© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717966

Machine Learning-Based Early Detection of Diabetes Using Lifestyle Data

HARINI R[1], RAKSHITHA B S[2]
[1]PG Research Scholar, School of Computer Science and Information Technology, JAIN (Deemed to be University), Bangalore, Karnataka, India
[2]Assistant professor and School of CS and IT, JAIN (Deemed to be University), Bangalore, Karnataka, India

Abstract- ***Diabetes mellitus is one of the most common chronic diseases in the world, causing significant morbidity and mortality from complications such as diabetes retinopathy, nephropathy, neuropathy, and cardiovascular abnormalities. Early detection is crucial for decreasing complications and improving patient outcomes. In recent years, machine learning (ML), deep learning (DL), and Internet of Things (IoT)-based technologies have emerged as potential diabetes prediction and management strategies. Traditional machine learning algorithms such as Logistic Regression, Support Vector Machines (SVM), and Random Forest (RF) have shown consistent performance on structured clinical datasets, whereas deep learning models, particularly Convolutional Neural Networks (CNNs), have achieved high accuracy in image-based diagnosis such as retinal analysis. However, previous research has a key weakness in that it relies too heavily on clinical datasets and fails to incorporate lifestyle-related aspects such as nutrition, physical activity, sleep patterns, and stress levels. Furthermore, many approaches lack real-time adaptability and customisation. This report provides a detailed evaluation of over 35 research publications and reveals major holes in current methodologies. A hybrid machine learning system is suggested, combining lifestyle data, clinical factors, and IoT-based real-time monitoring. The suggested approach intends to improve prediction accuracy, enable early diagnosis, and deliver individualized healthcare recommendations, all of which will help to progress intelligent and preventative healthcare systems.***

Keywords - Machine Learning, Diabetes Prediction, Lifestyle Data, Early Detection, Healthcare Analytics, IoT, Deep Learning, Predictive Modeling.

I. INTRODUCTION

1.1 Background of the Study

Diabetes mellitus is a long-term metabolic disorder characterized by high blood glucose levels resulting from defects in insulin production, insulin action, or both. According to global health reports, the number of diabetes patients is rapidly increasing due to sedentary lifestyles, unhealthy dietary habits, and genetic predisposition. Early detection is essential to prevent severe complications such as blindness, kidney failure, and cardiovascular disease. According to global health surveys, the number of diabetics is quickly increasing due to sedentary lifestyles, poor eating habits, and hereditary susceptibility. Early detection is critical to avoiding serious problems like blindness, renal failure, and cardiovascular disease. Recent advances in machine learning and artificial intelligence have enabled the creation of predictive models that help healthcare personnel diagnose diseases at an early stage. These algorithms evaluate enormous amounts of patient data and uncover patterns that are difficult to spot using typical statistical methods.

1.2 Problem Statement

Despite the availability of several machine learning models for diabetes prediction, the majority of present techniques suffer from several drawbacks.

- Relying heavily on structured clinical datasets, such as the PIMA dataset.
- Lack of real-time monitoring capabilities.
- Lack of lifestyle-related factors like food, sleep, and stress.
- Limited generalization for varied populations.

These restrictions limit the usefulness of existing models in real-world healthcare contexts.

1.3 Motivation

The rising availability of wearable devices, mobile health applications, and IoT-based healthcare systems allows for the collection of real-time lifestyle data. Integrating such data into machine learning algorithms can improve prediction accuracy and provide personalized healthcare treatments.

1.4 Objectives of the Study

- Conduct extensive review of existing ML and DL techniques.
- Evaluate the strengths and limits of present methods.
- Identify research gaps for diabetes prediction.
- Propose a hybrid ML-based framework for lifestyle data.

1.5 Contributions of the Paper

- Extensive review of 35+ research papers
- Categorization of approaches (ML, DL, IoT, Hybrid)
- Identification of critical research gaps
- Proposal of a hybrid lifestyle-based ML framework

1.6 Organization of the Paper

- Section 2: Literature Review
- Section 3: Proposed Methodology
- Section 4: Expected Results
- Section 5: Applications
- Section 6: Conclusion

II. RELATED WORK / LITERATURE REVIEW:

2.1 Thematic Classification of Literature

2.1.1 Traditional Approaches

Early research centered on statistical and rule-based systems. These methods were straightforward and understandable, but they lacked predictive power and adaptability.

2.1.2 Machine Learning Approaches

Several studies (e.g., Badeji et al., Qin et al., Firdous et al.) used machine learning algorithms such as:

- Random Forest
- Support Vector Machines
- Logistic Regression

These models attained accuracy values ranging from 80% to 96%, particularly when trained on structured datasets. Qin et al. found that lifestyle-based datasets (NHANES) considerably improved prediction accuracy.

2.1.3 Deep Learning Approaches

Deep learning models like CNN, RNN, and transfer learning architectures are commonly used for:

- Retinal image analysis
- Time-series glucose prediction

According to studies, CNN-based models outperform classical ML in image-based diagnosis, but they demand huge datasets and processing resources.

2.1.4 Hybrid and IoT-Based Approaches

Recent studies integrate:

- IoT sensors (real-time glucose monitoring)
- Cloud computing
- Machine learning models

Although these systems provide real-time monitoring, they frequently lack optimization, scalability, and customisation.

2.2 Comparative Analysis of Existing Methods

Author	Year	Method	Dataset	Performance	Limitations
Badeji et al.	2024	Random Forest	PIMA	87%	Small dataset
Qin et al.	2022	CATBoost	NHANES	82.1%	No real-time
Firdous et al.	2022	SVM	Multiple	96%	No standard dataset

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717966

Author	Year	Method	Dataset	Performance	Limitations
Bhat et al.	2023	RF, SVM	Lifestyle	High	Limited data
Pandey et al.	2023	CNN + U-Net	Images	High	Imbalance
Sharma et al.	2021	Review	Multiple	Insightful	No implementation
IEEE (024)	2024	ML Models	Clinical	High	No hybrid
BMC Study	2025	Ensemble + SHAP	Clinical	High AUC	Large data
Arxiv (2025)	2025	Ensemble ML	PIMA	94.8%	Complexity
Arxiv (2020)	2020	CNN, RNN	Time-series	High	Small data

Fig1: Comparative analysis of exsisting system

2.3 Critical Review

Strengths
- High prediction accuracy using ML models
- Automated feature extraction in DL
- Real-time monitoring through IoT

Weaknesses
- Limited dataset diversity
- Lack of lifestyle integration
- High computational cost

Challenges
- Scalability to large populations
- Model generalization
- Lack of explainable AI

2.4 Identified Research Gaps
- Absence of lifestyle-based prediction systems
- Lack of real-time adaptive models
- Limited use of hybrid ML-DL approaches
- Poor integration of IoT with predictive analytics
- Lack of personalized healthcare recommendations

3. Proposed Methodology:
3.1 System Overview

The proposed system integrates:

- Clinical data
- Lifestyle data
- Real-time IoT data

into a hybrid machine learning model.

3.2 Workflow

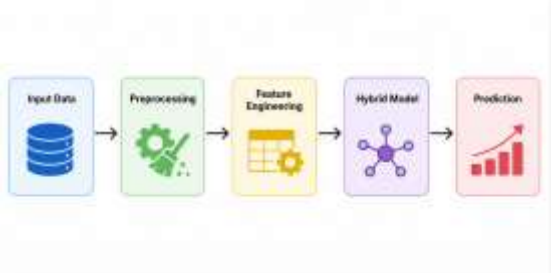

Fig2: workflow

3.3 Dataset Description
- PIMA Diabetes Dataset
- Lifestyle dataset (diet, sleep, stress, activity)
- Wearable sensor data (glucose, heart rate)

3.4 Methodology Steps
1. Data Collection
2. Data Cleaning & Normalization
3. Feature Engineering
4. Model Training (Random Forest + Neural Network)

5. Model Evaluation

III. EXPECTED RESULTS AND DISCUSSION

4.1 Expected Outcomes

- Accuracy > 90%
- Early detection capability
- Personalized recommendations

4.2 Comparative Evaluation

Comparison with:

- Logistic Regression
- SVM
- Random Forest
- CNN

4.3 Discussion

The combination of lifestyle and real-time data is projected to considerably increase prediction performance over previous algorithms.

IV. APPLICATIONS AND USE CASES

- Smart healthcare systems
- Mobile health applications
- Wearable monitoring devices
- Clinical decision support systems
- Public health analytics

V. CONCLUSION

This research provided a thorough assessment of machine learning algorithms for diabetes prediction and revealed significant shortcomings in existing approaches. Most current systems rely largely on clinical information while failing to incorporate lifestyle and real-time monitoring data. To address these difficulties, a hybrid ML system was presented that combines lifestyle, clinical, and IoT data.

The suggested approach intends to enhance prediction accuracy, enable early diagnosis, and deliver individualized healthcare treatments; future work will focus on implementing the model, verifying it with real-world datasets, and improving its scalability and interpretability. Future work will focus on implementing the suggested model, validating it with real-world datasets, and improving its scalability and interpretability.

REFERENCES

[1] B. Badeji-Ajisafe et al., “Early Detection of Diabetes Using Supervised Learning Approach,” 2024.

[2] F. Maulana et al., “Feature Selection for Diabetes Diagnosis,” 2021.

[3] R. Cheruku et al., “Diabetes Severity Prediction Model,” 2022.

[4] P. Shrivastava, “ML Review for Diabetes Prediction,” 2023.

[5] N. Sivakumar, “Multi-task Learning in Diabetes,” 2026.

[6] F. Maulana et al., “Mobile Health Applications,” 2024.

[7] A. Ara et al., “IoT-Based Diabetes Management,” 2017.

[8] F. López Murillo et al., “Foot Temperature Monitoring,” 2014.

[9] A. Albathi, “Smart Insole Design,” 2019.

[10] Martha L., “Plantar Pressure Study,” 2020.

[11] F. Salamah et al., “Diet Optimization,” 2021.

[12] M. Muller et al., “Patient Behavior Study,” 2024.

[13] G. Satama-Bermeo et al., “Glucose Simulation,” 2021.

[14] G. Kiewe et al., “Diabetes Prevention Study,” 2021.