IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

A Survey on Diabetes Prediction Using Machine Learning Techniques

Dr. Prashant Wakhare[1], Muskan Tamboli[2], Aditi Wakode[3], Sharvari Yadav[4]
Professor, Department of AI & DS[1]
Student, Department of AI & DS[2-4]
AISSMS Institute of Information Technology, Pune, Maharashtra, India
prashant.wakhare@aissmsioit.org[1], muskan.tamboli@aissmsioit.org[2],
aditi.wakode@aissmsioit.org[3], ysharvari01@gmail.com[4]

Abstract: *Diabetes is one of the most common chronic diseases worldwide and poses a serious threat to public health. Early prediction of diabetes can help in timely diagnosis and prevention of severe complications. Traditional diagnostic methods often rely on clinical tests and expert analysis, which can be time-consuming and expensive. In recent years, machine learning and artificial intelligence techniques have been widely used to improve the accuracy and efficiency of diabetes prediction systems. This survey paper reviews recent research (2019–2025) on diabetes prediction using machine learning and deep learning approaches. Various models such as Decision Trees, Support Vector Machines, Random Forest, Neural Networks, and ensemble techniques are analyzed and compared based on their performance, accuracy, and applicability. The study also discusses challenges such as data quality, model complexity, and real-world implementation issues. It is observed that machine learning-based systems significantly improve prediction accuracy, but further research is required to develop cost-effective and scalable healthcare solutions.*

Keywords: Diabetes Prediction, Machine Learning, Artificial Intelligence, Deep Learning, Healthcare Analytics, Classification Algorithms, Ensemble Learning, Early Diagnosis

I. INTRODUCTION

Diabetes is a long-term metabolic disorder that affects millions of people across the globe. It occurs when the body is unable to properly regulate blood glucose levels, leading to serious health complications such as heart disease, kidney failure, and nerve damage. Early detection and diagnosis are critical to managing the disease and improving patient outcomes.

Traditional methods for diagnosing diabetes mainly depend on laboratory tests and physician expertise. While effective, these methods may not always provide early prediction or preventive insights. With the growth of digital healthcare and data availability, machine learning techniques have emerged as powerful tools for analyzing medical data and predicting diseases.

Machine learning algorithms can process large datasets, identify hidden patterns, and make accurate predictions. Various approaches such as supervised learning, deep learning, and ensemble models have been applied in diabetes prediction. These systems can assist healthcare professionals by providing faster and more reliable results. However, challenges such as data imbalance, overfitting, and model interpretability still need to be addressed.

II. LITERATURE SURVEY

In paper **[1],** the authors used machine learning algorithms for diabetes prediction and achieved good accuracy. However, the model performance depends heavily on data quality and preprocessing techniques. In paper **[2],** a machine learning-based system was proposed for diabetes detection using classification algorithms. While the system is effective, it lacks scalability for large datasets. In paper **[3],** an ensemble learning approach combining multiple

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 18, April 2026 Impact Factor: 8.2

classifiers was introduced. This method improves prediction accuracy but increases computational complexity. In paper **[4],** various machine learning techniques were applied and compared. The study provides useful insights, but real-time implementation was not considered. In paper **[5],** different algorithms were evaluated for diabetes prediction. The results show variation in accuracy, indicating the need for model optimization. In paper **[6],** supervised machine learning techniques were used for prediction. Although the model performs well, it requires proper feature selection to avoid overfitting. In paper **[7],** a decision support system using machine learning and deep learning was developed. The system enhances prediction capability but requires high computational resources. In paper **[8],** a comprehensive review of machine learning-based diabetes prediction was presented. It highlights recent advancements but lacks practical implementation strategies. In paper **[9],** a comparative study between machine learning and deep learning approaches was conducted. Deep learning models showed better performance but required large datasets. In paper **[10],** a prediction model using advanced machine learning techniques was proposed. The model achieved high accuracy but is complex to implement. In paper **[11],** a detailed review of AI-based diabetes prediction and management systems was provided. It emphasizes the importance of integrating AI in healthcare. In paper **[12],** deep learning techniques were used for pediatric diabetes prediction. While effective, it requires specialized datasets. In paper **[13],** a machine learning model for diabetes prediction was developed, showing improved results but lacking real-time deployment. In paper **[14],** a mobile-based diabetes prediction system was proposed. It improves accessibility but depends on user input accuracy. In paper **[15],** a survey on machine learning techniques was presented, discussing trends and future directions. In paper **[16],** a comparative study of different models was conducted. The results highlight the importance of selecting appropriate algorithms. In paper **[17],** feature selection techniques were used to improve prediction accuracy. However, the process can be time-consuming. In paper **[18],** a comparison between machine learning and deep learning algorithms was performed. Deep learning provided better results but required more resources. In paper **[19],** data mining techniques were applied for diabetes prediction. The system performs well but depends on dataset size and quality. In paper **[20],** an ensemble deep learning model was proposed, achieving high accuracy but increasing system complexity.

III. ANALYSIS OF EXISTING SYSTEMS

Existing diabetes prediction systems primarily rely on machine learning algorithms such as Random Forest, Support Vector Machines (SVM), Decision Trees, and Artificial Neural Networks. These models have demonstrated high accuracy in controlled experimental settings and are widely adopted in research for predictive analysis. However, their effectiveness is often dependent on the availability of large, high-quality datasets, which may not always be accessible in real-world healthcare environments.

Furthermore, many of these systems involve complex computations, particularly in the case of deep learning models, making them resource-intensive and difficult to deploy in real-time applications. Another significant limitation is the lack of interpretability, especially in advanced models, which makes it challenging for healthcare professionals to trust and understand the predictions. Additionally, issues such as data imbalance, overfitting, and sensitivity to preprocessing techniques further impact the reliability and generalizability of these systems. As a result, despite achieving promising results in research settings, existing systems face considerable challenges in practical implementation.

IV. PROPOSED SYSTEM

To overcome the limitations of existing approaches, a more efficient and scalable diabetes prediction system is proposed. The proposed framework focuses on balancing accuracy, computational efficiency, and real-world applicability. It emphasizes the use of lightweight and optimized machine learning models that can perform effectively even with limited computational resources. By incorporating feature selection techniques, the system aims to reduce dimensionality and improve model performance while minimizing unnecessary complexity.

In addition, the proposed system considers the integration of ensemble learning methods to enhance prediction accuracy without significantly increasing computational cost. The design is oriented toward real-time deployment,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

making it suitable for practical healthcare applications such as mobile and web-based platforms. This approach ensures that the system remains cost-effective, accessible, and efficient, thereby addressing the key challenges associated with existing diabetes prediction models and improving its usability in real-world scenarios.

Sr.No	Author and Year (Citation)	Year	Journal/Conference	Advantages	Disadvantages
1	Mujumdar, A. & Vaidehi, V. (2019)	2019	Procedia Computer Science	Uses multiple ML algorithms for comparison	Limited dataset
2	Rani, K. J. (2020)	2020	International Journal of Scientific Research in Computer Science, Engineering and Information Technology	Simple and easy implementation	Less accuracy compared to ensemble models
3	Hasan, M. K. et al. (2020)	2020	IEEE Access	Ensemble classifiers improve accuracy	Computationally expensive
4	Singh, K. et al. (2019)	2019	ICICRS, IEEE	Comparative analysis of ML techniques	No deep learning used
5	Sonar, P. & JayaMalini, K. (2019)	2019	ICCMC, IEEE	Tests different ML approaches	Small-scale experiment
6	Febrian, M. E. et al. (2023)	2023	Procedia Computer Science	Uses supervised ML methods	Limited real-world validation
7	Yahyaoui, A. et al. (2019)	2019	UBMYK Conference	Combines ML and deep learning	More complex model
8	Jaiswal, V. et al. (2021)	2021	Primary Care Diabetes	Review of latest advances	No experimental model
9	Refat, M. A. R. et al. (2021)	2021	ISPCC, IEEE	ML vs DL comparison	Requires high computational resources
10	Modak, S. K. S. & Jha, V. K. (2024)	2024	Multimedia Tools and Applications	Improved prediction model	Complexity in implementation
11	Alam, M. A. et al. (2024)	2024	Journal of Next-Gen Engineering Systems	Comprehensive review	No practical implementation
12	El-Bashbishy, A. E. S. & El-Bakry, H. M. (2024)	2024	Scientific Reports	Focus on pediatric diabetes	Specific to children only
13	Jichkar, M. et al. (2024)	2024	SILCON 2024, IEEE	Recent ML-based approach	Conference paper, less detail
14	El-Sofany, H. et al. (2024)	2024	International Journal of Intelligent Systems	Mobile app integration	App usability concerns
15	Narasimharao, M. et al. (2025)	2025	Archives of Computational Methods in Engineering	Future trends and survey	No practical implementation
16	Sudha, K. et al. (2024)	2024	ICAECT, IEEE	Comparative study for management	Limited dataset

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

17	Abdollahi, J. & Aref, S. (2024)	2024	SN Computer Science	Feature selection improves accuracy	Depends on selected features
18	Tripathy, N. et al. (2024)	2024	ASSIC, IEEE	ML and DL comparative analysis	High training time
19	Shojaee-Mend, H. et al. (2024)	2024	Healthcare Informatics Research	Cross-sectional study with data mining	Limited sample size
20	Al Reshan, M. S. et al. (2024)	2024	IEEE Access	Ensemble deep learning decision support	Very complex and costly

V. FUTURE SCOPE

Future research in diabetes prediction should focus on improving model efficiency and accessibility. The use of lightweight machine learning models can help in real-time prediction with low computational cost. Integration with mobile applications and wearable devices can make diabetes prediction more accessible to patients.

Advanced technologies such as deep learning, hybrid models, and explainable AI can further enhance prediction accuracy and transparency. Moreover, the use of large and diverse datasets can improve model generalization and reliability. Developing cost-effective and user-friendly systems will play a key role in the widespread adoption of these technologies in healthcare.

VI. CONCLUSION

This survey paper reviewed various machine learning and deep learning techniques used for diabetes prediction. It is observed that these techniques significantly improve prediction accuracy and support early diagnosis. Ensemble and deep learning models show promising results, but they also increase system complexity.

Despite the advancements, challenges such as data dependency, computational cost, and implementation issues remain. Overall, machine learning-based diabetes prediction systems have great potential in modern healthcare. With further improvements in scalability, efficiency, and usability, these systems can contribute significantly to early diagnosis and better disease management.

REFERENCES

[1]. Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. Procedia computer science, 165, 292-299.

[2]. Rani, K. J. (2020). Diabetes prediction using machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 6(4), 294-305.

[3]. Hasan, M. K., Alam, M. A., Das, D., Hossain, E., & Hasan, M. (2020). Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access, 8, 76516-76531.

[4]. Singh, K., Rout, J. K., & Das, H. (2019, July). Diabetes prediction using machine learning techniques. In 2019 International Conference on Intelligent Computing and Remote Sensing (ICICRS) (pp. 1-6). IEEE.

[5]. Sonar, P., & JayaMalini, K. (2019, March). Diabetes prediction using different machine learning approaches. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 367-371). IEEE.

[6]. Febrian, M. E., Ferdinan, F. X., Sendani, G. P., Suryanigrum, K. M., & Yunanda, R. (2023). Diabetes prediction using supervised machine learning. Procedia Computer Science, 216, 21-30.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

[7]. Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. (2019, November). A decision support system for diabetes prediction using machine learning and deep learning techniques. In 2019 1st International informatics and software engineering conference (UBMYK) (pp. 1-4). IEEE.

[8]. Jaiswal, V., Negi, A., & Pal, T. (2021). A review on current advances in machine learning based diabetes prediction. Primary Care Diabetes, 15(3), 435-443.

[9]. Refat, M. A. R., Al Amin, M., Kaushal, C., Yeasmin, M. N., & Islam, M. K. (2021, October). A comparative analysis of early stage diabetes prediction using machine learning and deep learning approach. In 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) (pp. 654-659). IEEE.

[10]. Modak, S. K. S., & Jha, V. K. (2024). Diabetes prediction model using machine learning techniques. Multimedia Tools and Applications, 83(13), 38523-38549.

[11]. Alam, M. A., Sohel, A., Hasan, K. M., & Islam, M. A. (2024). Machine learning and artificial intelligence in diabetes prediction and management: A comprehensive review of models. Journal of Next-Gen Engineering Systems.

[12]. El-Bashbishy, A. E. S., & El-Bakry, H. M. (2024). Pediatric diabetes prediction using deep learning. Scientific reports, 14(1), 4206.

[13]. Jichkar, M., Shende, R., Bonde, O., Agrawal, P., Gupta, G. K., & Singh, A. K. (2024, November). Diabetes prediction using machine learning. In 2024 IEEE Silchar subsection conference (SILCON 2024) (pp. 1-6). IEEE.

[14]. El-Sofany, H., El-Seoud, S. A., Karam, O. H., Abd El-Latif, Y. M., & Taj-Eddin, I. A. (2024). A proposed technique using machine learning for the prediction of diabetes disease through a mobile app. International Journal of Intelligent Systems, 2024(1), 6688934.

[15]. Narasimharao, M., Swain, B., Priyadarshi, R., Nayak, P. P., & Bhuyan, S. (2025). A survey of machine learning techniques for diabetes prediction: current trends and future directions. Archives of Computational Methods in Engineering, 1-36.

[16]. Sudha, K., Lakshmipriya, C., Pajila, P. B., Venitha, E., & Anita, M. (2024, January). Enhancing Diabetes Prediction and Management through Machine Learning: A Comparative Study. In 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (pp. 1-6). IEEE.

[17]. Abdollahi, J., & Aref, S. (2024). Early prediction of diabetes using feature selection and machine learning algorithms. SN Computer Science, 5(2), 217.

[18]. Tripathy, N., Moharana, B., Balabantaray, S. K., Nayak, S. K., Pati, A., & Panigrahi, A. (2024, January). A comparative analysis of diabetes prediction using machine learning and deep learning algorithms in healthcare. In 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-6). IEEE.

[19]. 19)Shojaee-Mend, H., Velayati, F., Tayefi, B., & Babaee, E. (2024). Prediction of diabetes using data mining and machine learning algorithms: A cross-sectional study. Healthcare Informatics Research, 30(1), 73-82.

[20]. Al Reshan, M. S., Amin, S., Zeb, M. A., Sulaiman, A., Alshahrani, H., Shaikh, A., & Elmagzoub, M. A. (2024). An innovative ensemble deep learning clinical decision support system for diabetes prediction. IEEE Access, 12, 106193-106210