Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 7, Issue 4, Pages 18-24, October-December 2023

A Survey on Implementing Machine Learning Algorithm to Predict Diabetes Stages and Preventing Elevated Blood Glucose Levels

Kathiravan A.[1*], Dr. T. Ananth kumar[2] & Dr. P. Kanimozhi[3]

[1-3]*Department of Computer Science and Engineering, IFET College of Engineering, Tamilnadu, India.*
*Corresponding Author Email: kathiravan2002anbarasan@gmail.com**

DOI: https://doi.org/10.46759/IIJSR.2023.7403

Article Received: 09 August 2023 *Article Accepted: 25 October 2023* *Article Published: 06 November 2023*

ABSTRACT

Diabetes, a chronic metabolic disorder, A health condition with rising prevalence. Accurate prediction of diabetes stages and proactive management of elevated blood glucose levels are crucial for effective treatment and prevention of complications. This study investigates the implementation of machine learning algorithms, specifically Support Vector Machine (SVM), Random Forest, and k-Nearest Neighbors (KNN), to address these critical aspects of diabetes care. In this research, a comprehensive data-set comprising clinical and demographic information of individuals was utilized. Data preprocessing techniques, including feature selection, normalization, and handling of missing values, were employed to prepare the data-set for modeling. These Algorithms were trained and evaluated for their effectiveness in predicting diabetes stages and identifying individuals at risk of elevated blood glucose levels. The previous techniques used in diabetes prediction performs with moderate accuracy and optimization which will not sufficient to attain maximum level of prediction.

Keywords: Diabetes prediction; Level of blood sugar; Support vector machine; K-nearest neighbour.

1. Introduction

In an age characterized by an escalating global diabetes epidemic, the early identification and prevention of elevated blood glucose levels have become paramount in ensuring improved healthcare outcomes. This introduction lays the foundation for their primary emphasis has been on handling low-dimensional data, which restricts their capacity to fully represent the intricate and diverse aspects of diabetes. The previous model capacity to forecast outcomes and offer complex insights is hampered by this glaring restriction in the range of features that can be included. Given these difficulties, the previous research attempts only the shortfalls by creating a more reliable diabetes prediction system that only makes use of moderate outcomes by machine learning techniques to predict the diabetes condition, the working on these papers conveys us the Low end accuracy outcomes within the domain of diabetes prediction.

2. Literature Survey

A comprehensive survey of previous diabetes prediction techniques reveals a range of limitations that have spurred the need for more advanced and effective approaches. Many earlier methods faced significant challenges, predominantly in the form of lower accuracy rates, casting a shadow of doubt on the reliability of their predictive models. These accuracy issues raised concerns about the potential for misdiagnosis and delayed intervention, both critical aspects in the realm of diabetes management [1-3].

Moreover, the majority of these techniques were predominantly designed to process low-dimensional data, which inherently restricted their ability to capture the intricate and multifaceted nature of diabetes. This limitation in the scope of features employed limited their predictive power, hindering their potential to provide nuanced and comprehensive insights into the disease. As the understanding of diabetes continues to evolve, there's a growing

imperative to address these shortcomings. The limitations observed in previous methodologies underscore the necessity for developing a more robust and sophisticated diabetes prediction system. Such a system would harness advanced machine learning techniques while fully utilizing the spectrum of relevant features, ultimately enhancing prediction accuracy and improving the overall effectiveness of diabetes management. This research aims to bridge these gaps and contribute to the advancement of diabetes prediction.

This paper [4] Garg, Priyanka, and Navneet Duggal studies about Diabetes prediction where Diabetes mellitus is a prevalent and debilitating chronic disease with a significant impact on public health. In the quest for early diagnosis and intervention, ML methods have emerged as a potent tool. This literature survey explores the application of these technique in predicting the onset of diabetes, shedding light on notable developments and challenges in this critical domain. Numerous studies have demonstrated the feasibility and efficacy of machine learning algorithms in diabetes prediction. Decision Trees, Random Forests, and Neural Networks are frequently employed for their capacity to discern complex patterns in diverse datasets. These models utilize a wide array of features, such as genetic, clinical, and lifestyle data, to enhance prediction accuracy. Furthermore, feature selection and engineering play a pivotal role in model development [5]. Identifying the most informative features can improve prediction performance and reduce model complexity.

Tan, Kuo Ren [6] conducts a comprehensive review of machine learning approaches used in the prediction and risk assessment of diabetes. It covers a wide range of techniques and models employed in the field. The review emphasizes the significance of early diagnosis and personalized risk assessment in diabetes management. It offers insights into the potential for machine learning to predict diabetes, which can be a crucial step in preventing elevated blood glucose levels [7]. The study discusses various data sources, including electronic health records and wearable devices, and their utilization in training predictive models. Additionally, it delves into challenges such as data imbalance and feature selection in diabetes prediction, providing a well-rounded understanding machine learning for diabetes management [8].

This review [9] of Liu, Mei-Yuan focuses on predicting the problem of developing Type 2 diabetes using machine learning. Identifying individuals at risk is a crucial step in preventing elevated blood glucose levels and associated complications. The paper provides an extensive overview of machine learning methods and models applied to risk assessment. It highlights the significance of early intervention and lifestyle modifications in managing diabetes [10]. The study also examines data sources, including clinical records and genetic data, and the role they play in improving prediction accuracy. By summarizing the strengths and limitations of various algorithms, it aids researchers and healthcare professionals in selecting the most suitable approaches for early risk prediction and prevention of elevated blood glucose levels.

According to Chou, Chun-Yan in predicting readmissions of diabetes patients, a critical aspect of diabetes management [11], that can indicate disease progression and help prevent elevated blood glucose levels. The paper conducts a comparative analysis of various machine learning approaches used for this purpose. It explores how patient data, such as medical history, treatment details, and demographics, can be leveraged to predict readmissions. The study provides insights into model performance and identifies the most effective techniques

[12]. It emphasizes the importance of early intervention and personalized care to reduce readmissions. The findings are valuable for healthcare providers aiming to improve patient outcomes and for researchers working on enhancing the efficiency of diabetes management through machine learning [13].

This comprehensive review [14] by Binhowemel examines the role of machine learning in various aspects of diabetes management. It provides insights into diabetes prediction, monitoring, and treatment, with a focus on preventing elevated blood glucose levels. The paper covers a wide array of machine learning techniques, data sources, and applications, making it a valuable resource for understanding the broader context of machine learning in diabetes care. It emphasizes the importance of personalized medicine and data-driven decision-making in diabetes management [15]. The review explores the challenges and opportunities in integrating machine learning into clinical practice and offers innovative ideas in this domain. By providing an in-depth overview of the current state of machine learning in diabetes, the paper serves as a reference for healthcare professionals, researchers, and policymakers interested in enhancing diabetes care.

This review [16] focuses on machine learning applications in the prediction and management of Type 1 diabetes, an autoimmune condition that requires precise glucose level control. It highlights the unique challenges associated with Type 1 diabetes and how machine learning can address them. The paper discusses various data sources, including continuous glucose monitoring and insulin dosing data, used for prediction and management. It explores machine learning models such as such as SVM, that have been employed in Type 1 diabetes research [17]. The review emphasizes the need for real-time monitoring and closed-loop control systems to prevent elevated blood glucose levels and reduce the risk of hypoglycemia. It also discusses the potential for hybrid models that combine machine learning and physiological modeling. Researchers and healthcare professionals in the field of Type 1 diabetes will find this review valuable for its insights into current applications and future directions for machine learning.

Chapman, M. John [18] focuses on the prediction of Type 2 diabetes using machine learning. Early identification is a critical step in preventing elevated blood glucose levels and the associated complications of the disease. The study employs routine health check-up data, making it a practical and cost-effective approach. It discusses the features and risk factors considered for the prediction model, including parameters such as BMI, BP and family history. The paper [19] highlights the importance of intervention strategies and lifestyle modifications for individuals at risk. It also evaluates the model's performance in terms of sensitivity, specificity, and accuracy, indicating its clinical utility. Healthcare providers and researchers looking to improve early detection and prevention of Type 2 diabetes will benefit from the insights and approach presented in this paper.

A literature survey [20] on the topic "Classification and Prediction of Diabetes Disease Using Machine Learning Paradigm with a focus on Logistic Regression and Random Forest Algorithms" reveals the significance and advancements in this field. Researchers have made substantial progress in developing robust models for the early detection and prediction of diabetes. Numerous studies have explored the importance of machine learning techniques to address the challenges associated with diabetes prediction. Logistic Regression, a widely used classification algorithm, has been instrumental in this regard. It offers a straightforward yet effective method for

identifying diabetic patients based on their health attributes [21]. Random Forest, on the other hand, is a significant algorithm to improve prediction accuracy. Its ability to handle complex, high-dimensional data makes it a valuable tool in diabetes prediction. Researchers have demonstrated its effectiveness in improving the accuracy and reliability of diabetes classification models. Several studies have focused on these two algorithms to enhance diabetes prediction. For instance, they have employed Logistic Regression to analyze patient data and predict diabetes onset accurately.

Similarly, the Random Forest algorithm has been used to build ensemble models that provide robust predictions, reducing the risk of miscalculation. These models [22] have not only shown promising results but have also opened up opportunities for further research. Researchers aim to enhance the way of predicting this condition by incorporating additional features, optimizing model parameters, and exploring novel techniques [23]. In conclusion, Logistic Regression and Random Forest have emerged as powerful tools in this domain, with researchers striving to refine and extend these models for more accurate and early detection of diabetes, ultimately contributing to improved healthcare outcomes.

The comprehensive survey on diabetes prediction unfolds a compelling narrative that underscores the significance of leveraging machine learning techniques in addressing the escalating global diabetes crisis. By aggregating insights from a multitude of research studies and projects, this survey provides an all-encompassing overview of the state-of-the-art methodologies, models, and algorithms employed in diabetes prediction [24]. The survey begins by contextualizing the growing relevance of diabetes prediction within the broader healthcare landscape, emphasizing the importance of early diagnosis and intervention. It further delineates the pervasive challenges posed by diabetes and underscores the urgency for more accurate and timely predictive tools. The survey then delves into the core methodologies and algorithms, examining the multifaceted approaches employed in different studies.

The utilization of Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) algorithms, as exemplified in numerous projects, stands out as a formidable approach in enhancing the precision of diabetes prediction [25]. The survey underscores how these techniques are instrumental in developing predictive models capable of classifying diabetes stages and identifying individuals at risk of elevated blood glucose levels. In summary, this survey serves as a comprehensive repository of knowledge, offering valuable insights into the evolution and advancements in diabetes prediction. It illuminates the innovative methods and models that have the potential to revolutionize the field of healthcare, ultimately contributing to earlier interventions, improved patient care, and a more efficient healthcare ecosystem.

2.1. Problem Statement

The problem statement emanates from the inadequacies observed in previous diabetes prediction research. These studies have predominantly grappled with lower accuracy rates, casting uncertainty on the reliability of their predictive models. Moreover, they have been primarily focused on processing low-dimensional data, which limits their ability to capture the complex and multifaceted nature of diabetes. This stark limitation in the scope of features utilized hinders the models' predictive power and the ability to provide nuanced insights. In light of these challenges, our research aims to address these shortcomings by developing a more robust diabetes prediction

system that leverages advanced machine learning techniques, while harnessing the full spectrum of relevant features, thereby enhancing prediction accuracy and the overall effectiveness of diabetes management.

3. Conclusions

In conclusion, the survey of research papers on diabetes prediction has illuminated the various methodologies and challenges within this critical domain. It is evident that the field has witnessed significant advancements and innovations, while also grappling with limitations, such as accuracy concerns and reliance on low-dimensional data. The need for highly accurate, early prediction models that encompass a broader range of features is clear. As we move forward, it is essential to overcome these challenges and leverage advanced machine learning techniques to enhance the precision and reliability of diabetes prediction. By addressing these limitations, we can improve the quality of patient care and contribute to more effective diabetes management strategies.

Declarations

Source of Funding

This study has not received any funds from any organization.

Conflict of Interest

The authors declare that they have no conflict of interest.

Consent for Publication

The authors declare that they consented to the publication of this study.

Authors' Contribution

All the authors took part in literature review, research, and manuscript writing equally.

References

[1] Krishnamoorthi, Raja, Shubham Joshi, Hatim Z. Almarzouki, Piyush Kumar Shukla, Ali Rizwan, C. Kalpana & Basant Tiwari (2022). A novel diabetes healthcare disease prediction framework using machine learning techniques. Journal of Healthcare Engineering.

[2] Liu, Zhe, Nan Jia, Qianghuizi Zhang & Weiwei Liu (2023). Risk prediction models for postpartum glucose intolerance in women with a history of gestational diabetes mellitus: a scoping review. Journal of Diabetes & Metabolic Disorders, Pages 1-10.

[3] Staten, Lisa K., Tess D. Weathers, Celeste Nicholas, Tedd Grain, Dawn P. Haut, Patrice Duckett-Brown, Paul K. Halverson & Virginia Caine (2023). Design and Implementation of the Diabetes Impact Project: A Multisector Partnership to Reduce Diabetes Burden in Indianapolis Communities. Journal of Public Health Management and Practice, 29(5): 675-685.

[4] Garg, Priyanka & Navneet Duggal (2022). Type 2 diabetes mellitus, its impact on quality of life and how the disease can be managed-a review. Obesity Medicine.

[5] Saxena, Roshi, Sanjay Kumar Sharma, Manali Gupta & G.C. Sampada (2022). A Comprehensive review of various diabetic prediction models: a literature survey. Journal of Healthcare Engineering.

[6] Tan, Kuo Ren, Jun Jie Benjamin Seng, Yu Heng Kwan, Ying Jie Chen, Sueziani Binte Zainudin, Dionne Hui Fang Loh, Nan Liu & Lian Leng Low (2023). Evaluation of machine learning methods developed for prediction of diabetes complications: a systematic review. Journal of Diabetes Science and Technology, 17(2): 474-489.

[7] Ahmed, Arfan, Sarah Aziz, Alaa Abd-Alrazaq, Faisal Farooq, Mowafa Househ & Javaid Sheikh (2023). The Effectiveness of Wearable Devices Using Artificial Intelligence for Blood Glucose Level Forecasting or Prediction: Systematic Review. Journal of Medical Internet Research, 25: e40259.

[8] Shaukat, Zain, Wisal Zafar, Waqas Ahmad, Ihtisham Ul Haq, Ghassan Husnain, Mosleh Hmoud Al-Adhaileh, Yazeed Yasin Ghadi & Abdulmohsen Algarni (2023). Revolutionizing Diabetes Diagnosis: Machine Learning Techniques Unleashed. Healthcare, 11(21): 2864. https://doi.org/10.3390/healthcare11212864.

[9] Liu, Mei-Yuan, Chung-Feng Liu, Tzu-Chi Lin & Yu-Shan Ma (2023). Implementing a Novel Machine Learning System for Nutrition Education in Diabetes Mellitus Nutritional Clinic: Predicting 1-Year Blood Glucose Control. Bioengineering, 10(10): 1139. https://doi.org/10.3390/bioengineering10101139.

[10] Handelsman, Yehuda, Javed Butler, George L. Bakris, Ralph A. DeFronzo, Gregg C. Fonarow, Jennifer B. Green, George Grunberger et al. (2023). Early intervention and intensive management of patients with diabetes, cardiorenal, and metabolic diseases. Journal of Diabetes and its Complications.

[11] Sabbour, Hani & Aftab Ahmad (2023). Management of Heart Failure in Patients With Diabetes Mellitus in the UAE: A Call to Action. Journal of Cardiovascular Pharmacology and Therapeutics, 28: 10742484231162248.

[12] Chou, Chun-Yang, Ding-Yang Hsu & Chun-Hung Chou (2023). Predicting the onset of diabetes with machine learning methods. Journal of Personalized Medicine, 13(3): 406.

[13] Sabapathy, Sundaresan, Surendar Maruthu, Suresh Kumar Krishnadhas, Ananth Kumar Tamilarasan & Nishanth Raghavan (2022). Competent and Affordable Rehabilitation Robots for Nervous System Disorders Powered with Dynamic CNN and HMM. Intelligent Systems for Rehabilitation Engineering, Pages 57-93.

[14] Pina, Ana F., Maria João Meneses, Inês Sousa-Lima, Roberto Henriques, João F. Raposo & Maria Paula Macedo (2023). Big data and machine learning to tackle diabetes management. European Journal of Clinical Investigation, 53(1): e13890.

[15] Zou, Xiantong, Yingning Liu & Linong Ji (2023). Machine learning in precision pharmacotherapy of type 2 diabetes—A promising future or a glimpse of hope. Digital Health, 9: 20552076231203879.

[16] Binhowemel, Samar, Mohammed Alfakhri, Khoolod AlReshaid & Abdulrahman Alyani (2023). Role of Artificial Intelligence in Diabetes Research Diagnosis and Prognosis: A Narrative Review. Journal of Health Informatics in Developing Countries, 17(2).

[17] Afentakis, Ioannis, Rebecca Unsworth, Pau Herrero, Nick Oliver, Monika Reddy & Pantelis Georgiou (2023). Development and validation of binary classifiers to predict nocturnal hypoglycemia in adults with type 1 diabetes. Journal of Diabetes Science and Technology.

[18] Fujihara, Kazuya & Hirohito Sone (2023). Machine Learning Approach to Drug Treatment Strategy for Diabetes Care. Diabetes & Metabolism Journal, 47(3): 325-332.

[19] Chapman, M. John (2022). HDL functionality in type 1 and type 2 diabetes: new insights. Current Opinion in Endocrinology, Diabetes, and Obesity, 29(2): 112.

[20] Saxena, Roshi, Sanjay Kumar Sharma, Manali Gupta & G.C. Sampada (2022). A Comprehensive review of various diabetic prediction models: a literature survey. Journal of Healthcare Engineering.

[21] Moreira, Tatiana Rebouças, Francisca Diana da Silva Negreiros, Maria de Jesus Nascimento de Aquino, Lucilane Maria Sales da Silva, Thereza Maria Magalhães Moreira & Raimundo Augusto Martins Torres (2023). Digital technology and its effects on knowledge improvement for diabetes management: An integrative review. International Journal of Nursing Practice, 29(1): e13029.

[22] Mohsen, Farida, Hamada RH Al-Absi, Noha A. Yousri, Nady El Hajj & Zubair Shah (2023). A scoping review of artificial intelligence-based methods for diabetes risk prediction. NPJ Digital Medicine, 6(1): 197.

[23] Rahimi, Amir Kamel, Oliver J. Canfell, Wilkin Chan, Benjamin Sly, Jason D. Pole, Clair Sullivan & Sally Shrapnel (2022). Machine learning models for diabetes management in acute care using electronic medical records: A systematic review. International Journal of Medical Informatics, 162: 104758.

[24] Oladipupo, Matthew Abiola, Princewill Chima Obuzor, Babatunde Joseph Bamgbade, Abidemi Emmanuel Adeniyi, Kazeem M. Olagunju & Sunday Adeola Ajagbe (2023). An Automated Python Script for Data Cleaning and Labeling using Machine Learning Technique. Informatica, 47(6).

[25] Rahman, Md. Ashikur, Lway Faisal Abdulrazak, Md. Mamun Ali, Imran Mahmud, Kawsar Ahmed & Francis M. Bui (2023). Machine Learning-Based Approach for Predicting Diabetes Employing Socio-Demographic Characteristics. Algorithms, 16(11): 503. https://doi.org/10.3390/a16110503.