*International Journal of Modern Computation, Information, and Communication Technology*
*2025;8(1):26-32.*
***ISSN: 2581-5954***

# Analyzing Diabetes Risk Factors Using Support Vector Machine Algorithm

M. Moorthy[1*], B. Prasad[2]

[1]*Master of Computer Applications, Muthayammal Engineering College (Autonomous), Namakkal, Tamil Nadu, India.*
[2]*Department of Electronics and Communication Engineering, Sona College of Technology, Salem, Tamil Nadu, India.*

[*]*Corresponding author: marirenga2004@gmail.com*

**Abstract.** Diabetes mellitus is a persistent metabolic condition that presents considerable health issues globally. Active prediction and risk assessment are essential for mitigating problems and enhancing patient outcomes. This research examines diabetes risk variables with the Support Vector Machine (SVM) algorithm, an effective supervised machine learning method appropriate for classification applications. Diverse clinical and lifestyle characteristics, such as age, body mass index (BMI), glucose levels, blood pressure, and familial history, are used as input features. The SVM model is developed and evaluated to categorize people as diabetic or non-diabetic with high precision. Performance assessment is conducted with criteria accuracy, recall, and F1-score. The results indicate that SVM proficiently detects critical risk variables and delivers dependable predictions, underscoring its potential as a decision-support instrument in healthcare for the early identification of diabetes.

**Keywords:** Diabetes Mellitus, Support Vector Machine (SVM), Machine Learning, Classification, Prediction Model, Healthcare Analytics.

## INTRODUCTION

Diabetes mellitus is a very common chronic condition impacting millions worldwide. It is characterized by increased blood glucose levels due to the body's inability to manufacture or use insulin properly. The rising incidence of diabetes presents a significant worldwide health issue, resulting in severe consequences, including cardiovascular disease, renal failure, neuropathy, and eye impairment. The World Health Organization (WHO) anticipates that diabetes will rank among the leading causes of death in the next decades, underscoring the need for early detection and appropriate care. Diabetic risk factors include both hereditary and behavioral components. Common factors include age, gender, familial history, obesity, sedentary lifestyle, poor nutrition, hypertension, and abnormal glucose levels, which substantially affect an individual's probability of acquiring the disorder. Identifying and analyzing these risk variables is essential for developing preventative strategies and facilitating rapid treatments. Traditional diagnostic methods often depend on invasive procedures and may inadequately forecast hazards in time to avert disease development. This problem requires the use of sophisticated computational methods to reveal hidden patterns in medical information and enhance forecasting precision.

Machine learning (ML) has become a formidable instrument in healthcare, providing novel solutions for illness diagnosis, prognosis, and risk evaluation. Among the several learning methods, SVM has shown significant efficacy in classification tasks. SVM operates by transforming input characteristics into a higher-dimensional space and establishing an optimum hyperplane that differentiates data points into distinct categories. Its capacity to manage nonlinear interactions and high-dimensional datasets make it particularly effective for the analysis of complex medical data, such as diabetes risk prediction. The use of SVM in diabetes research centers on categorizing patients according to clinical and lifestyle variables to ascertain their classification as diabetic or non-diabetic. Through the training of the algorithm using empirical datasets, the model acquires substantial patterns linked to diabetes risk variables. The assessment of performance using measures like accuracy, precision, recall, and F1-score offers insights into the dependability of predictions. Moreover, SVM elucidates the most significant aspects leading to diabetes, assisting healthcare providers in making educated judgments. Consequently, using SVM for the analysis of diabetes risk variables improves predictive accuracy and aids in preventive healthcare initiatives. This method may alleviate the burden of diabetes by facilitating early diagnosis, directing lifestyle changes, and endorsing evidence-based treatments.

*Received: 13.02.2025 Revised: 02.04.2025 Accepted: 17.04.2025*

## LITERATURE SURVEY

The prediction of diabetes in children is increasingly reliant on early diagnosis informed by family genetic information. The use of sophisticated ML techniques in the examination of hereditary features and markers allows the creation of a highly cohesive predictive model grounded on the genetic impact of the parents. Five machine learning models—Naive Bayes, Random Forest, K-Nearest Neighbors (KNN), SVM, and Logistic Regression—are then evaluated for accuracy and dependability over 5, 10, and 15 epochs in the analysis of possible models [1]. A diabetes risk prediction methodology for women using an enhanced ANN model is presented, derived from the Pima Indians Diabetes dataset, to augment prediction accuracy and serve as a reference for clinical diagnosis. This research optimizes the classic ANN model by modifying hyperparameters, including the number of layers, number of neurons, activation function, and dropout rate, while using L2 regularization and an early stopping mechanism to mitigate overfitting [2]. This study uses random forest models to enhance the identification and prediction of diabetes. The dataset, including age, gender, polydipsia, and other features of 520 diabetic patients, obtained from the AIMCO website, has been partitioned into a training set and a test set in a 7:3 ratio. The random forest model was developed using the training set on the Spark big data computing platform, verified using the test set, and its predictive performance was assessed by several measures. Diabetes Mellitus (DM) is a huge worldwide health challenge, affecting millions and profoundly influencing quality of life. Timely identification and risk assessment are essential; yet traditional diagnostic techniques often incur high costs and are constrained by the accessibility of clinical indicators [4].

This paper tackles these problems by presenting an Optuna optimization framework for diabetes prediction and classification. Optuna is an effective hyperparameter optimization tool that systematically investigates the parameter space to determine optimum configurations for enhanced model performance [5]. Currently, diabetes mellitus is a global health issue that is increasingly prevalent. Consequently, the need for improved early detection approaches, especially in Machine Learning, is paramount. This article employs six machine learning algorithms to predict the probability of diabetes across three stages: prediabetes, diabetes, and absence of diabetes. Three metrics [6]. This research discusses the categorization and prediction of diabetes using a machine learning methodology. The proposed approach enhances the accuracy of diabetes prediction findings using ANN deep learning techniques. This thus enhances the overall classification results. ML methods are now successful for disease diagnosis [7]. Graph convolutional networks (GCNs), inspired by convolutional neural networks, have been used in several applications. However, the incorporation of GCNs into clinical environments, especially for complex health issues such as diabetes, remains far. This research presents a genome-specific graph convolutional network (Geno-GCN) using a multi-graph aggregator to predict the likelihood of Type 2 diabetes development from whole genome sequencing data. Geno-GCN integrates both positive and negative effects derived from graphs based on diabetes risk variables. [8].

This research focuses on the creation and testing of personalized, interpretable multilevel prediction models for hypoglycemia risk assessments in Type 1 Diabetes Mellitus (T1DM). The proposed hybrid methodology integrates mathematical models (MM) with diverse tree-based ML approaches and employs the Trees HAP interpretability method. A mathematical model for the activation of the insulin signaling pathway is used to elucidate the multifaceted processes of glucose-insulin dynamics. Multiple solutions for mitigating class inequality are examined [9]. The study investigates the use of ANN for forecasting diabetes risk, emphasizing various configurations of neuron units inside a single hidden layer. Utilizing the extensive dataset from the Behavioral Risk Factor Surveillance System (BRFSS), the work assesses ANN models with 5, 10, 15, and 20 neuron units, seeking to determine the most effective configuration for precise diabetes risk prediction [10].

Diabetes, classified as a chronic noncommunicable disease, has grown more frequently, affecting around 425 million adults globally as of 2020. The increasing worldwide incidence highlights the urgent need for extensive research endeavors to tackle this major health issue. Undiagnosed diabetes presents significant dangers, including severe consequences such as myocardial stroke and diabetic nephropathy, underscoring the vital need for early identification and care [11]. Diabetes mellitus presents a considerable global health concern, underscoring the need for early detection and care. This material presents an innovative method for diabetes prediction in healthcare using Convolutional Neural Network (CNN) algorithms [12].

Diabetes is a significant worldwide health issue, necessitating efficient prediction strategies to mitigate its widespread effects. This paper does a comprehensive comparison analysis of three notable machine learning approaches: Gaussian Naive Bayes, Decision Tree, and ANN for predicting diabetes risk [13]. An Adaptive

Weighted Decision Forest (AWDF) algorithm is presented to analyze human vital condition data and examination data, facilitating the analysis and prediction of diabetes and its consequences, while guiding diabetic patients in preventive and preemptive measures. The findings indicate that the accuracy and recall of the AWDF approach presented in this research surpass those of the individual machine learning algorithm model [14]. This project aims to create a prediction model for diabetes using risk factors and related disorders with ensemble machine learning techniques. This study aims to improve public health and implement appropriate measures. The study underscores the need for prompt identification and forecasting of diabetes to avert complications and enhance public health [15]. Artificial Intelligence (AI) and ML provide a substantial benefit in the early prediction of diabetes, allowing proactive healthcare measures that may reduce the likelihood of complications. This paper examines several AI-based methodologies for diabetes prediction, emphasizing a project that used many machine learning models, including SVM, Gradient Boosting, K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression [16].

The work advances the area of predictive healthcare by demonstrating the viability and efficacy of smart illness prediction algorithms on a web-based platform. The model can provide tailored reports with remarkable memory, accuracy, and a confusion matrix by enabling patients to enter their symptoms using the application's interface [17]. This research investigates the use of several machine learning classifiers to predict the development of diabetes and to determine the best successful model. A complete dataset including clinical and demographic attributes is used to train and assess classifiers, including Support Vector Machines, Random Forest, Decision Trees, Naive Bayes, and k-Nearest Neighbors. The research employs feature engineering and selection methods to improve model efficacy and decrease dimensionality [18].Diabetes is a chronic illness that results in fatalities globally and exacerbates if not identified in the early stages. Consequently, it is important to identify the illness in its first stages prior to inflicting harm on the human body. At this juncture, the detection of diabetes is a critical operation that must be executed with efficiency. Machine learning algorithms may be used to predict diabetes risk when provided with an appropriate dataset [19]. A predictive model is essential to mitigate the risk of cardiovascular disease onset and the financial burden associated with treatment expenses. This research has established a cardiovascular disease risk prediction model using individuals from the Endocrine and Cardiology Clinic at Hospital Canselor Tuanku Muhriz (HCTM) [20].

## PROPOSED SYSTEM

The proposed method aims to evaluate diabetes risk variables and forecast the probability of diabetes using the Support Vector Machine (SVM) method. The method uses clinical and lifestyle data to develop a prediction model that classifies people as diabetic or non-diabetic. This technique prioritizes computational intelligence for early risk assessment, contrasting with traditional diagnostic procedures that often depend on intrusive blood tests and physical exams, therefore facilitating preventative healthcare practices. The framework starts with the collection of data from credible sources, including medical libraries, patient records, and health surveys. The dataset generally includes significant parameters such as age, gender, familial history, body mass index (BMI), glucose levels, blood pressure, cholesterol, physical activity, and dietary practices. Due to the presence of missing values, outliers, or inconsistencies in raw medical data, preprocessing methods, such as data cleaning, normalization, and feature scaling, are used. This guarantees the dataset's accuracy, consistency, and appropriateness for training the SVM model. Figure 1 shows the sequential process of the proposed system, beginning with the collection of medical data, followed by data cleansing, feature selection, training and testing of the SVM model, and concluding with diabetes risk prediction.

Feature selection is essential in the proposed system. By pinpointing the most relevant factors affecting diabetes, the approach decreases computational complexity while enhancing model efficacy. Correlation analysis and principal component analysis (PCA) may be used to identify the most relevant traits. Glucose levels and BMI are regarded as major indicators, although lifestyle variables such as exercise and food provide supplementary context for risk assessment. The core of the proposed system is the execution of the SVM algorithm. SVM creates an optimum hyperplane that distinguishes between diabetes and non-diabetic data points. Kernel functions, including linear, polynomial, and radial basis function (RBF), are used to address both linear and nonlinear interactions in the dataset. Through the training of the model using a labelled dataset, the system acquires patterns that differentiate high-risk people from low-risk ones.

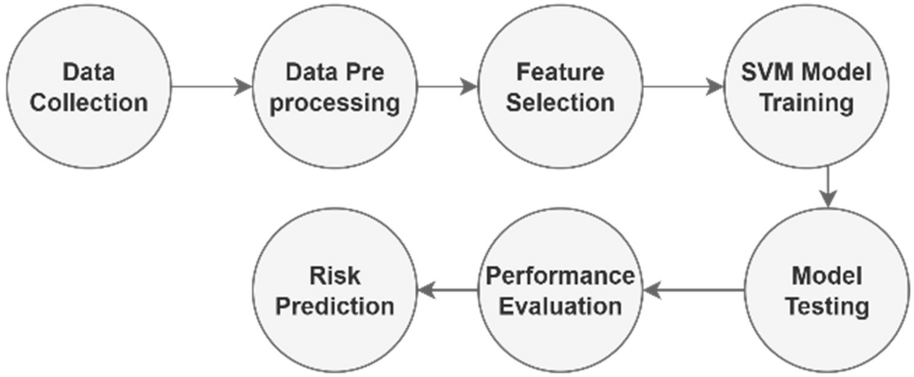


**FIGURE 1.** Workflow of the Proposed Diabetes Risk Prediction System

In the testing phase, the trained model assesses novel, unobserved data to accurately forecast diabetes risk.To evaluate the efficacy of the proposed system, performance measures including accuracy, sensitivity, specificity, precision, recall, and F1-score are calculated. These measures provide a thorough insight into the model's proficiency in accurately classifying diabetic and non-diabetic situations. Moreover, confusion matrix analysis facilitates the identification of misunderstandings and informs subsequent model optimization. The proposed system aims to provide an efficient, reliable, and scalable solution for diabetes prediction. Advanced technology amalgamates clinical expertise with machine learning abilities, assisting healthcare professionals in decision-making, facilitating early identification, and eventually alleviating the worldwide burden of diabetes.Table 1 highlights the components of the proposed system, including data collection and preprocessing, SVM model training, assessment, and risk prediction, highlighting the contribution of each stage to the accuracy of prediction.

**TABLE I.** Functional Modules of the Proposed Diabetes Risk Prediction System

| Module | Description |
|---|---|
| Data Collection | Collects medical and lifestyle data such as glucose, BMI, and family history. |
| Preprocessing | Cleans, normalizes, and scales the raw dataset for accurate analysis. |
| Feature Selection | Identifies significant risk factors like glucose and BMI to reduce complexity. |
| SVM Model Training | Builds a classification model using a labeled training dataset. |
| Model Evaluation | Test performance with accuracy, precision, recall, and F1-score. |
| Risk Prediction | Classifies patients as diabetic or non-diabetic based on a trained model. |

# RESULTS AND DISCUSSIONS

Diabetes is a disorder marked [21] by symptoms such as polyuria, polydipsia, abrupt weight loss, fatigue, polyphagia, vaginal thrush, blurred vision, pruritus, irritability, delayed wound healing, partial paresis, muscular rigidity, baldness, and obesity. These symptoms may be classified by age and gender, namely, as male or female. Polyuria is defined by excessive urine, while polydipsia is defined by excessive thirst. Sudden weight loss indicates unexplained weight reduction. Weakness refers to a general lack of physical strength, while polyphagia denotes an abnormal increase in appetite. Genital thrush is a candida infection that may induce pruritus, tenderness, and several other discomforts in the genital region. Visual blurring may indicate a specific illness, while itching is a generic symptom. Irritation constitutes an additional symptom. Prolonged wound healing is an additional sign. Partial paresis refers to a limited lack of voluntary motor function. Muscle rigidity is an additional symptom. Alopecia refers to the loss of hair. Obesity is a significant risk factor for diabetes.

Table 2 defines the distribution of a dataset including 520 samples, allocated 70% for training (364 samples) and 30% for testing (156 samples), so assuring a balanced representation of 320 positive and 200 negative instances.

**TABLE II.** Dataset Distribution for Training and Testing

| Category | Total Samples | Training Set (70%) | Testing Set (30%) |
|---|---|---|---|
| Positive | 320 | 224 | 96 |
| Negative | 200 | 140 | 60 |
| **Total** | **520** | **364** | **156** |

Table 3 shows that the SVM model accurately classified both positives and negatives, showing minimum errors and demonstrating robust classification performance.

**TABLE III.** Classification Results of the SVM Model Using the Confusion Matrix

| Actual / Predicted Class | Predicted Positive | Predicted Negative |
|---|---|---|
| Actual Positive | 95 | 1 |
| Actual Negative | 2 | 58 |

Figure 2 illustrates the performance metrics of the SVM model, which demonstrate the reliability and robustness of diabetes risk prediction. The model's high accuracy, precision, recall, and F1-score are verified.

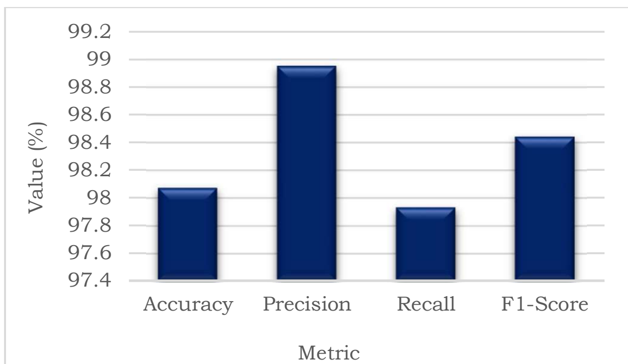


**FIGURE 2.** Performance Metrics of Diabetes Risk Prediction using SVM

Figure 3 presents a comparative analysis of model performance. The proposed SVM achieves higher accuracy, sensitivity, and specificity, markedly exceeding the performance of Logistic Regression, KNN, and Random Forest [22] in diabetes risk prediction and classification efficacy.

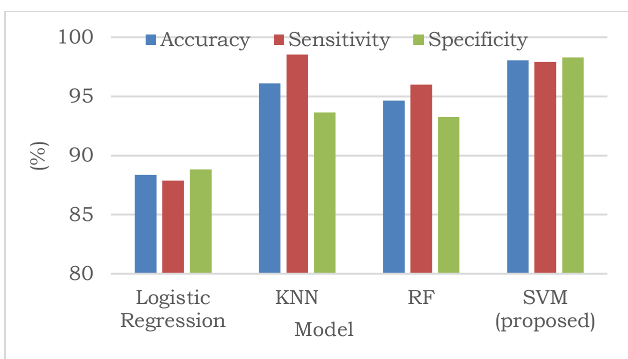


**FIGURE 3.** Quantitative Performance Comparison of Logistic Regression, KNN, Random Forest, and SVM

The proposed SVM-based diabetes risk prediction system has considerable accuracy and dependability; yet significant limitations persist. The model's efficacy is significantly influenced by the quality of the dataset; unbalanced or missing data might diminish predicted accuracy. Moreover, the system mostly depends on clinical and lifestyle characteristics, neglecting genetic, environmental, or socio-economic variables that significantly affect diabetes risk. Future research may concentrate on integrating extensive, multi-s datasets, including genomic and real-time IoT sensor data, to enhance generalization. Advanced ensemble learning or deep learning models may be investigated to improve resilience. Furthermore, creating intuitive interfaces and incorporating the model into healthcare apps will facilitate broader use in clinical practice for the early diagnosis and monitoring of diabetes.

## CONCLUSION

Diabetes is a persistent and escalating worldwide health issue that requires efficient early detection technologies to mitigate complications and enhance patient care. This work presented a support vector machine-based approach for predicting diabetes risk by analyzing clinical and lifestyle parameters to accurately classify people as diabetic or non-diabetic. Using preliminary processing, feature selection, and rigorous model training, the system exhibited significant reliability in detecting essential risk indicators, including glucose levels, BMI, and insulin concentration. The results of experiments showed that the SVM model attained enhanced accuracy, sensitivity, and specificity relative to conventional machine learning techniques, such as Logistic Regression, KNN, and Random Forest. The confusion matrix and performance indicators validated its robust classification capability, exhibiting few misclassifications. These results underscore the practical significance of the proposed approach in healthcare, where precise predictions are crucial for prompt diagnosis and preventative measures. The proposed SVM system provides a dependable, scalable, and interpretable solution for diabetes prediction. Using machine learning improves clinical decision-making and facilitates early detection measures. With enhancements via bigger datasets, IoT connections, and sophisticated learning techniques, the system has the potential to serve as an excellent decision-support tool in practical healthcare applications.

## REFERENCES

[1]. R. Samsani, K.S. Akash, and M. L. Pai, 2025, “A Genetic-Based framework for diabetes risk prediction using parental genetic contributions,” *5th International Conference on Pervasive Computing and Social Networking*, pp. 634-639.

[2]. W. Meng, Y. Zhao, X. Li, X. Yu, X. Jiao, F. Shi, and P. Qin, 2025, “Prediction of diabetes risk in women based on improved ANN modeling,” *5th International Conference on Neural Networks, Information and Communication Engineering*, pp. 1871-1875.

[3]. Z. Dai, Z. Hu, T. Shen, and Y. Zhang, 2023, “Risk prediction of diabetes based on spark and random forest algorithm,” *IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms*, pp. 535-539.

[4]. A. R. Kandula, V. V. Samhitha, V. Harshitha, R. N. Sindhuri, and S. A. Nagavalli, 2024, “Machine learning based screening for diabetes risk prediction,” *3rd International Conference on Automation, Computing and Renewable Systems*, pp. 1188-1193.

[5]. S. Kumari, V.C. Ranganayaki, K. Sathya Moorthy, K. Selvi, T.A. Mohanaprakash, and C. Tamilselvi, 2024, “Optuna-optimized machine learning technique for accurate diabetes prediction and classification,” *4th International Conference on Sustainable Expert Systems*, pp. 1478-1485.

[6]. H. K. Thakur, D. P. Sahu, R. K. Chaturvedi, G. K. Jha, and D. Kumari, 2025, “Predicting diabetes risk: A comparative analysis of machine learning algorithms,” *First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies*, pp. 431-434.

[7]. P. Singh, S. Silakari, and S. Agrawal, 2023, “An efficient deep learning technique for diabetes classification and prediction based on Indian diabetes dataset,” *3rd International Conference on Technological Advancements in Computational Sciences*, pp. 487-491.

[8]. K. Guo, H. Lin, M. Grosser, G. Zhang, and J. Lu, 2024, “Geno-GCN: a genome-specific graph convolutional network for diabetes prediction,” *46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1-4.

[9]. M. Athanasiou, N. Myropoulou, K. Zarkogianni, and K. S. Nikita, 2024, “Development of personalized interpretable multilevel prediction models for the risk assessment of hypoglycemia in Type 1 Diabetes,” *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1-5.

[10]. S. Y. Prasetyo, K. E. Setiawan, and H. A. Shiddiqi, 2024, “Assessing the efficacy of artificial neural networks for diabetes risk prediction,” *2nd International Symposium on Information Technology and Digital Innovation*, pp. 108-112.

[11]. S. Y. Prasetyo, Santy, and R. Yunanda, 2024, “Diabetes risk prediction exploration: uncovering patterns

and enhancing predictive accuracy through ensemble learning," *4th International Conference of Science and Information Technology in Smart Administration*, pp. 213-218.
[12]. A. Shahi, A. Wahi, and A. K. Dwivedi, 2024, "Diabetes prediction in healthcare using cnn algorithm," *International Conference on IoT, Communication and Automation Technology*, pp. 372-376.
[13]. S. Y. Prasetyo, Z. N. Izdihar, and G. Z. Nabiilah, 2024, "Analyzing machine learning approaches for diabetes risk prediction: Comparative performance assessment using BRFSS data," *7th International Conference on Informatics and Computational Sciences*, pp. 324-329.
[14]. R. Han, 2023, "A study of diabetes prediction based on adaptive weighted decision forest," *8th International Conference on Cloud Computing and Big Data Analytics*, pp. 203-207.
[15]. A. A. Endalew, A. Agegne Engda, A. O. Salau, and G. Alemu Baye, 2024, "Diabetes prediction based on risk factors and associated diseases using ensemble machine learning," *IEEE 5th International Conference on Electro-Computing Technologies for Humanity*, pp. 1-5.
[16]. D. Sehgal, I. Kaur, V. Sharma, B. Gautam, A. Singh, and N. Kumar, 2024, "AI-driven early diabetes prediction," *IEEE 3rd World Conference on Applied Intelligence and Computing*, pp. 40-48.
[17]. A. Rana, and P. Kaur, 2024, "Implementation of web-based intelligent system for diabetes prediction using machine learning," *2nd International Conference on Emerging Trends in Engineering and Medical Sciences*, pp. 462-467.
[18]. M. K. Daigavhane, and S. Gundewar, 2024, "Diabetes prediction using different machine learning classifiers," *Parul International Conference on Engineering and Technology*, pp. 1-6.
[19]. N. Gürler, E. Arslan, I. Shayea, and A. Mukhamedrakhimova, 2024, "Diabetes risk prediction with different machine learning algorithms," *IEEE 4th International Conference on Smart Information Systems and Technologies*, pp. 403-407.
[20]. O.T. Kee, H. Harun, N. Mustafa, N.A.A. Murad, S.F. Chin, R. Jaafar, N. Abdullah, 2023, "Cardiovascular disease risk in type 2 diabetes mellitus prediction model development using machine learning,'' *IEEE International Conference on Artificial Intelligence in Engineering and Technology*, pp. 61-65.
[21]. https://www.kaggle.com/datasets/rcratos/diabetes-risk-prediction
[22]. S. Afolabi, N. Ajadi, A. Jimoh, and I. Adenekan, 2025, "Predicting diabetes using supervised machine learning algorithms on E-health records," *Informatics and Health*, 2(1), pp. 9-16.