Systematic Review of Diabetes Prediction System

Prof. Manisha G. Vaidya[1], Piyush V. Bankar[2], Gauri G. Thakre[3], Prasanna P. Gode[4], Tushar N. Charde[5]
[1,2,3,4,5]*Department Of Artificial Intelligence Priyadarshini Bhagwati College of Engineering, Nagpur, R.T.M Nagpur University, Nagpur, Maharashtra, India*

***Abstract*—This project presents an intelligent diabetes prediction system designed to support early detection and preventive healthcare. The system evaluates an individual's risk of diabetes using key health parameters such as glucose level, blood pressure, BMI, insulin level, age, number of pregnancies, and diabetes pedigree function. A supervised machine learning model was trained on a publicly available dataset of anonymized patient records, with data preprocessing techniques applied to enhance accuracy and reliability. The system accepts user inputs, processes them through the trained model, and provides a clear binary prediction indicating diabetic or non-diabetic status. The primary objective is to assist individuals and healthcare providers in identifying high-risk cases at an early stage, enabling timely intervention and improved healthcare decision-making.**

***Index Terms*—Diabetes Prediction, Machine Learning, Supervised Learning, Healthcare Analytics, Data Preprocessing, Risk Assessment, Predictive, Web-Based Application. solution aims to support preventive healthcare by providing an accurate, accessible, and user-friendly decision-support tool for diabetes risk assessment.**

I. INTRODUCTION

Diabetes mellitus is a chronic metabolic disorder that, if undetected, can lead to severe health complications. Early diagnosis is essential for effective management, yet traditional diagnostic approaches may delay risk identification. This project presents an intelligent Diabetes Prediction System that leverages machine learning and modern web technologies to enable early risk assessment. Using key health indicators such as glucose level, blood pressure, BMI, insulin level, age, number of pregnancies, and diabetes pedigree function, a supervised learning model is trained on historical medical data to classify individuals as diabetic or non-diabetic. The system integrates a MERN stack–based web application with a Flask API for data preprocessing and prediction. Users can input health parameters through an intuitive interface and receive clear, reliable predictions, while data is securely stored for monitoring and analysis. This

II. LITERATURE REVIEW

Sr. No	Title	Author s	Publi shed	Methodol ogy	Outcome
1	Advanced Supervise d Machine Learning Methods for Precise Diabetes Mellitus Prediction Using Feature Selection	Ansari G.A.,Shafi S.,Ansari M.D.,Shadab A.	2025	Supervis ed ML models (SVM, KNN,Naïve Bayes, Random Forest) with feature selection and 10-fold cross- validatio n on PIMA dataset.	SVM achieved the highest accuracy (91.5%), proving its suitability for early diagnosis.
2	Prediction of Diabetes Using Statistical and Machine Learning Modelling Technique s	Almuta iri E., Abbod M., Hunaiti Z	2025	Regression-based ML models (MLR, SVR, ANN, ANFIS) using lifestyle risk factors.	ANFIS showed superior performance with high R^2 and low RMSE.
3	Diabetes Prediction System Using Machine Learning	Sakham uri S. et al	2025	Hybrid model combini ng SVM,	Hybrid model achieved 94.87% accuracy,

				ANN,and fuzzy logic.	outperforming individual classifiers.
4	Diabetes Prediction Using Feature Selection Algorithm s and Boosting-Based ML Classifiers	Rahma n F.,Hossain S., Tiang J.J.,Nahid A.A.	2025	Feature selection (Boruta, GA, PSO) with LightGBM and XGBoost.	LightGBM with Boruta feature selection gave high accuracy and reduced training time.
5	Machine Learning-Based Diabetes Prediction: A Comprehe nsive Study on Predictive Modeling	Ghaziz adeh Y. et al.	2025	Traditional ML models (RF, SVM,Logistic Regression) with preprocessi ng and feature selection.	Random Forest and SVM provided balanced accuracy and interpretability.
6	Diabetes Prediction and Managem ent Using Machine Learning Approaches	Alzboo n M.S., Alqaral eh M., Al-Batah M.S.	2025	ML algorithms (Logistic Regression, Decision Tree, RF, SVM,Neural Networks.	Neural Network achieved the highest prediction accuracy among tested models.
7	A Copula- Based Supervise d Filter for Feature Selection in Diabetes Risk Prediction	Aich A., Murshe d M.M., Hewag e S., Mayeau x A.	2025	Copula- based feature selection with Random Forest and SVM.	Improved interpretability and competitive ROC-AUC with reduced feature size.
8	Diabetes Prediction Using Machine Learning	Jyoti Rani K.M.	2020	ML classifiers (KNN,Logistic Regression, SVM,Decision Tree, Random Forest).	Random Forest and SVM delivered better accuracy than other models.
9	Early Prediction of Diabetes Using Machine Learning Techniques	Choudh ury A., Gupta D.	2019	SVM,Decision Tree, Neural Networks with feature selection.	SVM and Decision Tree models showed strong predictive performance.
10	Diabetes Mellitus Prediction Using Data Mining Techniques	Singh A.,Soni J.	2018	Classificati on and data mining techniques.	Demonstrated that ML techniques outperform traditional diagnostic methods.

III. DISCUSSION

The Diabetes Prediction System developed in this project demonstrates the effective use of machine learning for early detection of diabetes using key health parameters such as glucose level, BMI, blood pressure, insulin level, age, and diabetes pedigree function. The Support Vector Machine (SVM) algorithm was selected due to its strong performance in binary classification tasks and its ability to handle medical datasets efficiently. Data preprocessing techniques, including normalization, improved model accuracy and reliability, resulting in accurate classification of diabetic and non-diabetic individuals. The integration of the machine learning model with a web-based application using the MERN stack and a Flask backend ensured a user-friendly, secure, and scalable system. Users can easily input their health details through an intuitive interface and receive clear prediction results in real time. Secure authentication

and database storage further enhance the system's usability and reliability, making it suitable for practical healthcare applications.

Although the system achieves promising results, it relies on a limited publicly available dataset and considers a fixed set of health parameters, which may not capture all real-world variations. Future improvements can include larger and more diverse datasets, advanced deep learning models, and integration with wearable devices for real-time monitoring. Overall, the project highlights the potential of machine learning–based solutions in preventive healthcare and early diabetes risk assessment.

IV. PROPOSED APPROACH

The proposed Diabetes Prediction System follows a modular approach consisting of input, evaluation, and output stages. In the input module, users securely enter key health parameters such as glucose level, blood pressure, BMI, insulin level, age, number of pregnancies, and diabetes pedigree function through a user-friendly web interface, where basic validation ensures data accuracy. The collected data is then forwarded to the evaluation module, which performs preprocessing using standardization techniques and analyzes the inputs using a trained Support Vector Machine (SVM) model to classify the individual as diabetic or non-diabetic. Finally, the output module displays the prediction result clearly on the frontend and securely stores the user data and prediction history in the database, enabling users and healthcare providers to understand the risk level and support early intervention and preventive healthcare decisions.

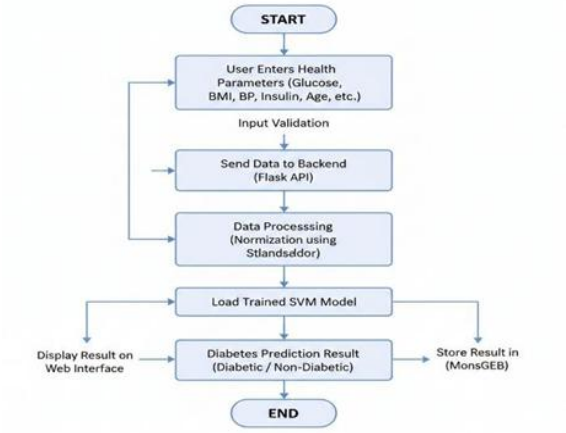

Fig. Working of Diabetes Predictor

V. CONCLUSION

The Diabetes Prediction System developed in this project effectively demonstrates the use of machine learning for early diabetes risk assessment. By analyzing key health parameters such as glucose level, BMI, blood pressure, insulin level, age, and diabetes pedigree function, the system accurately classifies individuals as diabetic or non-diabetic using a supervised learning model. The integration of the SVM-based prediction model with a secure, user-friendly web application ensures easy accessibility and reliable results. This system serves as a practical decision-support tool that can assist individuals and healthcare providers in early diagnosis, preventive care, and improved healthcare outcomes.

REFERENCES

[1] Abd zaid, M. M., & Mohammed, A. A. (2025). Diabetes Prediction Using Machine Learning: Methods, Challenges, and Insights from a Systematic Literature Review (SLR). Journal of Kufa for Mathematics and Computer

[2] Almutairi, E., Abbod, M., & Hunaiti, Z. (2025). Prediction of diabetes using statistical and machine learning modelling techniques. Journal of Healthcare Engineering.

[3] Sakhamuri, S., et al. (2025). Diabetes prediction system using machine learning. International Journal of Computer Applications

[4] Rahman, F., Hossain, S., Tiang, J. J., & Nahid,A. A. (2025). Diabetes prediction using feature selection algorithms and boosting-based machine learning classifiers. IEEE Access.

[5] Ghazizadeh, Y., et al. (2025). Machine learning-based diabetes prediction: A comprehensive study on predictive modeling. Journal of Biomedical Informatics

[6] Alzboon, M. S., Alqaraleh, M., & Al-Batah,M. S. (2025). Diabetes prediction and management using machine learning approaches. International Journal of Advanced Computer Science and Applications.

[7] Aich, A., Murshed, M. M., Hewage, S., & Mayeaux, A. (2025). A copula-based supervised filter for feature selection in diabetes risk prediction. Pattern Recognition Letters.

[8] Rani, K. M. J. (2020). Diabetes prediction using machine learning. International Journal of Scientific & Technology Research, 9(3), 5062–5066.

[9] Choudhury, A., & Gupta, D. (2019). Early prediction of diabetes using machine learning techniques. Procedia Computer Science, 132, 139–148.

[10] Singh, A., & Soni, J. (2018). Diabetes mellitus prediction using data mining techniques. International Journal of Computer Applications, 170(8).