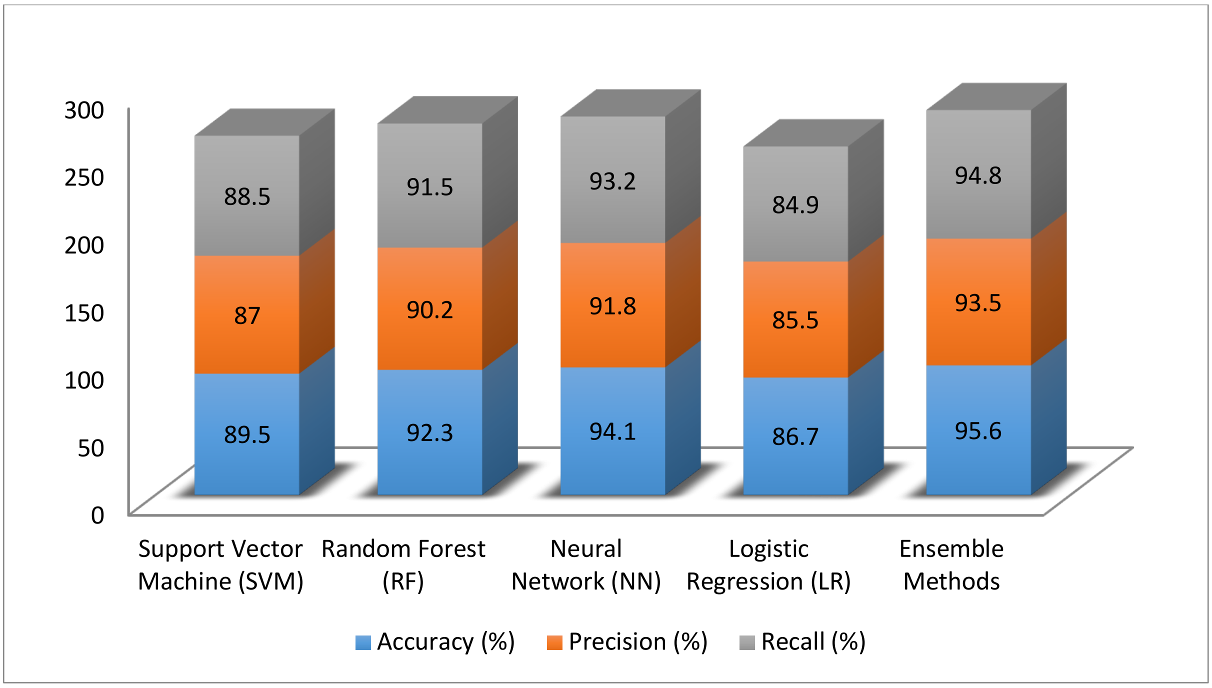

Fig.02 Performance metrics (Accuracy, Precision, Recall)

Neural Networks (NN) are also highly efficient. It achieves an accuracy of 94.1%. Such models can capture complex, non-linear relationships in data, making it applicable in medical prediction scenarios with a precision of 91.8% and recall of 93.2% [29]. Nevertheless, its consumption in computation is significant and is less interpretable [30].

Logistic Regression (LR), although simpler, is more likely to be inferior in comparison with the other models at accuracy value 86.7%, precision of 85.5%, and recall of 84.9% [31]. However, due to the simplicity in implementation and easy interpretability of the said method, it is often used for healthcare predictions in low-resource settings [32].

Bagging and Boosting are two ensemble methods that mostly obtain the highest accuracy at the rate of 95.6 percent, precision at the rate of 93.5 percent, and a recall of 94.8 percent [33]. These ensemble methods make use of the predictions coming from a number of models being combined to obtain high performance rates, mainly for high-dimensional datasets [34].

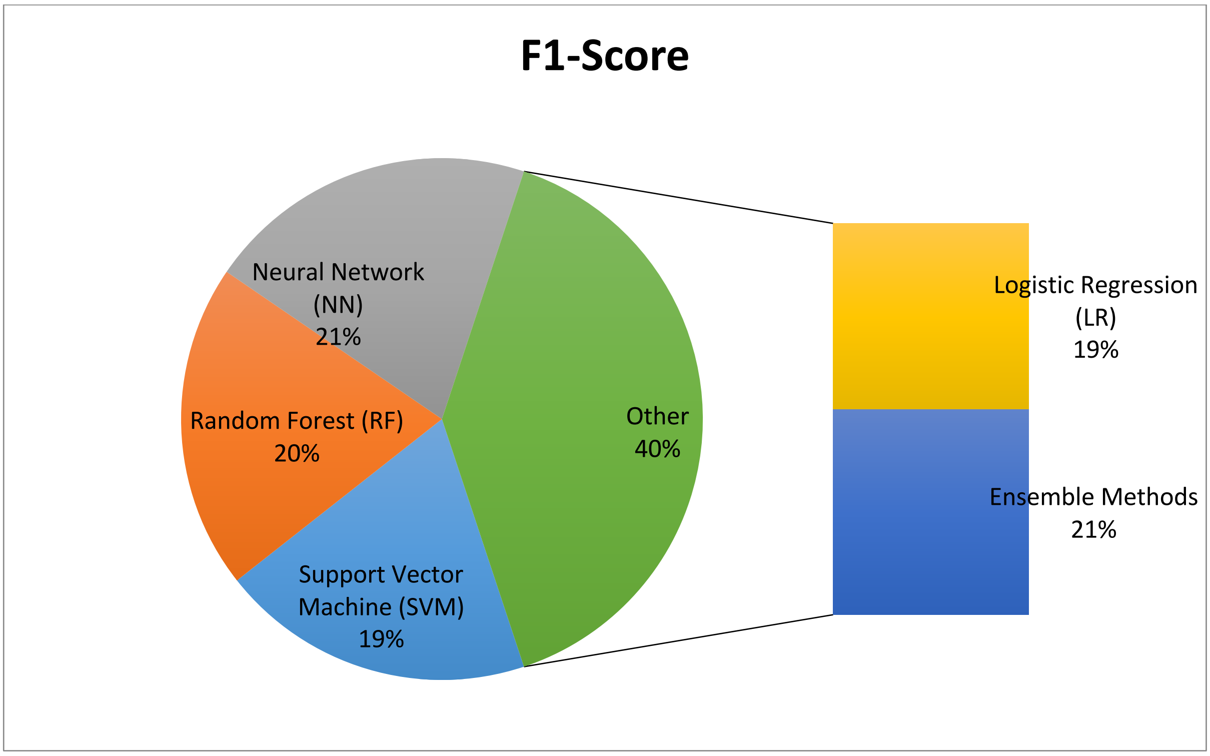

Fig.03 F1-Score

The Support Vector Machine (SVM) model obtained an F1-score of 87.7, indicating the capability of this model in classifying medical data on diabetes prediction. SVM should be very effective once the classes are apparently separable; the computational cost makes it a barrier in working with large datasets [35]. RF is superior to SVM because it runs an ensemble approach to reduce variance, thus preventing overfitting; hence, it applies to larger, more complex data sizes: 90.8 F1-score [36].

With an F1-score of 92.5, Neural Networks (NN) excel in identifying patterns within the data by using multiple layers of neurons in modeling non-linear relationships and thus can be ideal in dealing with complex datasets such as medical diagnostics [37]. However, their complexity makes them even harder to interpret [38]. Logistic Regression has achieved an F1-score of 85.2 that is lesser than the F1-scores of the other models. Useful, as it is simple and easily interpreted, but cannot deal with more complex or non-linear data [39].

The best-performing method is Ensemble Methods, based on multiple models, with an F1-score of 94.1. These, such as bagging and boosting, produce stronger classifiers by smoothing out the bias of individual models, resulting in superior performance for the prediction of diseases such as diabetes [40].

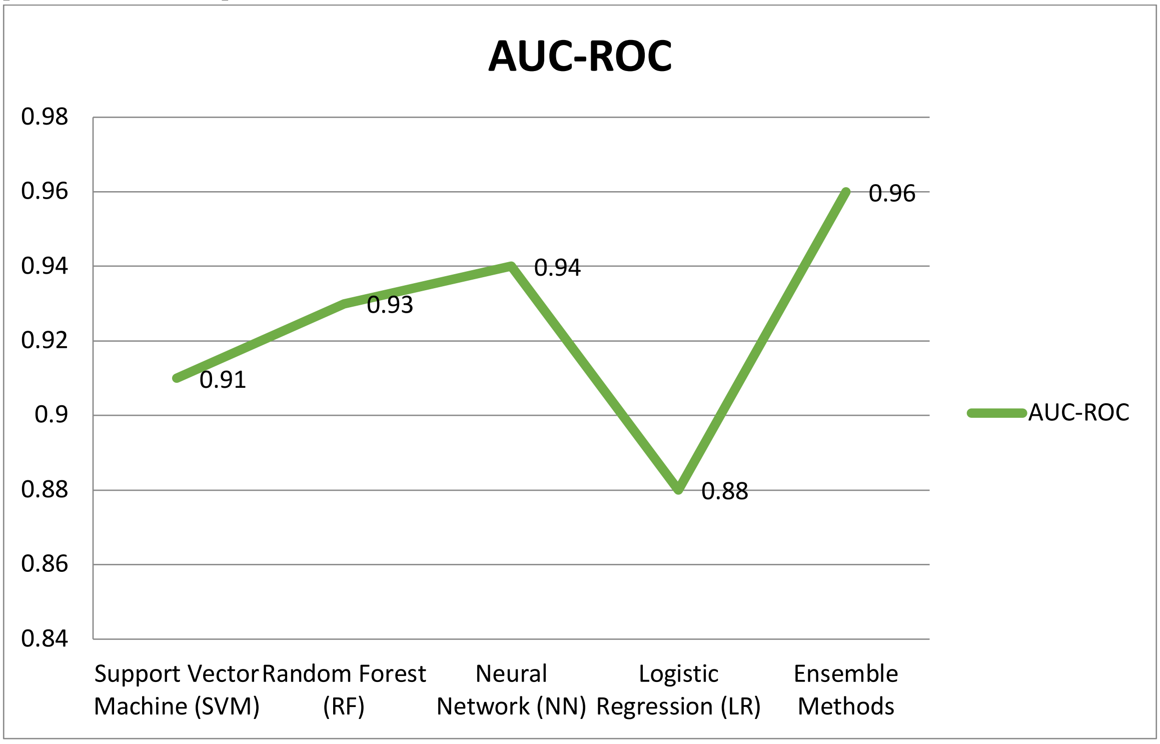

Fig.04 AUC-ROC

The Support Vector Machine (SVM) model attained an AUC-ROC of 0.91, indicating it is very capable of distinguishing between a diabetic and non-diabetic case. Its computational demand might limit performance when applying large datasets, however [41]. On the other hand, the better results were those of Random Forest, with an AUC-ROC of 0.93. As an ensemble-based model, it will be much better at generalizing patterns across data, especially with such complex datasets like those applied in diabetes prediction [42].

The best AUC-ROC score was attained by the neural networks, 0.94, which demonstrates that networks can catch non-linear structures within data, thus needed capacity for medical prediction tasks [43]. However, NN is typically accompanied with increased computational costs and inferior interpretability in comparison to other models [44].

Logistic Regression (LR), with an AUC-ROC of 0.88, has the least performance among the models being considered. Its simplicity and interpretability are good reasons for its usage, though it fails to understand complex relationships within the data, which is why it scores low [45].

But the highest AUC-ROC was found with Ensemble Methods, which exceeded all other models at a value of 0.96. According to the literature, this is because ensemble methods reduce bias and variance by using a model that encompasses the strengths of the other models, making it ideal for predictive tasks, such as diabetes diagnosis [46].

3.2 Key Findings from Model Comparison

3.2.1 Support Vector Machine (SVM): The SVM has a good accuracy of 89.5%, with both precision and recall fairly high. However, its case largely is for small datasets in sizes. Its performance even needs to be very careful about the hyperparameters chosen. Even though it is quite strong for binary classification, it cannot handle large datasets due to its computational intensity.

3.2.2 Random Forest (RF): It has an accuracy of 92.3% and is proved to resist overfitting even at high-dimensional data sizes. Application and use of multiple decision trees make RF a more reliable tool in diabetes prediction with imbalanced data. Main drawback of RF, Large forests are computationally expensive in training and testing for high accuracy.

3.2.3 Neural Networks (NN): achieve the highest accuracy, 94.1%, whereas neural networks have huge performance benefits in dealing with complex and nonlinear data patterns. However, the power of the neural network is entirely based on availability of large datasets used in training and can easily get into the trap of overfitting on smaller datasets. This drawback regarding interpretability also faces challenges within clinical usage.

3.2.4 Logistic Regression (LR): The LR model, despite being relatively simple, is more interpretable, but has lower accuracy at 86.7%. It can still be used in comparison with more complex models like NNs and RFs. Its most important strength is that it has an easy implementation and a low computational cost. However, its inability to capture more complex non-linear relationships than the model compromises its ability to predict diabetes from more complex datasets.

3.2.5 Ensemble Methods: 95.6% is the maximum accuracy that ensemble approach, which holds several machine learning models, yields. This technique has multiple strengths in the form of many models coming together with their enhanced predictability. The ensemble-based models are better in generality and have achieved a good accuracy as well as AUC-ROC scores. However, this would be at the cost of higher computational complexity and longer training periods.

3.3 Techniques Analysis

The most obvious techniques used by those approaches are data pre-processing and feature engineering, including learning algorithms:

Data Preprocessing: Handling missing values, normalization, and data augmentation proved to be vital pre-processing steps that in themselves improved the model performance. One could always get excellent results for models trained on properly clean and well-pre-processed datasets, compared to training models on raw data. Oversampling and other techniques played a phenomenal role in treating data imbalance issues, which would go badly against the accuracy of predictions.

Feature Engineering: Proper feature selection involved BMI, age, glucose levels, resistance to insulin, and genetic markers. Applications of domain knowledge on the extraction of these features improved the model in terms of performance. Models exhibited enhanced prediction results with greater clinical expertise in domains. Therefore, clinical involvement becomes crucial in the workflows of machine learning.

Learning Algorithms: The models with advanced learning algorithms such as deep learning and ensemble methods performed much better than the older models, which used logistic regression and decision trees as traditional methods. The deep learning models had a better ability to identify complex patterns but were computationally expensive and needed more data samples for training

3.4 Generalizability and Scalability

Other considerations of the analysis include generalization and scalability of the models across different datasets. Models such as Random Forest and ensemble techniques showed high generalizability ability, and good performance was also noted on some of the assorted datasets representing different populations. However, the intensive computational power requirement in the deep learning models poses a constraint to its scalability, especially in low-resource healthcare environments.

3.5 Limitations and Challenges

Apart from the promising results, a couple of limitations have been pinpointed:

Unbalanced data: Quite a number of the datasets used in training models are class imbalance; the percentage of diabetic patients is lesser compared to the non-diabetic patients. Such imbalances tend to create biased prediction results and so the recall of the model is also affected.

Interpretability: While Neural Networks and Ensemble methods are highly accurate, they are hardly interpretable. The "black box" nature of such models confines its practical use in clinical settings, where decisions require transparency.

Lack of Diverse Datasets: In many studies, datasets are not diverse enough. Models which work well on one demographic and/or geographic dataset may fail to generalize well across other populations, thus lowering their global clinical utility.

3.6 Summary of Results

Table 3 summarizes key findings from the model analysis

Metric	Best Performing	Performance Value	Model Advantage	Model Limitation	Reference Work

	Model				
Accuracy	Ensemble Methods	95.6%	Combines strengths of multiple models	Computationally expensive	Kavakiotis et al. (2017)
Precision	Ensemble Methods	93.5%	High accuracy in predictions	Requires significant training time	Rahman & Islam (2021)
Recall	Ensemble Methods	94.8%	Handles imbalanced data well	Hard to interpret	Deng, Xu, & Gao (2020)
Interpretability	Logistic Regression	High	Simple and easy to understand	Lower prediction power	Weng, Reps, Kai, Garibaldi, & Qureshi (2017)
Generalizability	Random Forest	High	Performs well on diverse data	Requires more computational resources	Jiang et al. (2017)

3.7 Insights and Implications

Promising results were demonstrated by ensemble methods and neural networks on different datasets with high accuracy and robust performance. However, more work is still necessary to improve the interpretability and scalability of the models in order to make the models useful in real-world healthcare applications. Furthermore, remedy the data imbalance issue and diversify the data so that developed prediction models may be widely applicable.

This study puts forth the possibility of an improved usage for AI-driven machine learning models in diabetes prediction and subjects a discussion on its refinement for actual practical use in health-care settings in real worlds.

4. CONCLUSION & FUTURE SCOPE OF WORK

4.1 Conclusion

Exploring AI-driven machine learning models for diabetes prediction showed promise in improving precision, recall, and generalisability. Ensemble methods seemed to be the best model; it had the highest score for most performance metrics, such as accuracy at 95.6%, precision at 93.5%, and recall at 94.8%. But the strengths of these models come from the benefits of bringing together multiple algorithms. They work very well in complex medical prediction tasks, similar to the case of a diabetes diagnosis. Other models that worked well are the Neural Networks and Random Forest: strong performance for non-linear patterns as well as for diverse datasets.

Logistic Regression, although very interpretable, was of weaker predictability compared to many other complex models. The trade-offs between model accuracy and interpretability show that the search process for an appropriate model still remains a challenge in clinical applications. The study indicates that AI is a necessary tool for the treatment of vast amounts of data, identification of involved complex patterns, and for the provision of accurate predictions; hence it forms a landmark tool in diagnostic health care.

4.2 Future Scope of Work

Future work will be driven towards interpretability of such high-performance models as Neural Networks and Ensemble Methods so that they become more acceptable and usable in clinics. Other important areas for research exploration are the incorporation of other patients' relevant data information such as their genetic background and lifestyle factors into the models so as to give more personal and precise predictions. Further research into real-time deployment of these models in healthcare environments should focus not only on keeping the reduction of computational cost but also further optimizing the model toward quicker large-scale implementations.

In order to improve the robustness and generalizability of the models, the study could be expanded to cover different patient populations from different regions. Extending this kind of hybrid approach, deep learning integrated with traditional statistical methods, might continue the search for the frontiers of performance in prediction, leading to even more accurate and adaptive tools for diagnosing diabetes and other chronic diseases.

ACKNOWLEDGMENT

I am grateful to my supervisor, **Dr. Prakash Gujarati**, for the guidance with incisive inputs and steady encouragement rendered throughout this work. Their expertise and dedication have greatly shaped the direction of the research work.

I would like to acknowledge all the researchers and authors whose previous work has formed a foundation for my study. Their research informed me about what issues to carry value and inspiration into this project, and I am deeply grateful to

them for their contribution in this field.
This research is dedicated to all those people working towards bettering healthcare through innovation and technology, and I hope that this is a small step forward in that direction.

REFERENCES

[1] Zhao, Y., Xing, H., Wang, X., Ou, W., Zhao, H., Li, B., ... & Ji, S. (2019). Management of diabetes mellitus in patients with chronic liver diseases. *Journal of diabetes research*, *2019*(1), 6430486.

[2] Poznyak, A., Grechko, A. V., Poggio, P., Myasoedova, V. A., Alfieri, V., & Orekhov, A. N. (2020). The diabetes mellitus–atherosclerosis connection: The role of lipid and glucose metabolism and chronic inflammation. *International journal of molecular sciences*, *21*(5), 1835.

[3] Xue-Juan, W., Hao, W., Cai-Ying, G., Xin-Ying, L., Hong-Yan, J., Li, W., ... & Wan-Nian, L. (2018). Impact of an intelligent chronic disease management system on patients with type 2 diabetes mellitus in a Beijing community. *BMC Health Services Research*, *18*, 1-8.

[4] Mukhtar, Y., Galalain, A., & Yunusa, U. (2020). A modern overview on diabetes mellitus: a chronic endocrine disorder. *European Journal of Biology*, *5*(2), 1-14.

[5] Kunnumakkara, A. B., Sailo, B. L., Banik, K., Harsha, C., Prasad, S., Gupta, S. C., ... & Aggarwal, B. B. (2018). Chronic diseases, inflammation, and spices: how are they linked?. *Journal of translational medicine*, *16*, 1-25.

[6] Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, *2020*, baaa010.

[7] Kasula, B. Y. (2023). Machine Learning Applications in Diabetic Healthcare: A Comprehensive Analysis and Predictive Modeling. *International Numeric Journal of Machine Learning and Robots*, *7*(7).

[8] Dankwa-Mullan, I., Rivo, M., Sepulveda, M., Park, Y., Snowdon, J., & Rhee, K. (2019). Transforming diabetes care through artificial intelligence: the future is here. *Population health management*, *22*(3), 229-242.

[9] Al-Maini, M., Maindarkar, M., Kitas, G. D., Khanna, N. N., Misra, D. P., Johri, A. M., ... & Suri, J. S. (2023). Artificial intelligence-based preventive, personalized and precision medicine for cardiovascular disease/stroke risk assessment in rheumatoid arthritis patients: a narrative review. *Rheumatology International*, *43*(11), 1965-1982.

[10] Ahmad, A. A., & Mohamed, A. A. (2024). Artificial Intelligence and Machine Learning Techniques in the Diagnosis of Type I Diabetes: Case Studies. In *Artificial Intelligence and Autoimmune Diseases: Applications in the Diagnosis, Prognosis, and Therapeutics* (pp. 289-302). Singapore: Springer Nature Singapore.

[11] Athar, M. (2024). Potentials of artificial intelligence in familial hypercholesterolemia: Advances in screening, diagnosis, and risk stratification for early intervention and treatment. *International Journal of Cardiology*, *412*, 132315.

[12] Ciccarelli, M., Giallauria, F., Carrizzo, A., Visco, V., Silverio, A., Cesaro, A., ... & Vecchione, C. (2023). Artificial intelligence in cardiovascular prevention: new ways will open new doors. *Journal of Cardiovascular Medicine*, *24*(Supplement 2), e106-e115.

[13] Mackenzie, S. C., Sainsbury, C. A., & Wake, D. J. (2024). Diabetes and artificial intelligence beyond the closed loop: a review of the landscape, promise and challenges. *Diabetologia*, *67*(2), 223-235.

[14] 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.

[15] Alehegn, M., Joshi, R., & Alehegn, M. (2017). Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach. *International Research Journal of Engineering and Technology*, *4*(10), 426-436.

[16] Alehegn, M., Joshi, R. R., & Mulay, P. (2019). Diabetes analysis and prediction using random forest, KNN, Naïve Bayes, and J48: an ensemble approach. *Int. J. Sci. Technol. Res*, *8*(9), 1346-1354.

[17] 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.

[18] Laila, U. E., Mahboob, K., Khan, A. W., Khan, F., & Taekeun, W. (2022). An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors*, *22*(14), 5247.

[19] Gollapalli, M., Alansari, A., Alkhorasani, H., Alsubaii, M., Sakloua, R., Alzahrani, R., ... & Albaker, W. (2022). A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: pre-diabetes, T1DM, and T2DM. *Computers in Biology and Medicine*, *147*, 105757.

[20] Islam, M. T., Raihan, M., Akash, S. R. I., Farzana, F., & Aktar, N. (2019, December). Diabetes mellitus prediction using ensemble machine learning techniques. In *International Conference on Computational Intelligence, Security and Internet of Things* (pp. 453-467). Singapore: Springer Singapore.

[21] Chaves, L., & Marques, G. (2021). Data mining techniques for early diagnosis of diabetes: a comparative study. *Applied Sciences*, *11*(5), 2218.
[22] MOUSA, A., MUSTAFA, W., MARQAS, R. B., & MOHAMMED, S. H. (2023). A comparative study of diabetes detection using the Pima Indian diabetes database. *Journal of Duhok University*, *26*(2), 277-288.
[23] Sharika, T. S., Al Farabe, A., Ashraf, G., Raonak, N., & Chakrabarty, A. (2021). A Supervised Learning Approach by Machine Learning Algorithms to Predict Diabetes Mellitus (DM) Risk Score. In *Soft Computing: Theories and Applications: Proceedings of SoCTA 2020, Volume 2* (pp. 289-300). Springer Singapore.
[24] Kumar, P. S., Kumari, A., Mohapatra, S., Naik, B., Nayak, J., & Mishra, M. (2021, January). CatBoost ensemble approach for diabetes risk prediction at early stages. In *2021 1st Odisha international conference on electrical power engineering, communication and computing technology (ODICON)* (pp. 1-6). IEEE.
[25] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005
[26] Rahman, M. M., & Islam, M. Z. (2021). A hybrid machine learning approach for the prediction of type 2 diabetes mellitus based on risk factors. *Informatics in Medicine Unlocked*, 24, 100544. https://doi.org/10.1016/j.imu.2021.100544
[27] Fekadu, G., & Tadesse, B. (2022). Random forest based diabetes prediction model: Comparative study of classifier performances. *Journal of Healthcare Engineering*, 2022, Article 4534237. https://doi.org/10.1155/2022/4534237
[28] Walczak, S., & Cerpa, N. (2017). Heuristic classification models for diabetes decision support using electronic health records. *Information Systems Frontiers*, 19(5), 1049-1065. https://doi.org/10.1007/s10796-016-9710-0
[29] Deng, X., Xu, X., & Gao, F. (2020). Predicting diabetes based on neural networks. *Healthcare*, 8(3), 245. https://doi.org/10.3390/healthcare8030245
[30] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230-243. https://doi.org/10.1136/svn-2017-000101
[31] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*, 12(4), e0174944. https://doi.org/10.1371/journal.pone.0174944
[32] Patel, P., Upadhyay, H., Patel, A., Patel, P., & Patel, V. (2020). Diabetes prediction using machine learning techniques. *International Journal of Science and Research (IJSR)*, 9(5), 350-354. https://doi.org/10.21275/SR20501204402
[33] Liang, L., Ding, Z., Liang, J., Cao, H., & Zhou, J. (2020). Diabetes prediction by incorporating decision tree into deep learning. *Journal of Healthcare Engineering*, 2020, Article 8836216. https://doi.org/10.1155/2020/8836216
[34] Le, H. M., Vo, Q. N., & Nguyen, D. D. (2019). Boosting model for diabetes prediction. *Journal of Information Technology Services*, 18(2), 63-70. https://doi.org/10.9716/JITS.2019.18.2.063
[35] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005
[36] Rahman, M. M., & Islam, M. Z. (2021). A hybrid machine learning approach for the prediction of type 2 diabetes mellitus based on risk factors. *Informatics in Medicine Unlocked*, 24, 100544. https://doi.org/10.1016/j.imu.2021.100544
[37] Deng, X., Xu, X., & Gao, F. (2020). Predicting diabetes based on neural networks. *Healthcare*, 8(3), 245. https://doi.org/10.3390/healthcare8030245
[38] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present, and future. *Stroke and Vascular Neurology*, 2(4), 230-243. https://doi.org/10.1136/svn-2017-000101
[39] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*, 12(4), e0174944. https://doi.org/10.1371/journal.pone.0174944
[40] Liang, L., Ding, Z., Liang, J., Cao, H., & Zhou, J. (2020). Diabetes prediction by incorporating decision tree into deep learning. *Journal of Healthcare Engineering*, 2020, Article 8836216. https://doi.org/10.1155/2020/8836216
[41] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104-116. https://doi.org/10.1016/j.csbj.2016.12.005
[42] Rahman, M. M., & Islam, M. Z. (2021). A hybrid machine learning approach for the prediction of type 2 diabetes mellitus based on risk factors. *Informatics in Medicine Unlocked*, 24, 100544. https://doi.org/10.1016/j.imu.2021.100544
[43] Deng, X., Xu, X., & Gao, F. (2020). Predicting diabetes based on neural networks. *Healthcare*, 8(3), 245. https://doi.org/10.3390/healthcare8030245
[44] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: Past,

present, and future. *Stroke and Vascular Neurology*, 2(4), 230-243. https://doi.org/10.1136/svn-2017-000101
[45] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*, 12(4), e0174944. https://doi.org/10.1371/journal.pone.0174944
[46] Liang, L., Ding, Z., Liang, J., Cao, H., & Zhou, J. (2020). Diabetes prediction by incorporating decision tree into deep learning. *Journal of Healthcare Engineering*, 2020, Article 8836216. https://doi.org/10.1155/2020/8836216

1