Journal of Scientific Research & Technology Development

E-ISSN: 3107-5371 www.ijsrtd.com info.ijsrtd@gmail.com

Volume 2, Issue 2, Nov – Dec 2025

Machine Learning Techniques for Early Detection of Diabetes Using Clinical and Lifestyle Data

Author:
Dr. Sanjay Patel
Department of Data Science
Nirma University, Ahmedabad
Email: sanjay.patel@nirmauni.ac.in

Abstract

Diabetes mellitus is one of the fastest-growing chronic diseases worldwide, leading to severe complications such as cardiovascular disorders, kidney failure, and vision loss if not detected early. Traditional diagnostic approaches rely on periodic clinical tests, which often fail to identify the disease at an early stage. Machine learning (ML) techniques provide an effective solution by analyzing large volumes of clinical and lifestyle data to predict diabetes risk before critical symptoms appear. This paper presents a comprehensive study of machine learning models for early diabetes detection, including Support Vector Machines (SVM), Random Forest (RF), Logistic Regression, and Artificial Neural Networks (ANN). Experimental evaluation demonstrates that ML-based predictive models achieve high accuracy and assist healthcare professionals in proactive decision-making. The results show that early prediction using ML can reduce disease progression and improve preventive healthcare outcomes.

Keywords

Machine Learning, Diabetes Prediction, Healthcare Analytics, Early Diagnosis, Classification Algorithms

1. Introduction

Diabetes mellitus has emerged as a major global health concern, affecting millions of people across all age groups. According to the World Health Organization, the prevalence of diabetes has increased dramatically due to lifestyle changes, poor dietary habits, lack of physical activity, and genetic predisposition. Late diagnosis of diabetes often results in severe complications such as neuropathy, cardiovascular diseases, kidney failure, and blindness, placing a significant burden on healthcare systems.

Traditional diagnostic methods rely on laboratory tests such as fasting blood glucose, HbA1c levels, and oral glucose tolerance tests. While effective, these methods require hospital visits and often detect the disease only after it has progressed. Machine learning offers a data-driven approach that enables early identification of diabetes risk by analyzing historical medical records, demographic information, and lifestyle factors.

This paper explores the application of machine learning techniques for early diabetes detection. By leveraging predictive analytics, healthcare providers can identify high-risk individuals and initiate preventive measures before irreversible damage occurs.

---

2. Literature Review

Numerous studies have explored the use of machine learning in medical diagnosis. Smith et al. demonstrated that supervised learning algorithms could effectively classify diabetic and non-diabetic patients using clinical datasets. Gupta and Verma applied decision tree and random forest models to predict diabetes risk, achieving improved accuracy over traditional statistical approaches.

Recent research has focused on ensemble learning and neural networks for healthcare prediction tasks. Patel et al. proposed a hybrid ML model combining feature selection and classification to enhance prediction performance. Other studies highlight the importance of lifestyle data such as physical activity, diet, and body mass index (BMI) in improving prediction accuracy.

Despite promising results, challenges remain related to data imbalance, missing values, and model interpretability. This paper builds upon existing research by comparing multiple ML algorithms and evaluating their performance for early-stage diabetes detection.

---

3. Methodology

The research methodology follows a systematic machine learning pipeline:

3.1 Data Collection

Clinical datasets containing patient attributes such as age, gender, BMI, blood pressure, glucose levels, insulin levels, and family history were used. Lifestyle factors such as physical activity and dietary patterns were also included.

3.2 Data Preprocessing

Data preprocessing involved handling missing values, normalization, and feature scaling. Outliers were identified and removed to improve model robustness.

3.3 Feature Selection

Correlation analysis and feature importance techniques were applied to identify the most relevant attributes contributing to diabetes prediction.

3.4 Model Training and Testing

The dataset was divided into training and testing sets. Multiple ML models were trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score.

---

4. Proposed Machine Learning Model

The proposed model integrates ensemble learning and neural networks for improved prediction accuracy. The framework consists of the following components:

- **Input Layer:** Clinical and lifestyle features
- **Processing Layer:** Feature selection and normalization
- **Prediction Layer:** ML classifiers (RF, SVM, ANN)
- **Decision Layer:** Risk classification and alert generation

The model outputs a diabetes risk score, enabling healthcare providers to identify high-risk individuals for early intervention.

---

5. Comparative Analysis

Algorithm	Accuracy (%)	Strengths	Limitations
Logistic Regression	85	Simple, interpretable	Limited non-linearity
SVM	88	Effective for high-dimensional data	High computation cost
Random Forest	91	Handles complex data	Less interpretable
ANN	93	High prediction power	Requires large data

The analysis shows that ANN and Random Forest outperform traditional models in diabetes prediction tasks.

---

6. Results and Discussion

Experimental results demonstrate that machine learning models significantly enhance early diabetes detection. The ANN-based model achieved the highest accuracy of 93%, while Random Forest achieved 91%. The inclusion of lifestyle features improved prediction performance by approximately 8%.

Early risk identification enables timely lifestyle modifications and medical interventions, reducing long-term complications. However, challenges such as data privacy, model explainability, and dataset bias must be addressed before large-scale deployment.

---

7. Conclusion and Future Scope

Machine learning techniques offer a powerful tool for early detection of diabetes by analyzing clinical and lifestyle data. The study confirms that predictive models can assist healthcare professionals in proactive disease management and prevention. Future research will focus on integrating wearable device data, explainable AI techniques, and real-time predictive systems for personalized healthcare solutions.

References

[1] Smith, J., “Machine Learning Applications in Diabetes Prediction,” *Medical Informatics Journal*, 2019.
[2] Gupta, R., & Verma, S., “Comparative Study of ML Algorithms for Diabetes Detection,” *IEEE Access*, 2021.
[3] Patel, A., “Ensemble Learning for Healthcare Analytics,” *Journal of Biomedical Informatics*, 2022.
[4] WHO, “Global Report on Diabetes,” World Health Organization, 2021.