IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Predicting Blood Levels: A Machine Learning Based Approach For Diabetes Management

C.V. Madhusudan Reddy [1], V Aishwarya [2], GK Srujana[3], M Srujana[4], U Asma Mubeen[5]

[1,2,3,4,5] Department of Computer Science Engineering (AI), St. Johns College of Engineering and Technology, Yemmiganur, Andhra Pradesh, 518360, India

Abstract: The paper reports the results of the analysis of large-scale study data of diabetes prediction by employing an array of machine learning models. The Random Forest model is found to be highly discriminatory on the training set if its Area Under the Curve (AUC) value is close to one. Overfitting or erroneous classification threshold may be one of the probable issues as it has low accuracy in properly tagging data. The research also contrasted the Gaussian Naive Bayes, XgBoost, CatBoost, Gradient Boosting, and Logistic Regression models. Consistent performance of logistic regression on datasets with mid-level accuracy and AUC revealed equally well-balanced capability in classification and ranking.

XgBoost and CatBoost did well in generalization and test data accuracy, and Gaussian Naive Bayes did well on the training data but suffered a dramatic decline in performance when executed on test and unseen data, possibly due to an overfitting problem. Gradient boosting also had a very close margin of accuracy when run on unseen data but had excellent discriminative capability on all the training, test, and unseen data and excellent generalization from the training data. The research ended with an agreement that which model would be used would be based on whether the specific requirement of the application was class discrimination, label prediction, or both being more salient.

Index Terms - Diabetes, Blood Glucose Monitoring, Machine Learning, IoT, Classification, Prediction, Healthcare Monitoring.

I. INTRODUCTION

Strengthening disease preventive practices and protecting populations against health threats are the core mandate of the critical public health discipline. Governments across the globe utilize huge parts of their GDP in public health programs, which have strengthened health care systems and boosted life expectancy, according to the world Health Organization [1]. Likewise, Williams et al. [1] emphasized the significance of regional and international action on diabetes health expenditure, reflecting the disease's growing cost to health care systems.

Chronic and genetic diseases, among which one of the most common and lethal is diabetes mellitus, are becoming novel risks among a variety of public health issues over recent decades. Diabetes mellitus, as the American Diabetes Association states [3], is a metabolic disorder with an impact on the body's capacity to metabolize glucose normally in the blood. Type 1 diabetes (T1D) resulting from an absolute lack of insulin secretion, and Type 2 diabetes (T2D) resulting from insulin resistance or deficiency of insulin formation are the two most important forms of the disease. Consistent with Acciaroli et al. [4], T2D accounts for 90–95% of all cases of diabetes and has grown at a quicker rate than T1D. Undiagnosed diabetes can cause critical conditions such as cardiovascular illnesses, stroke, kidney failure, and neuropathy, reports Tun et al. [5]. Ongoing blood glucose (BG) monitoring is essential to the management of diabetes in order to reduce its danger.

II. LITERATURE REVIEW

Several studies have explored diabetes classification and prediction using machine learning and IoT-based systems. The growing prevalence of diabetes worldwide has necessitated advancements in early detection and monitoring techniques.World Health Organization [1] highlighted the importance of physical activity in mitigating the risks associated with diabetes and other chronic diseases. Williams et al. [2] discussed the economic burden of diabetes on healthcare systems, emphasizing the need for cost-effective management solutions. American Diabetes Association[3] categorized diabetes based on insulin secretion and resistance, providing a foundation for machine learning-based classification approaches. Acciaroli et al.[4] reviewed continuous glucose monitoring (CGM) sensors, highlighting their role in real-time blood glucose tracking.[5] Tun et al. investigated the correlation between diabetes and stroke, stressing the significance of early diagnosis. [6] Davies et al. examined hyperglycemia management in type 2 diabetes, proposing various treatment methodologies.

[7] Bruen et al. explored glucose sensing innovations, while [8] Wadhwa and Babber applied artificial intelligence techniques for diabetes prediction. [9] Tedeschi and Sciancalepore analyzed the role of edge and fog computing in healthcare infrastructure, addressing real-time data processing challenges .The integration of machine learning models in healthcare was further examined.[10]Schaar et al., who demonstrated the effectiveness of AI-based methods in pandemic response.[11]Arnold discussed ethical concerns surrounding

AI in medicine, particularly regarding decision-making and patient privacy. Wearable sensor technologies have also been leveraged for diabetes monitoring. [12] Kim and Huh proposed an AI-based e-healthcare solution.[13] Ali et al. developed a framework incorporating wearable sensors and social networking data. [14] Saji et al. introduced an IoT- based healthcare module for real-time monitoring. Several classification models have been implemented for diabetes diagnosis. [15] Qawqzeh et al. utilized logistic regression with photoplethysmogram (PPG) waveform analysis.[16] Pethunachiyar employed support vector machines for patient classification. [17] Gupta et al. compared Naïve Bayes and SVM techniques for diabetes classification.

III. METHODOLOGIES

The proposed system employs a machine learning- based approach combined with Internet of Things (IoT) technology for efficient diabetes classification and blood glucose (BG) level prediction shown in fig 1. The methodology consists of the following key phases:

i. Data Collection and Preprocessing

- The system utilizes the PIMA Indian Diabetes Dataset for model training and evaluation.
- Data cleaning, normalization, and feature selection techniques are applied to enhance model performance.

ii. Machine Learning-Based Diabetes Classification

- The classification phase categorizes diabetes cases into predefined classes using three algorithms:
  1) Random Forest (RF)
  2) Multilayer Perceptron (MLP)
  3) Logistic Regression (LR)
- Each classifier is trained on historical diabetes data to recognize patterns associated with different diabetes types.

iii. Predictive Analysis of Blood Glucose Levels

- These models provide insights into diabetes progression and assist in proactive disease management.

iv. IoT-Enabled Real-Time Monitoring

- BLE-based sensors collect real-time physiological data (e.g., blood glucose levels, weight, vital signs).
- The data is transmitted via Bluetooth Low Energy (BLE) devices to a mobile application.
- The system integrates Apache Kafka (for real- time streaming) and MongoDB (for secure data storage).

v. Data Processing and Decision Support System

- The collected data is processed using cloud- based machine learning models.
- Patients receive personalized insights, alerts, and recommendations based on real-time and historical data.

- The system helps in early intervention by notifying users about abnormal glucose fluctuations.

IV. RESULTS AND DISCUSSION

The proposed system was evaluated using the PIMA Indian Diabetes Dataset, focusing on classification and prediction accuracy.

Diabetes Classification Performance: Three classifiers- Random Forest (RF), Multilayer Perceptron (MLP), and Logistic Regression (LR)—were used to classify diabetes cases. MLP achieved the highest accuracy of 86.083%, outperforming other models.

Diabetes Prediction Performance: Predictive models used include Long Short-Term Memory (LSTM), Moving Averages (MA), and Linear Regression (LR). LSTM achieved the best prediction accuracy of 87.26%.

Comparative Analysis: The proposed approach was compared with state-of-the-art methods. Results demonstrated improved accuracy, adaptability, and efficiency for healthcare applications.

IoT-Based Diabetes Monitoring System: A hypothetical BLE (Bluetooth Low Energy)-based real-time monitoring system was proposed. Apache Kafka (for data streaming) and MongoDB (for storage) were integrated for efficient health tracking. The system enables patients to self-monitor blood glucose levels and receive real-time insights.

Discussion: The study confirms that advanced ML models like MLP and LSTM significantly enhance diabetes classification and prediction. The integration of IoT-based monitoring further improves patient management and early detection. The results suggest that AI-driven healthcare solutions can be highly effective for diabetes management.

V. SYSTEM ARCHITECTURE

The system architecture begins with a raw dataset as the input, which undergoes a series of preprocessing steps to ensure data quality and suitability for machine learning. This includes replacing missing values, applying appropriate scaling methods to normalize the data, and employing rule-based approaches for initial data transformation. Following preprocessing, the dataset is split into training and testing subsets. The training subset is used to train machine learning models using various algorithms such as Naïve Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), and Histogram-Based Gradient Boosting (HBGB). During this phase, hyperparameter settings are applied to optimize model performance. The trained model is then validated against the testing data to evaluate its accuracy and generalization ability. Finally, the output of this process is a tested and validated machine learning model ready for deployment or further analysis.

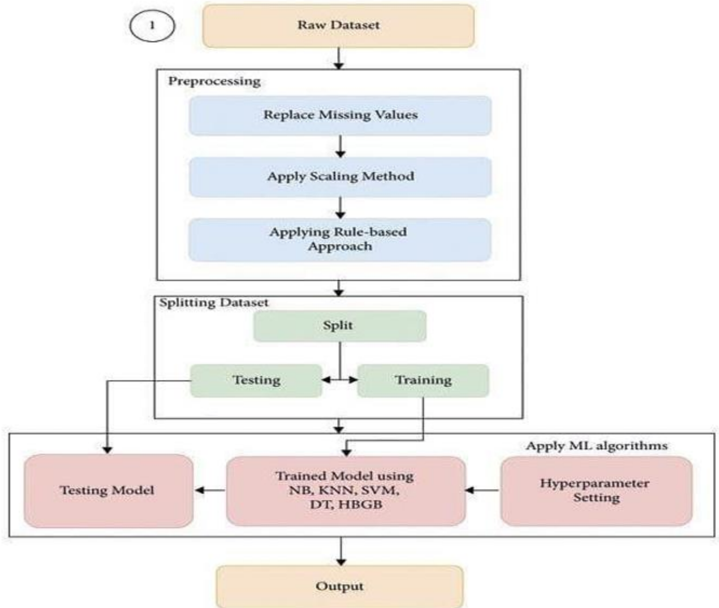

Fig.1 System Architecture

VI. CONCLUSION

The proposed system effectively enhances diabetes classification and prediction using machine learning and IoT-based solutions. Through the integration of random forest, multilayer perceptron (MLP), logistic regression, LSTM, and moving averages (MA), the approach has demonstrated high accuracy in diagnosing and predicting diabetes progression. Moreover, the IoT-enabled real-time monitoring system leverages Bluetooth Low Energy (BLE) devices, Apache Kafka for real-time data streaming, and MongoDB for storage, allowing diabetic patients to track and manage their blood glucose levels efficiently. This approach significantly improves patient engagement, early detection, and intervention. Despite achieving promising results, the study acknowledges the need for more extensive clinical validation and larger datasets to further refine accuracy. Future enhancements may include advanced deep learning models, personalized treatment recommendations, and blockchain-based security frameworks to protect sensitive patient data. In conclusion, the proposed system bridges the gap between traditional diabetes management and modern AI-driven healthcare solutions, offering a scalable, efficient, and data-driven approach to chronic disease management. With continuous advancements, this system could play a crucial role in enhancing healthcare services and improving the quality of life for diabetic patients worldwide.

VII. FUTURE SCOPE

The proposed system presents a significant step toward AI-driven diabetes management, but several areas require further exploration to enhance its effectiveness and real-world applicability. Enhancement of Machine Learning Models :The use of more complex deep learning architectures such as Transformers and Attention-based models could improve accuracy in diabetes prediction. Developing hybrid models that combine CNN, LSTM, and reinforcement learning can help in early-stage diabetes detection.

Real-World Deployment and Data Expansion: Future research should focus on collecting real-time data from hospitals, clinics, and IoT-based devices to improve model training. Multi-ethnic and diverse population datasets will enhance generalization and reduce bias in predictions.

Personalized Healthcare Solutions: AI can be utilized to provide individualized treatment recommendations based on patient history and lifestyle. Nutritional and physical activity guidance can be integrated into the system for improved diabetes control.

IoT and Edge Computing Integration: Future systems can leverage Edge Computing and 5G technology to reduce latency and improve real-time processing of patient data. Smart wearables equipped with continuous glucose monitoring (CGM) sensors can help patients track fluctuations efficiently.

Security, Privacy, and Blockchain Integration: Blockchain technology can be integrated to ensure secure and transparent medical data storage. The adoption of federated learning can allow AI models to be trained on decentralized patient data without compromising privacy.

Clinical Validation and Regulatory Approvals: Extensive clinical trials are needed to validate the effectiveness of AI models in real-world medical settings. Collaboration with healthcare regulatory authorities can facilitate approvals for AI-based diabetes management tools.

Future research should focus on enhancing AI-driven diabetes prediction with better data, advanced models, IoT integration, privacy measures, and real-world

implementation. These improvements will contribute to a more efficient, secure, and personalized approach to diabetes management.

VII. REFERENCES

[1] World Health Organization, *Global Action Plan on Physical Activity 2018-2030: More Active People for a Healthier World*, World Health Organization, Geneva, Switzerland, 2019.

[2] R. Williams, S. Karuranga, B. Malanda et al., "Global and regional estimates and projections of diabetes-related health expenditure: results from the International Diabetes Federation Diabetes Atlas," *Diabetes Research and Clinical Practice*, vol. 162, Article ID 108072.

[3] G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Calibration of minimally invasive continuous glucose monitoring sensors: state-of-the-art and current perspectives," *Biosensors*, vol. 8, no. 1, 2018.

[4] NN Tun, G. Arunagirinatha, SK Munshi, and JM Pappacha, "Diabetes mellitus and stroke: a clinical update," *World Journal of Diabetes*, vol. 8, no. 6, 2017.

[5] MJ Davies, DA D'Alessio, J. Fradkin et al., "Management of hyperglycemia in type 2 diabetes," *Diabetologia*, vol. 61, no. 12, pp. 2461–2498.

[6] D. Bruen, C. Delaney, L. Florea, and D. Diamond, "Glucose sensing for diabetes monitoring: recent developments," *Sensors*, vol. 17, no. 8, 2017.

[7] S. Wadhwa and K. Babber, "Artificial intelligence in healthcare: predictive analysis of diabetes using machine learning algorithms," in Proceedings of the International Conference on Computational Science and Its Applications, pp. 101-1 354–366, Springer, Cagliari, Italy, July 2020.

[8] P. Tedeschi and S. Sciancalepore, "Edge and fog computing in critical infrastructures: analysis, security threats, and research challenges," in Proceedings of the 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 101-1 1–10, IEEE, Stockholm, Sweden, June.

[9] MVD Schaar, AM Alaa, A. Floto et al., "How artificial intelligence and machine learning can help healthcare systems respond to COVID-19," Machine Learning, vol. 110, no. 1, pp. 1–14.

[10] MH Arnold, "Teasing out artificial intelligence in medicine: an ethical critique of artificial intelligence and machine learning in medicine," Journal of Bioethical Inquiry, vol. 18, no. 1, pp. 121–139.

[11] S. Kim and J. Huh, “Artificial intelligence-based electronic healthcare solution,” Advances in Computer Science and Ubiquitous Computing, Springer, Singapore, pp. 575–581.

[12] F. Ali, SE Sappagh, SMR Islam et al., “An intelligent healthcare monitoring framework using wearable sensors and social networking data,” Future Generation Computer Systems, vol. 114, pp. 114-1 23–43.

[13] M. Saji, M. Sridhar, A. Rajasekaran, RA Kumar, A.Suyampulingam, and N. Krishna, “IoT-based intelligent healthcare module,” Advances in Intelligent Systems and Computing, Springer, Singapore, pp. 765– 774.

[14] Qawqzeh YK, Bajahzar AS, Jemmali M, Autumn MM, and A. -Aljaoui, “Classification of diabetes using photoplethysmogram (PPG) waveform analysis: logistic regression modeling,” BioMed Research International, vol. 2020, Article ID 3764653, 6 pages.

[15] GA Pethunachiyar, “Classification of diabetes patients using kernel-based support vector machines,” in Proceedings of the 2020 International Conference on Computer Communication and Informatics (ICCCI), pp. 101-1 1–4, IEEE, Coimbatore, India, January.

[16] S. Gupta, HK Verma, and D. Bhardwaj, “Classification of diabetes using Naïve Bayes and support vector machine as a technique,” in Operations Management and Systems Engineering, Springer, Singapore, pp. 365–376.

[17] DK Choubey, M Kumar, V Shukla, S Tripathi, and VK Dhandhania, “Comparative analysis of classification methods with PCA and LDA for diabetes,” Current Diabetes Reviews, vol. 16, no. 8, pp. 833–850.

[18] M. Maniruzzaman, MJ Rahman, B. Ahammed, and MM Abedin, “Classification and prediction of diabetes.”