Diabetes Prediction Based on Machine Learning Techniques: A Review

Abstract: Diabetes is a metabolic disorder in the human body indicated by high sugar levels in blood or hyperglycemia. Machine Learning is a branch of Artificial Intelligence that enables machines to make predictions with no human interference by evaluating and interpreting data and learning from them. Health industry has recently seen a rise in the development of machine learning models to predict various diseases from the massive amount of data that has been recognized. However, health care systems can operate in a more efficacious manner and obtain a vast array of benefits by fetching useful insights from this data. The key objective of this review study is to analyze and critique the functioning and performance of different machine learning algorithms developed to diagnose diabetes. This paper highlights the role of machine learning in the healthcare industry.

Keywords: Diabetes, Artificial Intelligence, Machine Learning, Healthcare

Introduction :

Diabetes mellitus is a complex metabolic disorder that has emerged as one of the 21st century's most pressing health challenges. Commonly referred to as diabetes, is a chronic metabolic disorder caused by high blood glucose levels (hyperglycemia) due to the body’s inability to produce or effectively use insulin. Insulin is a hormone produced by the 𝛃-cells of pancreas. It regulates blood sugar by facilitating the absorption of glucose into cells for energy production. When this process is impaired, glucose accumulates in the bloodstream, leading to various health complications that affect almost every organ system and significantly influencing quality of life. **[38]**From ancient Egyptian physicians who first documented its sweet-tasting symptoms to modern-day researchers racing to find innovative treatments, diabetes has commanded the attention of medical professionals for millenia. Today, it affects hundreds of millions globally, crossing geographical boundaries and socioeconomic divisions, making it not just a medical condition but a critical public health priority that demands our urgent attention and understanding.



*Fig. 1: Classification of diabetes types*

Types of Diabetes : A Comprehensive Overview

The landscape of diabetes is diverse, with each type presenting its own unique characteristics, challenges and management approaches :

1. Type 1 Diabetes (T1D)

T1D also known as insulin-dependent diabetes mellitus (IDDM), occurs due to the autoimmune damage of the 𝛃-cells. This leads to suppression or cessation of the body’s capacity to produce insulin thereby creating an absolute insulin deficiency. The onset typically occurs before the age of 30, with a peak in adolescence, though it can develop at any age. Patients require careful monitoring as they face a risk of diabetic ketoacidosis, a severe complication that occurs when the body, lacking insulin, begins breaking down fat too rapidly. Management involves multiple daily insulin injections or insulin pump therapy, coupled with regular blood glucose monitoring and carbohydrate counting.

1. Type 2 Diabetes (T2D)

T2D, also known as non-insulin-dependent diabetes mellitus (IDDM), accounts for approximately 90 % of all diabetes cases. It develops through a complex interplay of genetic susceptibility and environmental factors. The condition progresses through several stages, it first begins with insulin resistance, where cells become less responsive to the effect of insulin. Initially, the pancreas compensates by producing excess insulin, but gradually over time, this compensation fails, leading to declining insulin production. T2D poses risks such as obesity(particularly central adiposity), physical inactivity, poor diet, advancing age, and certain ethnicities. The progression can be influenced by interventions at various stages, from lifestyle modifications in pre-diabetes to combinations of oral medications, injectable incretin-based therapies, eventually insulin therapy in advanced stages.

1. Gestational Diabetes (GDM)

GDM emerges during pregnancy due to placental hormones that create insulin resistance, challenging the ability of the mother to maintain normal blood glucose levels. This condition affects 2-10% of pregnancies approximately and requires careful monitoring as it can have an impact on both maternal and fetal health. Complications may include macrosomia (excessive fetal growth), increased risk of cesarean delivery and neonatal hypoglycemia. Women with GDM face a 35-60% chance of developing type 2 diabetes within 10-20 years post-pregnancy, necessitating regular screening after delivery.

1. Other Specific Types

Other specific distinct forms of diabetes include

1. Monogenic Diabetes : Including Maturity Onset Diabetes of the Young (MODY), characterized by genetic defects in the functionality of the 𝛃-cells. At least 14 different types of monogenic diabetes have been identified, each with specific genetic mutations affecting insulin production or glucose regulation
2. Neonatal Diabetes Mellitus (NDM) : NDM is a rare form of diabetes diagnosed within six months of life, presenting in two primary forms : Transient Neonatal Diabetes Mellitus (TNDM) and Permanent Neonatal Diabetes Mellitus (PNDM). TNDM typically resolves within the first few months but may recur later, often resulting in genetic abnormalities.

PNDM on the other hand is a lifelong condition requiring ongoing treatment.

In recent years, Artificial intelligence (AI) and machine learning (ML) have revolutionized diabetes prediction by analyzing vast datasets, identifying patterns, and delivering highly accurate predictions beyond traditional methods. These ML models integrate diverse data sources, from basic demographics to detailed medical histories, enabling a more comprehensive and precise approach to detecting diabetes.

AI and ML play a transformative role in diabetes prediction by enabling early detection through the analysis of vast datasets such as BMI, blood glucose, skin thickness etc. These AI-ML models not only predict the likelihood of developing diabetes but also offer great help to health professionals in minimizing its risk and offering personalized treatments. From traditional methods to Deep learning models the prediction of diabetes has come a long way. Techniques like explainable AI have now become a crucial part in the medical field in order to understand which factors are most influential in these predictions.Deep Learning models now use medical images to help detect early signs of diabetic retinopathy. Despite the varying accuracy across different models and datasets, AI-ML models in diabetes prediction signify a promising future for improved real-time diagnosis and monitoring.

Literature review:

This literature review provides a comprehensive overview of the existing research and advancements in the field of AI and machine learning to predict diabetes, highlighting the key methodologies, datasets, features and results. The study transitions from reviewing and analyzing traditional methodologies to deep learning techniques developed by researchers and professionals to predict diabetes mellitus.

Three supervised machine learning algorithms—Support Vector Machine (SVM), Logistic Regression (LR), and Artificial Neural Network (ANN)—were used by the authors of **[1]** to predict diabetes. The algorithms used the following features: age, diabetes, skinfold thickness, BMI, blood pressure, insulin levels, and diabetes spectrum function.SVM was demonstrated to be particularly good for binary classification tasks because of its capability to produce optimal hyperplanes, but LR produced easy-to-understand outcomes for binary result prediction. Because neural networks follow the patterns of the brain in forming networks, combining all these patterns proved to increase accuracy.

The study **[7]** used Decision Tree, K-nearest neighbor(KNN), Naive Bayes, and Random Forest on the PIMA dataset after a series of thorough preprocessing that included handling incomplete data and standardization. RF showed the best results with 86% accuracy and was found to be very well performing with noise and missing data. This study also proposed a cross of machine-learning models with real-time data collection from IoT sensors for enhanced healthcare applications.

To address the use of an ensemble model, the paper **[10]** highlighted the exceptional performance of Random Forest in diabetes prediction, emphasizing its ability to handle complex healthcare data. Ensemble models, like RF, integrate numerous decision trees to enhance prediction accuracy and also reduce overfitting risk. The authors performed intensive preprocessing, including normalization, handling class imbalance, and feature importance evaluation. They used measures such as precision, recall, F1-score, and accuracy to evaluate the performance of the model. This study highlighted the effectiveness of ensemble models, particularly random forests, in addressing the complexities of diabetes prediction and improving the predictive power of ML models in medical diagnostics.

The authors of the study **[12]** used three supervised machine learning algorithms on the PIMA India diabetes dataset: logistic regression, random forest (RF), and decision tree (DT). This study highlighted the significance of early diagnosis for effective diabetes control, with logistic regression having the highest accuracy of 76%, followed by RF with 75%. The results showed the best performance of logistic regression in handling the binary distribution function, while also validating the robustness of random forests to noise and their ability to represent relationships.

The authors of **[6]** indicated a move toward the use of deep learning models in the prediction of diabetes. This study used continuous oscillation deep neural networks to reduce overfitting and improve the prediction of blood glucose outcomes.

The study **[8]** showed significant results obtained by comparing five groups, namely Naïve Bayes, random forest, logistic regression, neural network, and SVM, using the PIMA dataset. Logistic regression showed the best performance with 77.2% accuracy, which was effective in classifying binary tasks, and showed that preliminary techniques such as process control hold utter importance.

The research paper **[15]** analyzed the use of machine learning algorithms for predicting diabetes by focusing on improving accuracy through data pre-processing and the use of SVM, RFC, and DNN algorithms. The study used data from the National Institute of Diabetes and Digestive and Kidney Diseases. The authors pre-processed the data with dummy variables and PCA, and achieved the highest accuracy of 89% with DNN, exceeding SVM and RFC. Despite emphasizing the significance of data pre-processing, this study identified various gaps, such as the limited exploration of ensemble techniques and relies on a single dataset, which may not fully represent the diverse population varying in demographics, lifestyle factors, and healthcare access to individuals.

The researchers of **[22]** examined various ML approaches applied to diabetic datasets to aid in the early diagnosis and management of Diabetes Mellitus. The paper highlighted the use of ML techniques with Big Data Analytics tools such as Hadoop and MapReduce, as well as classifiers like Naïve Bayes, Decision Trees, SVM, KNN, Random Forest, and Gradient Boosting.

The authors of the study **[18]** demonstrated that due to its ability to handle nonlinear interactions, XGBoost performed better overall. To provide personalized medical advice, the authors proposed creating an AI system that combines cognitive processing with a user interface.

Research by **[24]** demonstrated the use of various ML models for early diabetes diagnosis, including KNN, SVM, Gradient Boosting, Naïve Bayes, and LR. KNN was found to be the best-performing algorithm, with 75% accuracy, and had proven its usefulness when used with Flask for real-time prediction. This study addressed issues such as data quality and sampling bias while indicating the potential of AI to transform healthcare by providing insights into diabetes risk.

The research paper **[25]** studied the application of ML classifiers for predicting diabetes mellitus. The study utilized five ML models—Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and K-nearest neighbors (KNN)—to evaluate their effectiveness in diabetes prognosis. The RF classifier achieved the highest accuracy of 92.23%, demonstrating its ability to model complex, nonlinear relationships, while LR showed a lower accuracy of 74.42%. This study highlighted the potential of ML to enable early diagnosis through intensive data preprocessing, which included handling missing values, feature scaling, and correlation adjustments. The evaluation framework included accuracy, precision, recall, and F1-score measures which identified RF as the most effective model.

Various machine learning models have recently seen a rise in the prediction of diabetes in patients. One such study **[26]** focused on using genetic algorithm-based feature selection and classification methods to predict diabetes. This study addressed class imbalance using the ADASYN technique and demonstrated significant accuracy improvements with GA-based feature selection. The authors used two diabetic datasets in which the accuracy increased from 84.5% to 90.3% in Diabetic Dataset-1 (DD-1) and from 94.5% to 97.6% in Diabetic Dataset-2 (DD-2).

References

[1] Joshi, T. N., & Chawan, P. M. (2018). Diabetes Prediction Using Machine Learning Techniques. *Journal of Engineering Research and Application Www.Ijera.Com*, *8*, 2248–9622. https://doi.org/10.9790/9622-0801020913

[2] Prabhu, P., & Selvabharathi, S. (2019). Deep belief neural network model for prediction of diabetes mellitus. *2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC)*, 138–142. https://doi.org/10.1109/ICISPC.2019.8935838

[3] Faruque, M. F., Asaduzzaman, & Sarker, I. H. (2019). Performance analysis of machine learning techniques to predict diabetes mellitus. *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1-4. https://doi.org/10.1109/ECACE.2019.8679365

[4] Vijayan, V. V., & Anjali, C. (2016). Prediction and diagnosis of diabetes mellitus - A machine learning approach. *2015 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2015*, 122–127. https://doi.org/10.1109/RAICS.2015.7488400

[5] Lyngdoh, A. C., Choudhury, N. A., & Moulik, S. (2021). Diabetes Disease Prediction Using Machine Learning Algorithms. *Proceedings - 2020 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2020*, 517–521. https://doi.org/10.1109/IECBES48179.2021.9398759

[6] Bhargava, R., & Dinesh, J. (2021). Deep Learning based System Design for Diabetes Prediction. *2021 International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON 2021*. https://doi.org/10.1109/SMARTGENCON51891.2021.9645906

[7] Zaman, S. M. T., Paul, S. K., Paul, R. R., & Hamid, M. E. (2021). Detecting diabetes in human body using different machine learning techniques. *2021 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, 1–6. https://doi.org/10.1109/IC4ME253898.2021.9768501

[8] Saxena, R., Sharma, S. K., & Gupta, M. (2021). Analysis of machine learning algorithms in diabetes mellitus prediction. *Journal of Physics: Conference Series*, *1921*(1). https://doi.org/10.1088/1742-6596/1921/1/012073

[9] Pal, M., Parija, S., & Panda, G. (2021, August 5). Improved prediction of diabetes mellitus using machine learning based approach. *2nd International Conference on Range Technology, ICORT 2021*. https://doi.org/10.1109/ICORT52730.2021.9581774

[10] Xu, X., Huang, X., Ma, J., & Luo, X. (2021). Prediction of Diabetes with its Symptoms Based on Machine Learning. *2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering, CSAIEE 2021*, 147–156. https://doi.org/10.1109/CSAIEE54046.2021.9543343

[11] Samet, S., Laouar, M. R., & Bendib, I. (2021). Diabetes mellitus early stage risk prediction using machine learning algorithms. *5th International Conference on Networking and Advanced Systems, ICNAS 2021*. https://doi.org/10.1109/ICNAS53565.2021.9628955

[12] Buhari, B., & Adewara, A. A. (2022). *On the analysis of some machine learning algorithms for the prediction of diabetes*. *International Journal of Advanced Computer Science and Applications*, *13*(8), 123-130. https://doi.org/10.14569/IJACSA.2022.0130816

[13] Mangal, A., & Jain, V. (2022). Performance analysis of machine learning models for prediction of diabetes. *Proceedings - 2022 2nd International Conference on Innovative Sustainable Computational Technologies, CISCT 2022*. https://doi.org/10.1109/CISCT55310.2022.10046630

[14] Thabit, Q. Q., Fahad, T. O., & Dawood, A. I. (2022). Detecting Diabetes Using Machine Learning Algorithms. *2022 Iraqi International Conference on Communication and Information Technologies, IICCIT 2022*, 131–136. https://doi.org/10.1109/IICCIT55816.2022.10010408

[15] Awoniran, O. M., Oyelami, M. O., Ikono, R. N., Famutimi, R. F., & Famutimi, T. I. (2022). A Machine Learning Technique for Detection of Diabetes Mellitus. *Proceedings of the 5th International Conference on Information Technology for Education and Development: Changing the Narratives Through Building a Secure Society with Disruptive Technologies, ITED 2022*. https://doi.org/10.1109/ITED56637.2022.10051439

[16] Kuriakose, S. M., Basa Pati, P., & Singh, T. (2022). Prediction of Diabetes Using Machine Learning: Analysis of 70,000 Clinical Database Patient Record. *2022 13th International Conference on Computing Communication and Networking Technologies, ICCCNT 2022*. https://doi.org/10.1109/ICCCNT54827.2022.9984264

[17] Sivaraman, M., & Sumitha, J. (2022). Diabetes Prediction based on Supervised and Unsupervised Learning Techniques - A Review. *3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings*, 1292–1296. https://doi.org/10.1109/ICOSEC54921.2022.9952107

[18] Omoora, E. S., Altaweil, H. A., Nagem, T., & Bozed, K. A. (2023). Diabetes Mellitus Prediction Based on Machine Learning Techniques. *2023 IEEE 11th International Conference on Systems and Control, ICSC 2023*, 225–231. https://doi.org/10.1109/ICSC58660.2023.10449831

[19] Shampa, S. A., Islam, M. S., & Nesa, A. (2023). Machine Learning-based Diabetes Prediction: A Cross-Country Perspective. *2023 International Conference on Next-Generation Computing, IoT and Machine Learning, NCIM 2023*. https://doi.org/10.1109/NCIM59001.2023.10212596

[20] Sarkar, P., & Pawar, S. (2023). Machine Learning based Early Predication and Detection of Diabetes Mellitus. *International Conference on Artificial Intelligence for Innovations in Healthcare Industries, ICAIIHI 2023*. https://doi.org/10.1109/ICAIIHI57871.2023.10489259

[21] Vrindavanam, J., Haarika, R., MG, S., & Kumar, K. S. (2023). Diabetes prediction in teenagers using machine learning algorithms. *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 343-347. IEEE.

[22] Krishna Manaswini, T., Nayak, P., Harshitha, V. S., & Barlapudi, S. (2023). Predictions of Diabetic Mellitus using ML Techniques: A Systematic Overview. *International Conference on Sustainable Computing and Smart Systems, ICSCSS 2023 - Proceedings*, 43–47. https://doi.org/10.1109/ICSCSS57650.2023.10169244

[23] Parimala, G., Kayalvizhi, R., & Nithiya, S. (2023). Diabetes Prediction using Machine Learning. *2023 International Conference on Computer Communication and Informatics, ICCCI 2023*. https://doi.org/10.1109/ICCCI56745.2023.10128216

[24] Sehgal, D., Gautam, B., Kaur, I., Singh, A., Sharma, V., & Kumar, N. (n.d.). *AI-Driven Early Diabetes Prediction*. https://doi.org/10.1109/AIC.2024.7

[25] Kumar, A., Gill, A. S., Singh, J. P., & Ghosh, D. (2024). A Comprehensive and Comparative Examination of Machine Learning Techniques for Diabetes Mellitus Prediction. *2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024*. https://doi.org/10.1109/ICCCNT61001.2024.10725693

[26] Ravi Kiran, T. S., Srisaila, A., Shankar, G. S., Sowjanya, B., & Lakshmanarao, A. (2024). Machine Learning Approach for Diabetes Prediction using Genetic Algorithm based Feature selection. *2024 3rd International Conference for Innovation in Technology, INOCON 2024*. https://doi.org/10.1109/INOCON60754.2024.10511558

[27] Chandra, T. B., Reddy, A. S., Adarsh, A., Jabbar, M. A., & Jyothi, B. N. (2024). Diabetes Prediction Using Gaussian Naive Bayes and Artificial Neural Network. *International Conference on Distributed Computing and Optimization Techniques, ICDCOT 2024*. https://doi.org/10.1109/ICDCOT61034.2024.10516226

[28] Senthil, J., Akbar, S., Akiladevi, N., Praveena, S., Ravindar, K., & Banupriya, V. (2024). Early Prediction of Diabetes and its Risk Factors based on ARIMA-ELMAN ANN Network. *2nd International Conference on Intelligent Data Communication Technologies and Internet of Things, IDCIoT 2024*, 1376–1381. https://doi.org/10.1109/IDCIoT59759.2024.10468045

[29] Kaur, I., & Ali, A. (2024). An In-Depth Exploration of Machine Learning Algorithms and Performance Evaluation Approaches for Personalized Diabetes Prediction. *Proceedings - 2024 International Conference on Emerging Innovations and Advanced Computing, INNOCOMP 2024*, 532–538. https://doi.org/10.1109/INNOCOMP63224.2024.00093

[30] Aruna Devi, B., & Karthik, N. (2024). Explainable Artificial Intelligence for Prediction of Diabetes using Stacking Classifier. *Proceedings of CONECCT 2024 - 10th IEEE International Conference on Electronics, Computing and Communication Technologies*. https://doi.org/10.1109/CONECCT62155.2024.10677165

[31] Aouamria, S., Boughareb, D., Nemissi, M., Kouahla, Z., & Seridi, H. (2024). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING An Ensemble Deep Learning Model for Diabetes Disease Prediction. In *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE* (Vol. 2024, Issue 4). www.ijisae.org

[32] Katiyar, N., Thakur, H. K., & Ghatak, A. (2024). Recent advancements using machine learning & deep learning approaches for diabetes detection: a systematic review. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *9*. https://doi.org/10.1016/j.prime.2024.100661

[33] Singh Gill, K., Anand, V., Chauhan, R., & Pokhariya, H. S. (2024). Using Machine Learning-based SMOTE Analysis with the Light GBM Classification Method to Classify Diabetic Patients. *2024 3rd International Conference for Innovation in Technology, INOCON 2024*. https://doi.org/10.1109/INOCON60754.2024.10511914

[34] Dattangire, R., Pamulaparthyvenkata, S., Mandvikar, S., Balakrishnan, A., & Chintale, P. (2024). AI/ML-Based Diabetes Application using Hybrid Grey Wolf and Dipper Throated Optimization Algorithm. *International Conference on Intelligent Algorithms for Computational Intelligence Systems, IACIS 2024*. https://doi.org/10.1109/IACIS61494.2024.10721678

[35] Chandra Sekhar Reddy, L., Gottipalli, M., Sravanthi, P., Rajanikanth, J., Yalamarthi, G., & Gurrapu, N. (2024). Bridging Horizons in Diabetes Prediction: A Comparative Exploration of Machine Learning and Deep Learning Approaches in Pima Indian Women. *Proceedings - 2nd International Conference on Advancement in Computation and Computer Technologies, InCACCT 2024*, 386–391. https://doi.org/10.1109/InCACCT61598.2024.10550977

[36] Gilani, S. A. H., Syed, M. H., & Anjum, A. (2024). Effective Diabetes Prediction: Integrating Ensemble Learning with LIME for Robust Results. *2024 International Conference on Frontiers of Information Technology (FIT)*, 1–6. https://doi.org/10.1109/FIT63703.2024.10838461

[37] Bardia, V., & Sophiya, E. (2024). Diabetes Prediction Using Machine Learning Algorithm: A Comparative Analysis. *10th International Conference on Advanced Computing and Communication Systems, ICACCS 2024*, 1973–1979. https://doi.org/10.1109/ICACCS60874.2024.10717264

[38] Rubin, G., & M King, K. (2013). A history of diabetes: from antiquity to discovering insulin. British *Journal of Nursing, www.magonlinelibrary.com, 12.* https://doi.org/10.12968/bjon.2003.12.18.11775