Diabetes Prediction and Management Using Machine Learning

A Project Based Learning Report Submitted in partial fulfilment of the requirements for the award of the degree

of

Bachelor of Technology

in The Department of AI&DS

DATA MANAGEMENT AND WAREHOUSING 23AD2204R/E

Submitted by

2310080009: MEENAKSHI VARMA
2310080065: KOTA PHANI TEJA
2310080021: SAMITH REDDY
2310080052: DEVAANAND KUMAR
2310080027: SAI SUVVIDH KARRI

Under the guidance of

Dr.Purushottam Komaravalli

Department of Artificial Intelligence and Data Science Engineering

Koneru Lakshmaiah Education Foundation, Aziz Nagar

Aziz Nagar – 500075

MAR - 2025.

1. Introduction

Diabetes is a chronic metabolic disorder that affects millions of people worldwide. It is characterized by high blood glucose levels and can lead to severe complications such as cardiovascular disease, kidney failure, neuropathy, and retinopathy. Early detection and effective management of diabetes are crucial for improving patient outcomes and reducing healthcare burdens.

Traditional diagnostic methods for diabetes rely on clinical assessments and laboratory tests, which may be time-consuming and less effective in handling complex patient data. The emergence of machine learning (ML) techniques has provided new opportunities for improving diabetes prediction, risk assessment, and disease management. This review explores various studies that leverage ML models for diabetes classification and prediction, highlighting their methodologies, performance metrics, and potential applications.

2. Literature Review

2.1 A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques

- **Dataset:** Pima Indian Diabetes Database (PIDD) (768 instances, 9 attributes) [1].
- **Methods:** Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM).
- **Findings:** The proposed Intelligent Diabetes Mellitus Prediction Framework (IDMPF) achieved 83% accuracy. However, the study was retracted due to concerns over data integrity and peer-review manipulation [1].

2.2 Diabetes Prediction Using Machine Learning Techniques

- **Dataset:** PIDD (768 instances, 9 attributes) [2].
- **Methods:** SVM, k-Nearest Neighbors (k-NN), RF, Artificial Neural Networks (ANN).
- **Findings:** RF and SVM showed the highest predictive performance, demonstrating the effectiveness of ensemble methods in diabetes classification [2].

2.3 Early Prediction Model for Type-2 Diabetes Based on Lifestyle Factors

- **Methods:** Random Forest, SVM, Convolutional Neural Networks (CNN).
- **Findings:** ML-based models effectively predict Type-2 diabetes using lifestyle factors. Future research should focus on integrating real-time health monitoring [3].

2.4 Explainable Stacking-Based Model for Predicting Hospital Readmission for Diabetic Patients

- **Methods:** Stacking ensemble (Random Forest, AdaBoost, XGBoost, CatBoost, Logistic Regression as meta-learner).
- **Findings:** The model achieved high accuracy in predicting 30-day hospital readmission and identified key risk factors [4].

2.5 Analysis of Prediction Accuracy of Diabetes Using Classifier and Hybrid Machine Learning Techniques

- **Dataset:** PIDD (768 instances, 9 attributes) [5].
- **Methods:** Random Forest (RF), XGBoost.
- **Findings:** XGBoost outperformed RF with 74.10% accuracy, highlighting the importance of feature selection and model optimization [5].

2.6 An Ensemble Approach for Classification and Prediction of Diabetes Mellitus Using Soft Voting Classifier

- **Dataset:** PIDD (768 instances, 9 attributes) [6].
- **Methods:** Soft Voting Classifier (Random Forest, Logistic Regression, Naïve Bayes).
- **Findings:** The ensemble model achieved 79.08% accuracy, surpassing individual classifiers [6].

2.7 Comparative Anatomization of Data Mining and Fuzzy Logic Techniques in Diabetes Prognosis

- **Dataset:** PIDD, LARS Diabetes Dataset, Abel Vikas's Diabetes Dataset [7].
- **Methods:** Random Forest, SVM, Decision Trees, k-NN, Fuzzy Expert Systems (FES).
- **Findings:** RF achieved 99.7% accuracy, while Fuzzy Logic-based systems reached 96% accuracy [7].

2.8 Predictive Models for Diabetes Management Using Big Data Analytics

- **Methods:** Statistical models, machine learning, pattern recognition.
- **Findings:** Predictive models aid in early diagnosis and risk assessment. However, clinical validation remains a challenge [8].

2.9 Data Analytics Suite for Type 2 Diabetes Management

- **Methods:** Multi-tier classification, risk prediction models, treatment response prediction.
- **Findings:** ML-based analytics improve clinical decision-making, but scalability and validation need improvement [9].

2.10 Enhancing Diabetes Prediction with Explainable AI and Cloud-Based Deployment

- **Methods:** Hybrid ML models, cloud integration.
- **Findings:** Cloud-based AI systems improve scalability and accessibility in diabetes management [10].

3. Comparison of Studies

Study	Methodology	Accuracy (%)	Key Findings
Retracted Study	Decision Tree, RF, SVM	83 (retracted)	Data integrity issues [1]
Kavakiotis et al.	RF, SVM, k-NN, ANN	High	RF and SVM performed best [2]
Type-2 Lifestyle Model	RF, SVM, CNN	High	Lifestyle-based prediction is effective [3]
Hospital Readmission Model	RF, AdaBoost, XGBoost, CatBoost	High	Identified key predictors for readmission [4]
Hybrid Classifier Study	RF, XGBoost	74.10	XGBoost outperformed RF [5]
Ensemble Voting Classifier	RF, LR, Naïve Bayes	79.08	Ensemble learning improved accuracy [6]
Data Mining vs. Fuzzy Logic	RF, k-NN, Decision Trees, FES	99.7 (RF)	RF excelled in structured data [7]
Big Data Predictive Models	Statistical models, ML	N/A	Potential for risk forecasting in diabetes [8]
Data Analytics Suite	ML-based analytics	N/A	Improves decision-making but lacks scalability [9]
Cloud-Based AI	Hybrid ML, Cloud	High	Enhances accessibility for diabetes prediction [10]

4. Conclusion

Machine learning has significantly enhanced diabetes prediction, classification, and management. The studies reviewed demonstrate the effectiveness of ML models such as Random Forest, Support Vector Machines, XGBoost, and ensemble techniques. Advanced methods like explainable AI and cloud-based solutions show promise in improving clinical decision-making and scalability.

Despite these advancements, challenges remain. Many studies rely on a single dataset, limiting generalizability. Computational complexity, ethical concerns, and the need for clinical validation pose additional hurdles. Future research should focus on real-world implementation, multi-institutional validation, and hybrid models that integrate deep learning and traditional statistical approaches.

5. References

[1] Krishnamoorthi (2022). A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques (Retracted). [2] Kavakiotis et al. (2017). Diabetes Prediction Using Machine Learning Techniques. [3] Hirnak et al. (2020). Early Prediction Model for Type-2 Diabetes Based on Lifestyle Factors. [4] Lu & Uddin (2022). Explainable Stacking-Based Model for Predicting Hospital Readmission. [5] Hybrid Classifier Study (Year Unknown). [6] Ensemble Approach for Diabetes Classification (Year Unknown). [7] Comparative Anatomization of Data Mining and Fuzzy Logic Techniques (2020). [8] Predictive Models for Diabetes Management Using Big Data Analytics (Year Unknown). [9] Philip et al. (2022). A Data Analytics Suite for Type 2 Diabetes. [10] Enhancing Diabetes Prediction with Explainable AI and Cloud-Based Deployment (Year Unknown).