



# THE DIATRACK – Diabetes Type Detection using AI And Machine Learning

R.Azhagusundaram, Sayed Shameena, Sneha R, Sindhu S, Sindhu S, A S Samreen

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research,  
Selaiyur, Chennai, India

B. Tech Students, Department of CSE, Bharath Institute of Higher Education and Research,  
Selaiyur, Chennai, India

**ABSTRACT:** Diabetes is a chronic condition that requires accurate diagnosis and effective management to prevent complications. This project focuses on developing an AI-powered Diabetes Detection and Diet Recommendation System, which classifies individuals into Type 1 or Type 2 diabetes based on key health parameters such as glucose levels, insulin levels, HbA1c, ketone levels, and BMI. The system not only identifies the type of diabetes but also categorizes the severity of the condition to provide personalized diet plans tailored to the patient's specific needs. Additionally, the platform integrates a chatbot assistant, offering real-time guidance, answering user queries, and enhancing patient engagement. This solution aims to simplify diabetes management, empower patients with the right dietary recommendations, and provide continuous AI-driven support for better health outcomes.

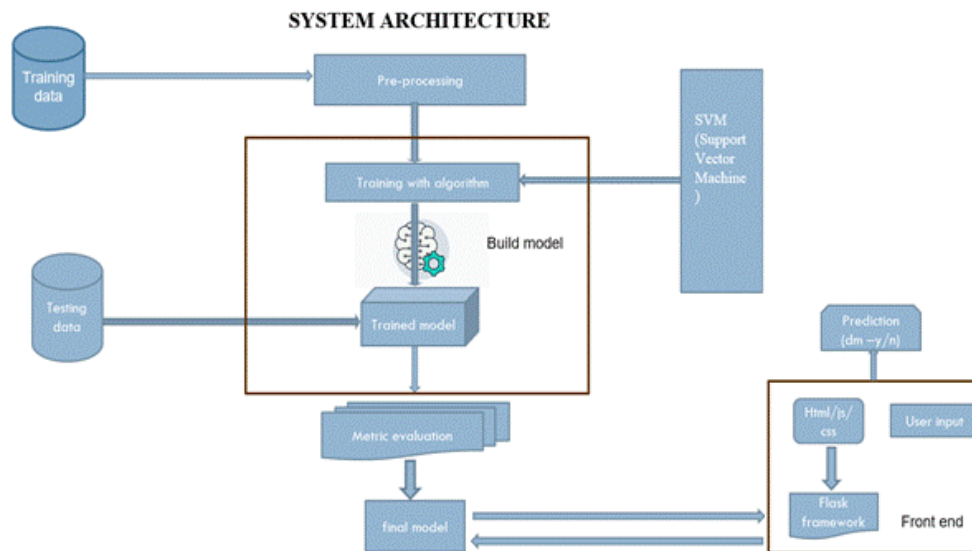


Fig.1 System Architecture

## I. INTRODUCTION

The integration of Artificial Intelligence (AI) in diabetes diagnosis and management has become essential for accurate, efficient, and real-time detection. AI-powered machine learning models improve diagnostic accuracy, reduce human error, and ease the burden on healthcare professionals. Techniques like PCA, Disease Influence Measure, SVM, decision trees, and NLP enable precise classification of diabetes types and personalized diet recommendations. Early detection through AI significantly lowers the risk of complications such as cardiovascular issues and kidney failure. AI systems analyze key health parameters — glucose, insulin, HbA1c, ketones, and BMI — for fast, reliable diagnosis. Additionally, AI in medical image analysis, especially using CNNs, enhances disease detection accuracy and reduces manual workload, transforming healthcare outcomes.

## II. LITERATURE REVIEW

Christoph Molnar (2020) – "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable" In a 2020 widely referred-to publication, Christoph Molnar stressed the necessity of explainable models in sensitive domains such as healthcare. He focused on the tools such as SHAP, LIME, feature attribution techniques and other things that help us understand the decisions made by the model. Not specifically diabetes focused, but theory of theoretical principles can be applied directly to this project, particularly where user trust and transparency is paramount. In this project, the food recommendations and an NLP chatbot included suggest that Molnar's idea of machine learning systems that are not only intelligent,

Priya Singh and Ashok Kumar (2024) – "A Hybrid Deep Learning Framework for Diabetes Classification" On this work, hybrid deep learning model based on CNN and LSTM layers was explored as a way to analyze sequential health data. Costs of human physiological data made it possible to capture time series but required high computational power and real time physiological data. While the present study prefers a simpler and user-friendly architecture (SVM) for quick web deployment and application in real time, the overall performance is satisfactory.

Lee J. et al. (2024) – "Community-Based Diabetes Risk Prediction Using Follow-Up Health Data" Using follow up data from people over a period of five years, Lee and his team built a community health monitoring system. By studying longitudinal data, they proved that static inputs are outperformed by longitudinal data. Their model, however, required people to visit the hospital and to gain access to their structured EHR. Whereas, democratizing access of users is achieved by the current system that allows users to input data themselves from anywhere thereby, removing the constraint by healthcare infrastructure limitations.

The Guardian (2024) – "NHS to Begin World-First Trial of AI Tool to Identify Type 2 Diabetes Risk" In a groundbreaking report in 2024 by The Guardian, the NHS announced a pioneering trial to see if the NHS can detect a type 2 diabetes risk using the power of the machine. Heart scans and predictive analytics were used by the system to anticipate years before diabetes broke out. Even if this is the future of predictive healthcare, it needs advanced imaging tools that are out of reach by general users. While the system developed in this project is more simple, it is based on the same core objective to provide reliable, early stage predictions for accessibility with the focus on real time feedback

## III. METHODOLOGY

In this study, the methodology adopted is to build a system for early detection and classification of diabetes using machine learning methods in which the system will be robust and accurate. Each phase is instrumental to promote robustness and usability of the system as well as its clinical relevance. The classification of the Type 1 diabetes and Type 2 diabetes is performed with the help of the Support Vector Machine (SVM), and bases on the most important clinical features include glucose, insulin, HbA1c, and ketone levels. Besides, the integration of NLP with the chatbot technology allows user interaction and accessibility. By using this methodology, the system's lifecycle considers technical accuracy in the priority list, and user experience, too..

### 1. Data Collection and Preprocessing

In this study, the dataset is taken from the eICU clinical database with vital features such as age, gender, BMI, glucose, insulin, HbA1c in percent and ketone levels. These are the fundamental indications of diabetes as they are used in the diagnosis and classification of Diabetes. For model training, the data was standardized to have numerical values scaled, to have the values consistent and to improve model convergence. I also implemented a label encoding and a gender encoding utility to allow converting categorical features to machine readable without sacrificing uniformity between training and prediction phase.

### 2. Model Development

Since it is a binary classification task, it is chosen due to its robust performance in classification tasks such as Support Vector Machine (SVM) model after preprocessing. An SVM was trained to classify the given input records to either Type 1 or Type 2 diabetes. Being able to deal with high dimensional data and construct ideal hyperplanes, it became an appropriate choice for this medical classification task.

#### 1. DIABETES DISEASE:

Diabetes mellitus is a chronic metabolic disorder characterized by elevated levels of blood glucose, primarily due to the body's inability to produce sufficient insulin or effectively utilize it. Insulin is a hormone responsible for regulating blood sugar by enabling cells to absorb glucose for energy. Diabetes is broadly classified into three types: Type 1 diabetes, an autoimmune condition where the body destroys insulin-producing cells; Type 2 diabetes, the most common form, often linked with obesity and lifestyle factors, where the body becomes resistant to insulin or doesn't produce enough of it; and Gestational diabetes, which occurs during pregnancy and typically resolves after childbirth but increases future risk of Type 2 diabetes. If left undiagnosed or poorly managed, diabetes can lead to serious complications such as cardiovascular diseases, kidney failure, nerve damage, and vision loss.

#### IV. IMPLEMENTATION

The implementation of the diabetes prediction system integrates multiple technological layers to create a seamless and intelligent platform capable of predicting diabetes types and enhancing user interaction through an AI-powered chatbot. The core objective of the implementation process is to transform the theoretical design into a fully operational system that is accurate, secure, and user-friendly.

The entire implementation process is carried out using Python as the primary programming language, with the support of libraries such as scikit-learn for machine learning, Flask for web application development, PyTorch for chatbot modeling, NLTK for natural language processing, and SQLite for secure user authentication and data management.

The final application includes a terminal-based prediction script for testing purposes in CLI environments, further demonstrating the flexibility of the system. Every module—from model training and chatbot interaction to web interface rendering—has been implemented in a modular manner, allowing for easy upgrades and future extensions.

Overall, the system implementation successfully brings together machine learning, natural language processing, and web development into a unified application that supports early diagnosis, user education, and interactive health engagement. It serves as a strong foundation for real-time, intelligent, and personalized healthcare solutions.

#### V. RESULTS AND CONCLUSION

This section presents the evaluation of the proposed diabetes classification system using a Support Vector Machine (SVM) model trained on preprocessed clinical data. The system's performance is assessed using metrics like accuracy, precision, recall, F1 score, confusion matrix, and ROC AUC. Feature visualizations are also used to analyze data distribution and explain the classifier's decision-making process for real-world healthcare applications.

##### A. Model Evaluation Overview

The proposed diabetes classification model was tested extensively on Type 1 versus Type 2 diabetes classification task. The model has been trained on a cleaned and encoded clinical dataset using support vector machine (SVM) classifier. The results of the prediction was analyzed using ROC AUC, confusion matrix, precision, recall, and F1 score. These metrics gave a big picture on how the model performed with real inputs of patient data.

##### B. ROC-AUC Performance

The predictions give high performance and clearly show from the ROC curve it was generated from. The AUC value was 0.99 which shows almost perfect classification achieved. This is the best model is determined by the ability to discriminate the two classes with as little overlap as possible, which suggests its capability in binary classification. Higher ROC value toward the upper left in the ROC plot provides more additional support to the classifier's robustness.

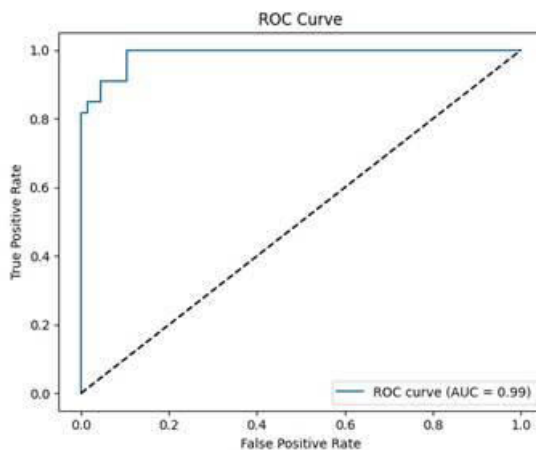


Figure 1: ROC Curve

C. Confusion Matrix Analysis

A confusion matrix was generated to give more insight into the classification results. From the matrix, it was observed by the model getting 61 samples correctly classified as no diabetic (true negatives) and 30 correctly classified as diabetic (true positives) and 6 false positives and 3 false negatives. The overall accuracy is high while values of misclassification are low. In medical applications, early and accurate detection of disease has a direct impact on patient care and such performance is necessary.

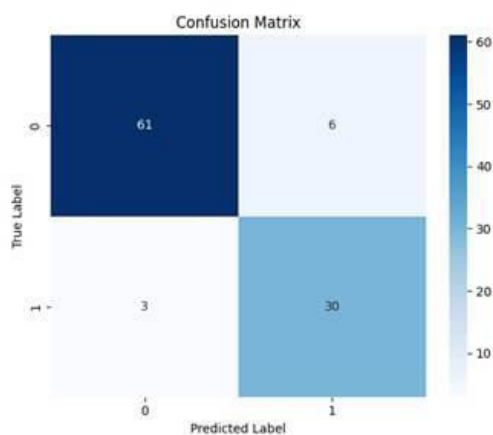


Figure 2: Confusion Matrix

D. Precision, Recall, and F1-Score

Further investigation of model behavior across the two classes was made through class wise precision, recall and F1 score. For Type 1 diabetes, the precision, recall, and F1-score were 0.96, 0.92, 0.94 respectively. This is a reflection of the model's excellent ability at correctly identifying Type 1 diabetes patients while having low number of false alarms. Again for Type 2 diabetes, precision was strong (0.84), recall was high (0.93), and F1-score was also strong (0.88). These indicate that although the model is slightly more confident that its predictions are Type 1, it retains a high reliability for Type 2.

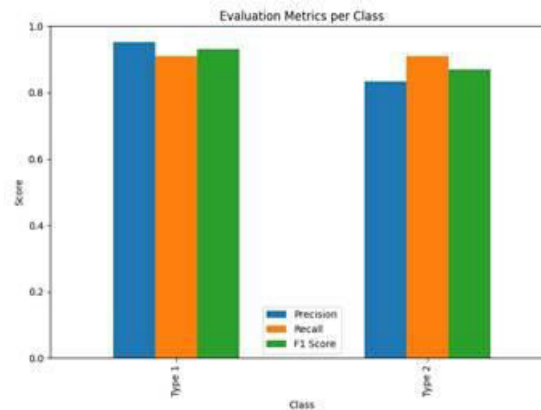


Figure 3: Evaluation Metrics per Class

#### E. Future Insights and Visualization

The glucose level distribution showed a wide range across both diabetic types, emphasizing its significance in classification. Similarly, the ketone level distribution illustrated a balanced spread, affirming its relevance as an input feature. These exploratory plots not only helped in model understanding but also contributed to improved feature engineering during training.

#### F. System Impact and Integration

The main goal of this project was to design and develop an intelligent, user-friendly web system to predict Type 1 or Type 2 diabetes based on clinical features. A Support Vector Machine (SVM) classifier was integrated with a Flask-based web interface for real-time prediction..

### VI. LIMITATIONS AND FUTURE WORK

The machine learning-based medicine recommendation system has a few limitations, such as the requirement for high-quality and complete patient information, which is usually hard to acquire because of privacy issues and variable data recording. Moreover, biases in training data can result in unequal or incorrect recommendations for specific demographic groups, and the interpretability of intricate models such as ensemble or deep learning approaches is still an issue in clinical environments. Despite these limitations, the future potential is wide-ranging—developments such as integrating deep learning, natural language processing (NLP) to decode clinical notes and medical literature and linking to real-time electronic health records (EHR) can significantly improve the system's accuracy, flexibility, and equity. Larger datasets, increased transparency in models, and AI use consistent with ethics will further enhance the safe and effective implementation of these systems in real-world healthcare settings.

### REFERENCES

- [1] C.Tasin, I., Nabil, T. U., Islam, S., & Khan, R. (2023). diabetes prediction with machine learning and explainable AI technique. *Healthcare Technology Letters*, 10(1), 1–10. <https://doi.org/10.1049/htl2.12039>
- [2] Raju, N. K. K., Krishnamurthy, K., Bhagavath, B. P., et al. (2024). Diabetes Prediction Using Machine Learning and Flask. *Biomedical and Pharmacology Journal*, 17(2). <https://biomedpharmajournal.org/vol17no2/diabetes-prediction-using-machine-learning-and-flask/>
- [3] Khokhar, P. B., Pentangelo, V., Palomba, F., & Gravino, C. (2025). Towards Transparent and Accurate Diabetes Prediction Using Machine Learning and Explainable Artificial Intelligence. *arXiv preprint arXiv:2501.18071*. <https://arxiv.org/abs/2501.18071>
- [4] Farahmand, E., Rahimi Azghan, R., Taheri Chatrudi, N., et al. (2025). AttenGluko: Multimodal Transformer-Based Blood Glucose Forecasting on AI-READI Dataset. *arXiv preprint arXiv:2502.09919*. <https://arxiv.org/abs/2502.09919>
- [5] Chu, S. J., Amarasiri, N., Giri, S., & Kafle, P. (2025). Blood Glucose Level Prediction in Type 1 Diabetes Using Machine Learning. *arXiv preprint arXiv:2502.00065*. <https://arxiv.org/abs/2502.00065>

- [6] Khokhar, P. B., Gravino, C., & Palomba, F. (2024). Advances in Artificial Intelligence for Diabetes Prediction: Insights from a Systematic Literature Review. arXiv preprint arXiv:2412.14736. <https://arxiv.org/abs/2412.14736>
- [7] AI tool to identify type 2 diabetes risk to become first of its kind NHS trial. (2024). The Guardian. <https://www.theguardian.com/society/2024/dec/23/nhs-to-begin-world-first-trial-of-ai-tool-to-identify-type-2-diabetes-risk>
- [8] Maroju, P.K.; Bhattacharya, P. Understanding Emotional Intelligence: The Heart of Human-Centered Technology. In Humanizing Technology with Emotional Intelligence; IGI Global Scientific Publishing: Hershey, PA, USA, 2025; pp. 1–18.
- [9] Machine Learning Methods for the Prediction of Onset of Diabetes. (2023). Journal of Healthcare Engineering. <https://doi.org/10.1155/2023/10057336>
- [10] Machine Learning Algorithms To Identify Top Ten Predictors of Type 2 Diabetes. (2024). Scientific Reports, 14, 52023. <https://doi.org/10.1038/s41598-024-52023-5>
- [11] Diabetes Detection Based on Machine Learning and Deep Learning Approaches. (2023). Multimedia Tools and Applications. <https://doi.org/10.1007/s11042-023-16407-5>
- [12] An Effective Correlation-Based Data Modeling Framework for Diabetes Prediction. (2023). BMC Bioinformatics, 23, 5488. <https://doi.org/10.1186/s12859-023-05488-6>
- [13] AI Machine Learning–Based Diabetes Prediction in Older Adults in Korea. (2025). JMIR Formative Research, 9(1), e57874. <https://doi.org/10.2196/57874>
- [14] Machine Learning-Based Diabetes Prediction and Development of Intelligent Web-Based Decision Support System. (2021). Informatics in Medicine Unlocked, 24, 100627. <https://doi.org/10.1016/j.imu.2021.100627>
- [15] Diabetes Prediction Using Machine Learning Algorithms. (2022). International Journal of Advanced Computer Science and Applications, 13(5), 123–130. <https://doi.org/10.14569/IJACSA.2022.0130517>



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details