International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Predictive Modelling And Artificial Intelligence Strategies For Early Identification Of Type 2 Diabetes Risk: A Review

Preeti Verma
Research Scholar, PhD (Computer Science & Engineering), Integral University Lucknow
vermapree168@gmail.com
Dr. Mohd Akbar
Associate Professor, Integral University, Lucknow
akbar@iul.ac.in

ABSTRACT
Type 2 Diabetes Mellitus (T2DM) is an alarmingly emerging issue in the health care of the world and needs effective measures to be taken to detect and prevent this deadly illness at an early stage. Recent progress in artificial intelligence (AI) and predictive modelling has led to the creation of intelligent systems that can be used to identify persons at risk before the clinical manifestations of the condition develop. In this review paper, it is observed that machine learning and deep learning techniques play a vital role in the early prediction of Type 2 Diabetes through clinical, lifestyle, and behavioural data. The research also emphasizes frequently employed predictive algorithms, the most important evaluation metrics, and the opportunities of AI-powered systems to facilitate an early risk evaluation and preventive health. Also, the paper addresses significant issues in AI-based diabetes prediction such as imbalanced data, insufficient feature integration, inability of models to provide interpretability, and the insufficient use of them in practice. The review highlights the importance of creating interpretable, scalable, and generalizable AI-based predictive models to improve screening in its initial phases and decrease the burden of Type 2 Diabetes in the world.
Keywords: Type 2 Diabetes Mellitus, Artificial Intelligence, Predictive Modelling, Machine Learning, Early Risk Prediction, Healthcare Analytics.

1. INTRODUCTION

The Type 2 Diabetes Mellitus (T2DM) is a major population health challenge that affects millions of lives across the world. The reports done on global health have revealed that the level of diabetes has been increasing due to high urbanization rates, sedentary lifestyle, poor dietary habits and excessive obesity [1]. The T2DM condition is characterized by insulin resistance and lack of glucose metabolism which can lead to severe cardiovascular disease, kidney failure, neuropathy and loss of vision unless promptly diagnosed and treated at the initial stage.

Traditional diagnosing procedures tend to identify diabetes only at the point when the unusual sugar level is detected, or when the clinical symptoms appear. However, the process of preventive measures and management of the disease can be improved significantly through the assistance of early detection of individuals who are going to be affected with diabetes at the pre-diabetes levels [2].

The past several years has seen the discussion of the use of Artificial Intelligence (AI) and predictive modelling methods extensively in the context of the healthcare research. The complex patterns that would have been challenging to identify within the traditional statistics could be identified in deep learning algorithms and machine learning, and they could be utilized to analyze large volumes of medical data [3]. The relevant technologies can be used to create intelligent decision-support systems which are capable of forecasting the danger of diseases in advance and assist in early clinical intervention.

Figure 1: Key Challenges in Diabetes Management and Healthcare Decision-Making

No matter how far they have gone in these advancements, there are still a number of challenges regarding the creation of dependable AI-based systems to predict diabetes at the early stage. Numerous available researches are concentrated on the classification of the diseases instead of early indicators of risks [4]. Further, the fact that lifestyle and behavioral factors are not much integrated, that predictive models do not lend to interpretation, and that the systems are not well generalizable across populations also limits the practical use of such systems.

Thus, the given review paper will review existing AI-driven predictive models of Type 2 Diabetes risk, define existing gaps in the field of research, and outline the possibilities of establishing more sustainable and scalable early detection models [5].

1.1.Research Objectives

The main aim of the study is to develop and test an artificial intelligence-sensitive framework to predict Type 2 Diabetes Mellitus (T2DM) at its onset utilizing organized clinical, lifestyle, and behavioural information.

The specific objectives of the study are as follows:

- To design and develop AI-driven predictive models using machine learning and deep learning techniques capable of identifying early risk patterns associated with Type 2 Diabetes prior to clinical diagnosis.
- To evaluate and compare the performance of multiple AI models including traditional machine learning algorithms and deep learning architectures based on standard evaluation metrics such as accuracy, precision, recall, F1-score, and ROC–AUC.

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

- To address data imbalance, bias, and generalization challenges through advanced preprocessing, feature engineering, and validation strategies, ensuring reliable performance across diverse populations.
- To propose a scalable and deployable early warning system for Type 2 Diabetes risk assessment that can be effectively utilized in real-world healthcare and preventive screening environments, including resource-constrained settings.

2. OVERVIEW OF TYPE 2 DIABETES AND RISK FACTORS

Diabetes Mellitus type 2 is a disease that is acquired over time because of a complex of genetic, metabolic, lifestyle, and environmental influences. Being able to detect the disease at an early stage is especially difficult since people may not show any symptoms at the initial stages of the disease [6].

There are some risk factors that have been identified to cause T2DM. These are age, obesity, family history of diabetes, hypertension, sedentary lifestyle, poor dietary habits, smoking and high cholesterol levels [7]. Also, common clinical predictors include metabolite-related values, including fasting blood glucose, body mass index (BMI), insulin resistance, and lipids.

Non-new risk assessment techniques are based on statistical models and clinical screening instruments. Nonetheless, these methods are frequently inadequate to describe interplay between various risk variables. Presuming a more sophisticated alternative, AI-based predictive models can analyze multidimensional data sets and reveal hidden relationships between variables.

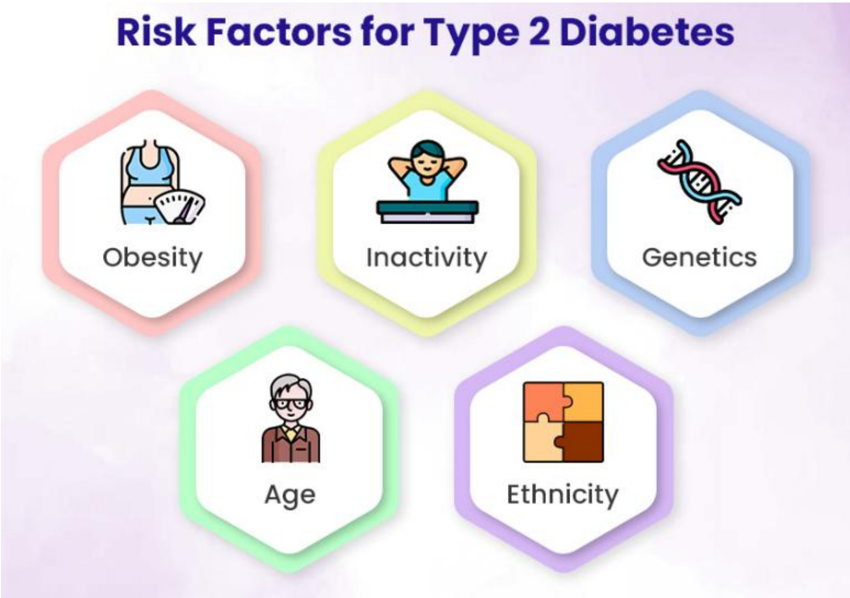

Figure 2: Risk factors for Type 2 diabetes

3. ARTIFICIAL INTELLIGENCE IN HEALTHCARE

Implementation of Artificial Intelligence has revolutionized various fields of healthcare including medical imaging, disease diagnosis, drug discovery and predictive analytics at a relatively rapid rate [8]. Based on calculational algorithms, AI systems learn and predict or recommend based on the data using patterns.

In the context of diabetes prediction, AI techniques are primarily categorized into two groups:

3.1. Machine Learning Approaches

Machine learning algorithms learn patterns from historical data to predict future outcomes. Commonly used algorithms include:

- Logistic Regression
- Decision Trees
- Random Forest
- Support Vector Machines (SVM)
- K-Nearest Neighbors (KNN)
- Gradient Boosting Algorithms

These algorithms are popular because of their capability to process structured medical data and also because they make the comparably understandable forecasts.

3.2. Deep Learning Techniques

Deep learning models rely on artificial neural networks with more than one hidden layer to be able to analyze high-dimensional complicated data. The standard architectures that are deployed to predict diabetes are:

- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Deep Belief Networks

Deep learning models are particularly effective when large datasets are available, although they often require higher computational resources and may lack interpretability.

Healthcare analytics has used various artificial intelligence algorithms to predict the risk of diabetes. Such methods vary in their complexity, interpretation, and prediction [9]. Table 1 summarises some of the widely used AI methods in the prediction of diabetes.

Table 1: Artificial Intelligence Techniques Used in Diabetes Prediction

This table summarizes different algorithms used in previous studies.

AI Technique	Type	Key Characteristics	Application in Diabetes Prediction
Logistic Regression	Machine Learning	Simple and interpretable statistical model	Used for baseline diabetes risk prediction
Decision Tree	Machine Learning	Rule-based classification model	Identifies important clinical features
Random Forest	Ensemble Learning	Combines multiple decision trees	Improves prediction accuracy
Support Vector Machine (SVM)	Machine Learning	Effective for high-dimensional data	Used for classification of diabetic and non-diabetic patients
Artificial Neural Network (ANN)	Deep Learning	Multi-layer neural architecture	Detects complex nonlinear relationships

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Convolutional Neural Network (CNN)	Deep Learning	Feature extraction from structured data	Applied in medical image-based diagnosis

4. PREDICTIVE MODELLING TECHNIQUES FOR DIABETES RISK ASSESSMENT

Predictive modelling is described as the act of forecasting future results through statistical and computational methods using data of the past. In diabetes prediction studies, medical datasets, which comprise demographic, physiological, and behavioral variables are used in the training of predictive models [10].

The majority of studies adhere to an organised predictive modelling pipeline that has the following steps:

4.1. Data Collection and Preprocessing

The medical data is gathered like in hospitals, health surveys, or even in repositories. Some of the typical preprocessing actions are the treatment of missing values, normalization, feature scaling, and coding of categorical variables.

4.2. Feature Selection and Engineering

The predictions of the most significant predictors of diabetes risk are determined with the help of the methods of the feature selection [11]. Some of the popular techniques that have been used to improve the performance of the model include Correlation analysis, principal component analysis and recursive feature elimination.

4.3. Model Training

Machine learning algorithms are trained using labelled datasets to identify patterns that are connected to the risk of diabetes [12]. Training involves changing the model parameters in order to minimize the error in prediction.

Table 2: Review of Literature on AI-Based Predictive Modelling for Early Detection of Type 2 Diabetes

Author Name	Topic Covered	Research Study Title
Manik et al. (2025) [13]	AI & ML for early diabetes detection	AI-powered predictive modelling for early Type 2 diabetes
Adua et al. (2021) [14]	ML for early detection & risk factors	Predictive model and feature significance for Type 2 diabetes
Kopitar et al. (2020) [15]	ML models for early risk identification	Early Type 2 diabetes identification using ML models
Fazakis et al. (2021) [16]	ML & ensemble methods for risk prediction	ML methods for long-term Type 2 diabetes risk
Khalifa & Albadawy (2024) [17]	AI in diabetes diagnosis & management	AI for diabetes diagnosis, prevention, and management
Farran et al. (2013) [18]	ML for diabetes & hypertension prediction	Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity
Kaur & Kumari (2022) [19]	ML-based diabetes prediction	Predictive modelling and analytics for diabetes

Kengne et al. (2014) [20]	Risk models for diabetes & CVD	Risk predictive modelling for diabetes and cardiovascular disease
Lagani et al. (2013) [21]	Models for diabetes complications	Predictive risk models for diabetes complications
Lai et al. (2019) [22]	ML techniques for diabetes prediction	Predictive models for diabetes mellitus

5. CHALLENGES IN AI-BASED DIABETES PREDICTION

Despite the proven potential of the artificial intelligence and machine learning methods on the risk prediction of Type 2 Diabetes, the application of those techniques is still faced with a number of limitations that restrict their extensive use in clinical settings [23]. These difficulties are due to the problems of the data quality, model design, interpretability, and its practical implementation in the healthcare settings. To create effective predictive systems that are reliable and clinically applicable, these limitations should be addressed [24].

5.1. Data Imbalance

Predictive modelling lack of balance of data is one of the most common issues in medical predictive modelling. In most medical datasets, there is a vast population of non-diabetic patients compared to the diabetic cases. This is an imbalance since the prevalence of diabetes among the general population is comparatively low as compared to healthy people.

In instances where machine learning models are trained using such unbalanced datasets, they will be skewed towards the majority population (non-diabetic cases). Consequently, the model might be highly accurate when taken as a whole but not able to detect correctly those that are actually under the threat of developing diabetes. This is especially problematic with the use of healthcare applications where the correct identification of minority cases is more relevant than accuracy in general [25].

In order to resolve data imbalance, some sampling methods have been advanced by the research community including oversampling, under sampling, and synthetic data generation techniques such as SMOTE. Moreover, learning and ensemble methods that are sensitive to cost can also be employed to enhance minority cases detection. Imbalanced data should be properly handled so that the predictive models can be effective and reliable to detect diabetes early in the future.

5.2. Limited Feature Integration

The other significant limitation that can be noticed in most AI-based diabetes prediction studies is the limitation on the number of predictive features used. The majority of the existing models are mostly based on the clinical and physiological evidence, including blood glucose, BMI, blood pressure, and cholesterol levels [26]. Although these are significant clinical variables that indicate metabolic wellbeing, they do not capture all the complexities leading to the development of diabetes.

The Type 2 Diabetes is a condition that is brought about by lifestyle, behavioural, environmental, and socio-demographic factors. Physical lack of exercise, unhealthy eating habits, smoking, alcohol use, stress, sleep quality and work conditions may have severe influence on metabolic wellbeing. On the same note, demographic factors like age, sex, income, education, and the level of urbanization also determine the risk of diseases [27].

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Nonetheless, a large number of predictive models do not consider these variables because of the limited data or usage of structured clinical datasets. Lack of these factors makes the models to be incapable of capturing the entire spectrum of determinants affecting the risk of diabetes. The inclusion of multi-source information such as clinical, behavioural, and socio-demographic data has the potential to contribute critically to the ability of AI-based models to predict and can increase the accuracy of early detection.

5.3. Lack of Model Interpretability

Significant issue of AI-based healthcare is model interpretability. Most modern machine learning systems, especially deep learning networks, are complicated computing architectures with numerous hidden layers. Although these models are usually highly predictive, they are black-box models, i.e. the way they do it is not easily comprehensible [28].

In clinical settings, healthcare practitioners need clear and understandable models that ensure that certain variables lead to disease forecasting. Medical practitioners and physicians should be capable of decoding the results of predictive systems so that they make the right decisions in terms of diagnosing and treating patients [29]. The other disadvantage is that AI models cannot offer clear explanations to be used in healthcare settings.

To solve this problem, researchers are paying increased attention to the XAI approaches. The significance of features analysis, SHAP and LIME methods help a researcher comprehend the importance of particular features in the predictivity of the model. The model's interpretability can be enhanced to promote the trust, transparency and acceptance of the AI-based decision support systems in clinical practice.

5.4. Poor Generalization Across Populations

Generalization is the capacity of predictive model to act efficiently on new or unseen data. The current models of diagnosing diabetes are also created on the basis of the datasets received in certain geographic areas, hospitals, or the population of a particular demographic type. Consequently, these models can be suitable in capturing the trends that can be unique to a specific people but cannot be universal.

As an illustration, differences in genetic composition, lifestyles, access to healthcare and socio-economic statuses can determine the risk of diabetes among diverse individuals. A predictive model that was trained using data of a given country or a specific group of people may, thus, give wrong predictions when used to forecast other groups.

This is one of the weaknesses that emphasize the need to come up with models based on various and representative datasets comprising several demographic groups. The predictive models can be enhanced in terms of robustness and reliability through cross-population validation and multi-center datasets. To deploy AI-based diabetes prediction systems at the global scale, it is important to ensure that they are better generalized.

5.5. Limited Real-World Deployment

Although the amount of research studies on AI-based diabetes prediction is on the rise, the majority of predictive models are still limited to laboratory set-ups. Numerous studies are more interested in the development of algorithms and testing their performance based on benchmark datasets, not carrying out these systems in the real medical practice [30].

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 7.2 Website: https://ijarmt.com ISSN No.: 3048-9458

Practical implementation entails the incorporation of predictive models in clinical practice, electronic health records and healthcare decision-support systems. Nevertheless, there are a few obstacles to this transition, and they are the privacy of data, regulations, infrastructure, and insufficient cooperation between AI scientists and medical practitioners.

Moreover, resource-constrained areas are likely to have inadequate technological facilities to deploy sophisticated AI systems in the healthcare systems. This causes numerous promising predictive models to not be translated into viable instruments of early screening and preventing diabetes.

To solve this problem, future studies need to consider developing scalable, easy-to-use and clinically validated AI systems, which can be incorporated into mainstream healthcare settings. Data scientists, clinicians, policymakers and healthcare institutions will need to collaborate to make sure that the real-world implementation is successful.

6. CONCLUSION

The artificial intelligence and predictive modelling are becoming significant in the design of early intervention to control T2DM. AI systems may be able to study extensive and multifaceted healthcare data and identify patterns and risk factors that are associated with the onset stages of diabetes development, with the help of the approach of machine learning and deep learning. The technologies possess much potential in preventing healthcare as it can perform the early risk assessment and result in timely clinical response. However, despite the fact that the current findings of the research are encouraging, there are still several limitations, which limit the effective application of AI-based models of diabetes prediction. Such problems as data imbalance, small inclusion of lifestyle and socio-demographic variables, lack of model transparency, inadequate generalization in dissimilar populations, and deficient validation in practice healthcare settings are still major issues. To overcome such limitations by enhancing the data integration, the use of explainable artificial intelligence methods and utilizing various and representative datasets will be crucial in the development of predictive systems that can be reliable and clinically useful. Most studies done in the future ought to aim at developing scalable, explainable, and resilient AI systems that can be incorporated into healthcare systems to improve early screening and preventive measures of Type 2 Diabetes.

REFERENCES

1. Lotfi, Z., Haji Hosseini, R., & Aminipour, M. (2025). Artificial Intelligence–Driven Approaches for Prediction, Management, and Complication Risk in Type 2 Diabetes: A Systematic Review. InfoScience Trends, 2(6), 1-17.
2. Kumar, M., Ang, L. T., Ho, C., Soh, S. E., Tan, K. H., Chan, J. K. Y., ... & Karnani, N. (2022). Machine learning–derived prenatal predictive risk model to guide intervention and prevent the progression of gestational diabetes mellitus to type 2 diabetes: Prediction model development study. JMIR diabetes, 7(3), e32366.
3. Mohsen, F., Al-Absi, H. R., Yousri, N. A., El Hajj, N., & Shah, Z. (2023). A scoping review of artificial intelligence-based methods for diabetes risk prediction. npj Digital Medicine, 6(1), 197.

4. Kiran, M., Xie, Y., Anjum, N., Ball, G., Pierscionek, B., & Russell, D. (2025). Machine learning and artificial intelligence in type 2 diabetes prediction: a comprehensive 33-year bibliometric and literature analysis. Frontiers in digital health, 7, 1557467.
5. Xie, Z., Nikolayeva, O., Luo, J., & Li, D. (2019). Building risk prediction models for type 2 diabetes using machine learning techniques. Preventing chronic disease, 16, E130.
6. Ismail, L., Materwala, H., Tayefi, M., Ngo, P., & Karduck, A. P. (2022). Type 2 diabetes with artificial intelligence machine learning: methods and evaluation. Archives of Computational Methods in Engineering, 29(1), 313-333.
7. Fregoso-Aparicio, L., Noguez, J., Montesinos, L., & García-García, J. A. (2021). Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. Diabetology & metabolic syndrome, 13(1), 148.
8. Ellahham, S. (2020). Artificial intelligence: the future for diabetes care. The American journal of medicine, 133(8), 895-900.
9. Hossain, M. E., Uddin, S., & Khan, A. (2021). Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes. Expert Systems with Applications, 164, 113918.
10. Oikonomou, E. K., & Khera, R. (2023). Machine learning in precision diabetes care and cardiovascular risk prediction. Cardiovascular Diabetology, 22(1), 259.
11. Peddinti, G., Cobb, J., Yengo, L., Froguel, P., Kravić, J., Balkau, B., ... & Groop, L. (2017). Early metabolic markers identify potential targets for the prevention of type 2 diabetes. Diabetologia, 60(9), 1740-1750.
12. Dankwa-Mullan, I., Rivo, M., Sepulveda, M., Park, Y., Snowdon, J., & Rhee, K. (2019). Transforming diabetes care through artificial intelligence: the future is here. Population health management, 22(3), 229-242.
13. Manik, M. M. T. G., Saimon, A. S. M., Ahmed, M. K., Hossain, S., Moniruzzaman, M., & Islam, M. S. (2025, May). Predictive modelling for early detection of type 2 diabetes using AI-driven machine learning algorithms and big data analytics. In International Conference on AI and Robotics (pp. 422-433). Cham: Springer Nature Switzerland.
14. Adua, E., Kolog, E. A., Afrifa-Yamoah, E., Amankwah, B., Obirikorang, C., Anto, E. O., ... & Tetteh, A. Y. (2021). Predictive model and feature importance for early detection of type II diabetes mellitus. Translational Medicine Communications, 6(1), 17.
15. Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A., & Stiglic, G. (2020). Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Scientific reports, 10(1), 11981.
16. Fazakis, N., Kocsis, O., Dritsas, E., Alexiou, S., Fakotakis, N., & Moustakas, K. (2021). Machine learning tools for long-term type 2 diabetes risk prediction. ieee Access, 9, 103737-103757.
17. Khalifa, M., & Albadawy, M. (2024). Artificial intelligence for diabetes: Enhancing prevention, diagnosis, and effective management. Computer methods and programs in biomedicine update, 5, 100141.

18. Farran, B., Channanath, A. M., Behbehani, K., & Thanaraj, T. A. (2013). Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait—a cohort study. BMJ open, 3(5), e002457.

19. Kaur, H., & Kumari, V. (2022). Predictive modelling and analytics for diabetes using a machine learning approach. Applied computing and informatics, 18(1-2), 90-100.

20. Kengne, A. P., Masconi, K., Mbanya, V. N., Lekoubou, A., Echouffo-Tcheugui, J. B., & Matsha, T. E. (2014). Risk predictive modelling for diabetes and cardiovascular disease. Critical reviews in clinical laboratory sciences, 51(1), 1-12.

21. Lagani, V., Koumakis, L., Chiarugi, F., Lakasing, E., & Tsamardinos, I. (2013). A systematic review of predictive risk models for diabetes complications based on large scale clinical studies. Journal of Diabetes and its Complications, 27(4), 407-413.

22. Lai, H., Huang, H., Keshavjee, K., Guergachi, A., & Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. BMC endocrine disorders, 19(1), 101.

23. Rigla, M., García-Sáez, G., Pons, B., & Hernando, M. E. (2018). Artificial intelligence methodologies and their application to diabetes. Journal of diabetes science and technology, 12(2), 303-310.

24. Zhang, L., Wang, Y., Niu, M., Wang, C., & Wang, Z. (2020). Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: The Henan Rural Cohort Study. Scientific reports, 10(1), 4406.

25. Contreras, I., & Vehi, J. (2018). Artificial intelligence for diabetes management and decision support: literature review. Journal of medical Internet research, 20(5), e10775.

26. Chang, V., Ganatra, M. A., Hall, K., Golightly, L., & Xu, Q. A. (2022). An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators. Healthcare Analytics, 2, 100118.

27. Rau, H. H., Hsu, C. Y., Lin, Y. A., Atique, S., Fuad, A., Wei, L. M., & Hsu, M. H. (2016). Development of a web-based liver cancer prediction model for type II diabetes patients by using an artificial neural network. Computer methods and programs in biomedicine, 125, 58-65.

28. Lu, H., Uddin, S., Hajati, F., Moni, M. A., & Khushi, M. (2022). A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. Applied Intelligence, 52(3), 2411-2422.

29. Nguyen, B. P., Pham, H. N., Tran, H., Nghiem, N., Nguyen, Q. H., Do, T. T., ... & Simpson, C. R. (2019). Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. Computer methods and programs in biomedicine, 182, 105055.

30. Deberneh, H. M., & Kim, I. (2021). Prediction of type 2 diabetes based on machine learning algorithm. International journal of environmental research and public health, 18(6), 3317.