



DEVELOPMENT OF A PREDICTIVE MODEL FOR UNCONTROLLED BLOOD SUGAR RISK IN TYPE 2 DIABETES MELLITUS PATIENTS: A LITERATURE REVIEW 2020–2025

Andi Amanda Amalia Agus Arifin¹, Sitti Raodhah², & Dwi Santy Damayati³

^{1,2,3}Universitas Islam Negeri Alauddin Makassar, Indonesia

Correspondence Email: amaliahamanda@gmail.com

ABSTRACT

Type 2 diabetes mellitus represents a chronic condition with continuously rising global prevalence and serves as a leading contributor to morbidity and mortality worldwide. Managing this disease presents substantial challenges, particularly due to the significant number of patients experiencing inadequate glycemic control, which elevates the likelihood of developing cardiovascular complications, nephropathy, neuropathy, and retinopathy. Recent advancements in analytical technology and the emergence of large-scale data availability have enabled numerous studies conducted between 2020 and 2025 to demonstrate that machine learning methodologies can identify intricate risk patterns with superior accuracy compared to traditional statistical approaches. This study provides a comprehensive examination of glycemic control risk prediction model development, key contributing factors, and the performance characteristics of various algorithms including logistic regression, random forest, gradient boosting, neural networks, and ensemble learning methods, based on an extensive literature review. Research findings indicate that clinical parameters such as HbA1c, glucose levels, BMI, and blood pressure, alongside behavioral factors including physical activity, dietary patterns, and treatment adherence, as well as sociodemographic characteristics, significantly influence the risk of poor glycemic control. Prediction models developed across multiple studies demonstrate robust performance metrics and hold considerable promise for implementation in primary healthcare settings, particularly within developing nations where laboratory resources are constrained. This investigation underscores the critical need for developing accurate, adaptable, and easily implementable prediction models to facilitate early detection and enable more precisely targeted intervention strategies in type 2 diabetes mellitus management.

Keywords: Type 2 diabetes mellitus; Glycemic; Machine learning; Risk Prediction Model; HbA1c

1. INTRODUCTION

Type 2 diabetes mellitus is a chronic disease with a growing global health burden that has a significant impact on mortality and morbidity. In 2017, an estimated 462 million people were living with type 2 diabetes, equivalent to 6.28 percent of the world's population, and this number has continued to grow in recent years (Khan et al., 2019). A global report in 2021 reaffirmed that 529 million people live with diabetes, and the prevalence has increased in almost all age groups (Ong et al., 2023). More than 90 percent of diabetes cases are type 2 diabetes mellitus, which is closely related to lifestyle factors, obesity, and predisposition (Hossain et al., 2024). Changes in people's lifestyles, rapid urbanization, and increasing obesity are the main factors driving the increase in the number of people vulnerable to diabetes. The World Health Organization emphasizes that controlling blood sugar levels is crucial in preventing potentially fatal long-term complications in patients with type 2 diabetes mellitus (WHO, 2024).

Uncontrolled glycemic control remains a major challenge in the management of type 2 diabetes mellitus. Hemoglobin A1c is the primary indicator for assessing long-term blood sugar control, and research consistently shows a strong relationship between HbA1c values and the risk of microangiopathy and macroangiopathy complications. Achieving an HbA1c target below seven percent has been shown to significantly reduce the risk of complications (ElSayed et al., 2023). Conversely, HbA1c levels of nine percent or higher are categorized as poor glycemic control, increasing the risk of severe complications (Sherwani et al., 2016). Even a one-percent increase in HbA1c is associated with a thirty-percent increase in all-cause mortality and a forty-percent increase in cardiovascular mortality (Dubale et al., 2023). Although clinical guidelines have established clear therapeutic targets, many patients continue to experience poor glycemic control, contributing to increased incidence of cardiovascular disease, kidney failure, neuropathy, and retinopathy (IDF, 2025).

This situation highlights the need for more effective approaches to predicting the risk of uncontrolled blood sugar so that interventions can be provided early and more precisely. Advances in analytical technology in recent years, particularly the development of machine learning and large-scale data processing, have opened up new opportunities to develop more accurate prediction models capable of capturing complex patterns in patient data. A review by Maciag et al. (2021) shows that machine learning can identify risk factors that influence blood glucose fluctuations, including diet, weight status, laboratory parameters, and daily lifestyle.

Recent developments in glucose monitoring technology have also shown promising results in diabetes management. Research by Reddy et al. (2023) demonstrates that continuous glucose monitoring (CGM) systems and digital biomarkers from wearable devices can provide superior outcomes compared to routine self-monitoring, particularly when these systems deliver immediate feedback to patients and decision support tools for healthcare providers. However, these technological advances present significant implementation challenges in resource-limited settings. The CGM-based approaches described by Reddy et al. (2023) primarily focus on real-time glucose monitoring in controlled clinical environments, requiring expensive equipment, continuous device adherence, and robust technical infrastructure that may not be readily available in developing countries. Furthermore, their emphasis on real-time monitoring differs fundamentally from predictive risk stratification approaches that can identify patients at risk of poor glycemic control before complications arise. This gap presents a critical opportunity to develop alternative prediction models that utilize more accessible clinical, behavioral, and demographic variables without relying on costly continuous monitoring devices. Such models show great potential in personalizing glycemic interventions based on individual risk profiles, enabling diabetes control to be carried out in a more adaptive and responsive manner across

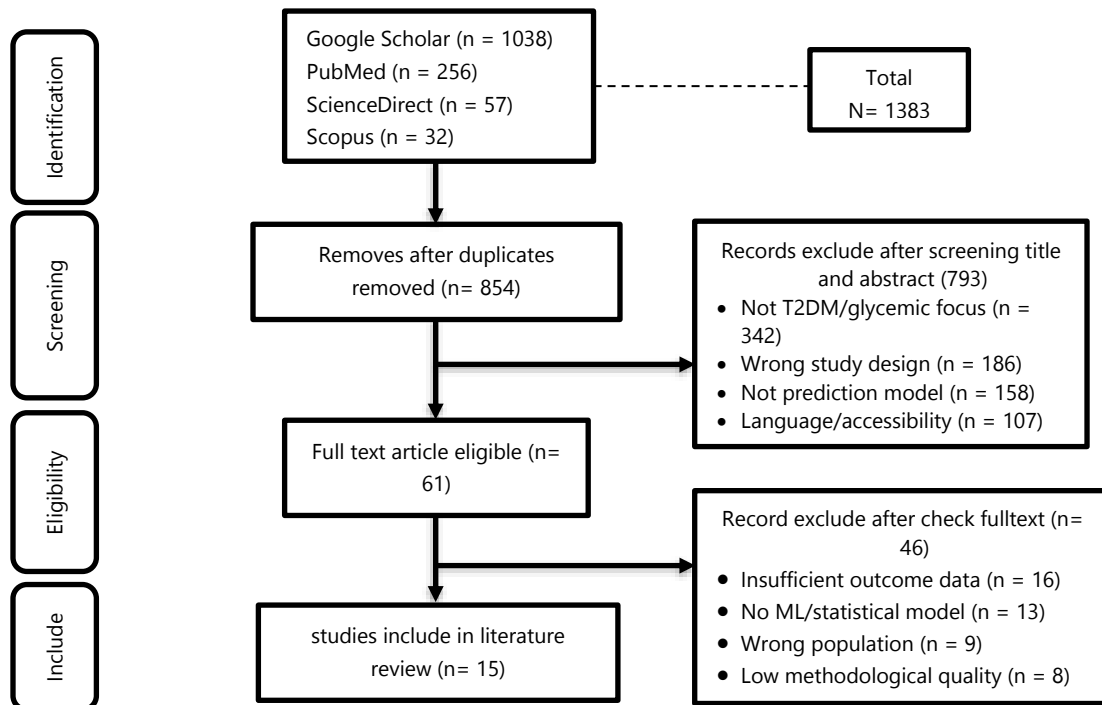
diverse healthcare contexts, particularly in primary care settings where advanced monitoring technologies remain limited.

Social and behavioral factors have also been shown to play an important role in determining the success of glycemic control. Alfageeh et al. (2024) reported that medication adherence, socioeconomic status, stress levels, and physical activity affect patients' ability to achieve good blood sugar control. Predictive models that integrate clinical, social, and behavioral factors show better performance than models that only use laboratory data. These results emphasize the need for a comprehensive approach in designing glycemic control risk prediction models.

In Indonesia and other developing countries, the need for glycemic control risk prediction models is becoming increasingly urgent. Limited healthcare personnel, suboptimal distribution of laboratory facilities, and the increasing burden of diabetes make early detection of risk crucial in primary healthcare services. Predictive models that are easy to implement and do not rely on expensive laboratory tests can serve as strategic tools to facilitate rapid intervention and prevent long-term complications.

Based on these developments, this study aims to examine the latest predictor variables and analytical models used to predict the risk of uncontrolled blood sugar control in patients with type 2 diabetes mellitus. The study was conducted by reviewing scientific literature published between 2020 and 2025 in order to provide an up-to-date overview of advances in glycemic prediction technology. This study also aims to identify existing research gaps and formulate directions for developing more accurate, adaptive, and relevant prediction models for the primary healthcare context. With a more comprehensive understanding of risk prediction models, this study is expected to make a real contribution to improving type 2 diabetes mellitus management strategies and supporting efforts to improve patients' quality of life.

2. METHODS



Picture 1. PRISMA diagram *Literature Review*

This study employs a narrative literature review approach following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines to systematically identify and synthesize scientific findings regarding risk factors and prediction models for uncontrolled blood sugar in type 2 diabetes mellitus patients. Literature searches were conducted in November 2025 across four major databases: Google Scholar (n = 1,038), PubMed (n = 256), ScienceDirect (n = 57), and Scopus (n = 32), using keyword combinations including "type 2 diabetes mellitus", "poor glycemic control", "uncontrolled blood glucose", "risk prediction model", "machine learning", and "HbA1c". The publication timeframe was restricted to 2020-2025 to capture the latest developments in diabetes management and predictive modeling technology. The initial search yielded 1,383 records, which after removing 529 duplicates, resulted in 854 unique records for screening.

Title and abstract screening was conducted independently by two reviewers, excluding studies that did not focus on type 2 diabetes mellitus or glycemic control (n = 342), employed inappropriate study designs such as animal studies or purely qualitative research (n = 186), lacked prediction model development or risk analysis (n = 158), and had language or accessibility barriers (n = 107). Sixty-one full-text articles underwent detailed assessment against inclusion criteria: studies investigating type 2 diabetes patients, use of standard glycemic parameters (HbA1c or fasting glucose), reporting of risk factors or predictor variables, prediction model development using statistical or machine learning approaches, and English-language full-text availability. Forty-six articles were subsequently excluded due to insufficient outcome data (n = 16), absence of prediction modeling approaches (n = 13), wrong population characteristics (n = 9), and low methodological quality (n = 8). Ultimately, 15 studies meeting all criteria were included for comprehensive review, as illustrated in the PRISMA flow diagram (Figure 1).

Data extraction was performed systematically using standardized forms, capturing study characteristics, participant demographics, predictor variables (clinical, behavioral, demographic, laboratory), prediction algorithms (logistic regression, random forest, gradient boosting, neural networks, ensemble methods), and performance metrics (accuracy, sensitivity, specificity, AUC). Variables were categorized into clinical factors (HbA1c, glucose, blood pressure, BMI, comorbidities), demographic characteristics (age, gender, socioeconomic status), behavioral factors (physical activity, diet, medication adherence), and laboratory parameters (lipids, creatinine, electrolytes). Narrative synthesis was employed to integrate findings across studies, evaluating each variable based on strength of association with poor glycemic control, consistency across multiple studies, clinical relevance, and applicability in resource-limited primary healthcare settings. Quality assessment examined methodological rigor, sample size adequacy, statistical appropriateness, and potential bias, with findings compared across different study designs and geographic contexts to ensure objectivity and comprehensive interpretation.

3. RESULTS AND DISCUSSION

Findings from a literature review of various studies that developed risk prediction models for diabetes and blood sugar control in patients with type 2 diabetes mellitus using statistical approaches and machine learning algorithms. All studies analyzed included a variety of designs, methods, sample sizes, population characteristics, and predictive algorithms, enabling comprehensive mapping of key predictor variables, model performance, development challenges, and implications for efforts to build a risk prediction model for uncontrolled blood sugar control that is appropriate for the Indonesian context. The analysis was conducted thematically by

grouping findings based on model performance strength, predictor variable contribution, research gaps, and knowledge integration in risk prediction model development.

No.	Title	Authors	Year	Methods	Result
1.	Prognostic Nomogram for Uncontrolled Type 2 Diabetes Using Thailand Nation-wide Cross-sectional Studies	Teeraboon Lertwanich-wattana, Picha Suwannahi-tatorn, Mathirut Mungthin, and Ram Rangsin	2024	cross-sectional with a predictive modeling and validation approach.	In 2018, 24% of the 38,568 participants in the development group had uncontrolled type 2 DM (defined as HbA1c $\geq 9\%$). The predictive nomogram for uncontrolled diabetes consisted of demographic characteristics, prescription medications, history of diabetes complications, and laboratory results (C-statistic 0.77). The goodness-of-fit test and DCA showed good agreement between the results and clinical application for type 2 DM.
2.	Application of Machine Learning Algorithms to Predict Uncontrolled Diabetes Using the All of Us Research Program Data	Tadesse M Abegaz, Muk-tar Ahmed, Fati-mah Sherbeny, Vakaramoko Diaby, Hongmei Chi, and Askal Ayalew Ali,	2023	Machine Learning (ML) / Predictive Modeling Study	The results show that the random forest model provides the best performance with an accuracy of 0.80 (95% CI: 0.79–0.81) and an area under the receiver operating characteristic curve (AUROC) of 0.77, higher than extreme gradient boost (accuracy 0.74), weighted ensemble model (accuracy 0.77), and logistic regression (accuracy 0.64 with the lowest AUROC of 0.7). The most important predictor variables in predicting uncontrolled diabetes are potassium levels, body weight, aspartate aminotransferase, height, and heart rate. This study concludes that machine learning techniques, particularly random forest, can be used effectively to predict uncontrolled diabetes by utilizing clinical characteristics such as serum electrolytes and physical measurements of patients.
3.	Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: the Henan	Liying Zhang, Yikang Wang, Miaomiao Niu, Chongjian Wang & Zhenfei Wang	2020	Cohort study with a machine learning predictive modeling approach	The results showed that all models had strong predictive performance, with AUC values ranging from 0.811 to 0.872 when using laboratory data and 0.767 to 0.817 without laboratory data, with the GBM model showing the best performance (AUC 0.872 with laboratory data and 0.817 without laboratory data). Model

	Rural Cohort Study				performance plateaued when using 30 variables, except for the CART model. The ten most important predictor variables identified through the Shapley additive explanations approach included sweet taste, urinary glucose, age, heart rate, creatinine, waist circumference, uric acid, pulse pressure, insulin, and hypertension. This study successfully identified new risk factors such as urine indicators and sweetness preference that were not found in previous risk prediction methods, demonstrating that machine learning methods can provide deeper insights into disease risk factors without a priori causality assumptions.
4.	A Machine-Learning Model Accurately Predicts Projected Blood Glucose	Daniel R. Goldner, Chandra Y. Osborn, Lindsay Sears, Brian Huddleston, & Jeff Dachis	2018	Retrospective Observational Study dengan Machine Learning Predictive Modeling	The evaluation results on the holdout data (September-November 2017) show that the model has high accuracy with a median absolute error of 14.2 mg/dL and a mean absolute error of 21.3 mg/dL, where 91% of predictions are within the range of ±50 mg/dL. This study concludes that a machine learning model with minimal input can accurately predict future blood glucose values in type 2 diabetes patients who do not use insulin, although further empirical investigation is needed to determine whether users will modify their behavior after learning that certain behaviors can result in undesirable blood glucose values.
5.	Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques	Zidian Xie, Olga Nikolayeva, Jiebo Luo, & Dongmei Li	2019	Cross-sectional Study with Machine Learning Comparative Analysis	The results showed that all predictive models achieved high performance with an area under the curve (AUC) value ranging from 0.7182 to 0.7949, where the neural network model had the highest accuracy (82.4%), highest specificity (90.2%), and highest AUC (0.7949), while the decision tree model showed the highest sensitivity (51.6%) for type 2 diabetes. Risk factor analysis identified that individuals who slept 9 hours or more per day had a higher risk of developing type 2 diabetes with an

					adjusted odds ratio (aOR) of 1.13 (95% CI: 1.03-1.25), and individuals with health check-up frequency less than 1 year showed a significantly higher risk with an aOR of 2.31 (95% CI: 1.86-2.85), indicating that excessive sleep duration and low health check-up frequency are important risk factors for type 2 diabetes.
6.	Predictive Model for Early Detection of Type 2 Diabetes Using Patients' Clinical Symptoms, Demographic Features, and Knowledge of Diabetes	Taiwo Adetola Ojurongbe, Habeeb Abiodun Afolabi, Adesola Oyekale, Kehinde Adekunle Bashiru, Olubunmi Ayelagbe, Olusola Ojurongbe, Saddam Akber Abbasi, & Nurudeen A Adegoke	2024	Multivariable penalized logistic regression	As a result, the predictive model developed showed very high accuracy, with an Area Under the Curve (AUC) of 99% on the training data and 94% on the test data. Key findings revealed that increased WHR and AIP were significantly associated with a higher risk of diabetes, while high catalase levels were correlated with a reduced risk. On the other hand, triglyceride (TG) levels did not show a significant association with this disease. In conclusion, this study emphasizes the importance of specific clinical and biochemical markers for early detection, which was then realized in the development of a web application to facilitate the identification of individuals at risk.
7.	Diabetes Risk Prediction Model Based on Community Follow-Up Data Using Machine Learning	Liangjun Jiang, Zhenhua Xia, Ronghui Zhu, Haimei Gong, Jing Wang, Juan Li, & Lei Wang	2023	Follow-up patient records using a Random Forest Classifier-based model.	The results showed that this model achieved an accuracy of 91.24% and an AUC value of 97% on the ROC curve. To improve its application in clinical and real-life settings, a diabetes risk score card was also designed and tested with real data, showing higher accuracy (95.15%) and high model reliability. In conclusion, this community follow-up big data mining-based diabetes risk prediction model can be used for large-scale risk screening by community doctors and also for wearable devices or smart biosensors for patient self-examination.
8.	Prediction of Type 2 Diabetes Mellitus with Logistic Regression Algorithm for	Sepriyan Gunawan, Rini Astuti, Willy Priharto, & Ryan Hamonangan	2025	Quantitative with Logistic Regression algorithm	The results of the study show that Logistic Regression performs well with an accuracy of 82.51%, precision of 88.04%, recall of 89.76%, F1-score of 88.89%, AUC of 83.12%, and RMSE of 0.4182. The main features that influence the

	Early Detection				prediction are age, diet (lack of vegetables/fruit), and diastolic blood pressure. This model is effective for early detection of Type 2 Diabetes Mellitus risk and can be used as a basis for decision making in primary health facilities.
9.	Development of Diabetes Risk Prediction Model Using AdaBoost Approach and SMOTE Oversampling Technique	Sofian Sidiq, Alfian, & Nur Shobi Maburur	2025	AdaBoost and SMOTE	The results show that the AdaBoost model without SMOTE has an accuracy of 81.87% and an ROC-AUC value of 0.9031, but its performance tends to be higher in the majority class than in the minority class. After applying AdaBoost with SMOTE, the model's performance in the minority class improved significantly, with more balanced precision, recall, and F1-score between the two classes. The overall accuracy of the model increased to 82.83%, and the ROC-AUC value to 0.9058. The combination of AdaBoost and SMOTE proved to be effective in handling data imbalance, providing more balanced predictions between the majority and minority classes.
10.	Prediction Modeling of Blood Sugar Levels in Diabetes Patients Using Linear Regression Method	Hanan Hanafi, Ganjar Alfian, Tri Widodo, & Muhammad Syafrudin	2024	Linear regression method	The results of the prediction model in this study were implemented and visualized in a website-based information system. In this system, users can predict future blood sugar levels based on blood sugar levels 30 minutes prior. Users can also view data visualizations of blood sugar level movements based on specific time ranges. This system is expected to help diabetes patients predict future blood sugar levels, enabling them to control their blood sugar levels and avoid poor health conditions in the future.
11.	Prediction of Type I and Type II Diabetes Mellitus Using the KNN Method at the Dharma Husada Clinic	Richa Nanda Fitria, Wahyu Sugianto, & Amalia Cemara Nur'aidha	2024	K-Nearest Neighbor (KNN)	The results show that the KNN method can indeed be applied to classify DM. However, its accuracy is limited, especially in this small dataset and for patients aged 30 years. This study reveals the challenge of selecting the optimal K value; a K value that is too small can cause overfitting, while a K value that is too large can cause

12.	Analysis of the Performance of Random Forest and CatBoost Models with the SMOTE Technique in Diabetes Risk Prediction	Rony Irfannandhy, Lekso Budi Handoko, & Noval Ariyanto	2024	Random Forest (RF) and CatBoost models with SMOTE technique	underfitting, where the choice of K has a significant impact when the amount of data is limited. The results showed that both models had highly competitive accuracy, namely 82% for SMOTE-RF and 81% for SMOTE-CatBoost. Further analysis revealed that glucose, BMI, and age were the main predictors that most influenced the risk of diabetes, with glucose being the most important indicator. Practically, these predictive models have the potential to assist medical personnel in detecting diabetes early and more accurately to prevent more serious complications.
13.	Early Detection Modeling of Diabetes Mellitus Using the Ensemble Learning Approach	Syaikhina Usabili & Uce Indahyanti	2024	Ensemble learning classification method	Modeling for early detection of diabetes mellitus yielded results with ensemble learning classification models and imbalanced data using SMOTE Upsampling in type 2 diabetes mellitus classification. Random Forest based on Confusion Matrix obtained a result of 87.30%. where the value produced by Random Forest is categorized as excellent classification.
14.	Prediction of Diabetes Mellitus Using Machine Learning Algorithms	Egi Safitri, Dani Rofianto, Neni Purwati, Hendra Kurniawana, & Sri Karnila	2024	Three machine learning algorithms: Random Forest, Logistic Regression, and Decision Tree.	The results showed that Logistic Regression achieved the highest accuracy (75%) and balanced performance in identifying positive and negative cases. Decision Tree excelled in recall, while Random Forest showed a slightly lower balance between precision and recall. ROC curve analysis revealed that Random Forest had the highest AUC (0.82), followed by Logistic Regression (0.81) and Decision Tree (0.73). This study confirms that machine learning algorithms can effectively predict diabetes, providing a valuable tool for early detection and intervention, which could ultimately reduce the global burden of diabetes mellitus.
15.	Application of the XGBoost	Erliyan Redy Susanto & Agum Cahyana	2025	XGBoost Algorithm	XGBoost Algorithm As a result, the optimized XGBoost model showed outstanding performance

Algorithm for Diabetes Prediction: Confusion Matrix and ROC Curve Analysis	with an accuracy of 96.33%, precision of 93.4%, recall of 97.16%, F1 score of 95.7%, and ROC-AUC score of 0.99. Confusion matrix analysis confirmed the model's diagnostic superiority with only 6 false negatives, demonstrating its reliable ability to detect positive cases. In conclusion, this model has the potential to be a valuable clinical decision support tool for early diabetes detection, although further efforts to minimize false negatives are still needed for optimal patient safety.
--	--

The study by Lertwanichwattana et al. (2024) is an example of a model that combines a cross-sectional approach with predictive modeling and produces a nomogram with good performance. With a C-statistic of 0.77, this model demonstrates adequate ability to identify patients with type 2 diabetes with uncontrolled blood sugar ($\text{HbA1c} \geq 9\%$). This finding is important because it demonstrates the practical value of integrating demographic variables, prescription drugs, complication history, and laboratory results in formulating a prediction tool that is not only accurate but also easy to use in clinics. The study underscores the importance of selecting predictor variables based on broad population data, as national-scale data allows the model to have better predictive robustness. These results are further supported by the study by Abegaz et al. (2023), which showed that the Random Forest algorithm outperformed other models in predicting uncontrolled diabetes with an accuracy of 0.80 and an AUROC of 0.77. In the context of clinical prediction, these figures are considered good. Interestingly, the study found the role of variables that are rarely explored in glycemic control prediction, such as potassium levels, AST, height, weight, and heart rate. These variables show that blood sugar control is inseparable from broader physiological conditions, including electrolyte function and other organs that indirectly affect glucose homeostasis.

One study that has made a significant contribution to providing a new perspective on diabetes risk prediction is the study by Zhang et al. (2020) conducted on a rural population in China. Using various machine learning algorithms, this study showed that the Gradient Boosting Machine (GBM) model provided the best predictive performance with an AUC of 0.872 when using laboratory data and 0.817 when laboratory data was not used. Shapley Additive Explanations (SHAP) analysis in this study identified ten most important predictor variables, including sweet taste preference, urinary glucose, age, heart rate, creatinine, waist circumference, and pulse pressure. The identification of variables such as sweet taste preference and urine indicators is an important finding because it broadens the understanding that glycemic control is greatly influenced by behavioral factors and sensory intake, which have been under-explored in traditional risk prediction approaches. This study shows that in developing a prediction model for uncontrolled blood sugar, non-laboratory variables can make a significant contribution when used optimally.

Research by Goldner et al. (2018) further demonstrates the potential of ML in predicting blood glucose with minimal input. Using only a short historical data set, this model was able to predict future blood glucose values with a median absolute error of 14.2 mg/dL and 91% of predictions falling within the range of ± 50 mg/dL. Although the focus of this study is different because it predicts short-term glucose, these findings are still relevant to the development of blood sugar control predictions in general because they show that glucose patterns can be analyzed and estimated accurately with artificial intelligence even using minimal data. High-performance models such as this have the potential to be applied in self-monitoring systems or implemented in wearable devices to provide early warnings to patients about increased glucose risk.

The study by Xie et al. (2019) also shows that ML-based diabetes risk prediction models perform strongly. The neural network model in this study showed the highest accuracy of 82.4%, with an AUC of 0.7949 and specificity of 90.2%. The study also found interesting risk factors such as sleep duration exceeding 9 hours and low frequency of health check-ups (<1 year). These results indicate that risk prediction does not only depend on biomarkers but is also influenced by seemingly simple lifestyle habits. Adding behavioral variables to the prediction model has great potential to enhance the richness of prediction patterns that are more relevant in the context of public health.

Meanwhile, research by Ojurongbe et al. (2024) shows the extraordinary capabilities of multivariable penalized logistic regression with an AUC reaching 99% on training data and 94% on test data. Despite the possibility of overfitting, this study provides important insights that creative variables such as Waist-to-Hip Ratio (WHR), Atherogenic Index of Plasma (AIP), and catalase levels can be used for more accurate diabetes risk prediction. This suggests that the use of oxidative biomarkers and lipid indices could be a new direction in the development of risk prediction models for poor blood sugar control.

The study by Jiang et al. (2023) makes a relevant contribution in the context of integrating prediction models into clinical practice. With a model accuracy of 91.24% and an AUC of 97%, this Random Forest-based model was then developed into an easy-to-use risk score card. Interestingly, the score card showed higher accuracy than the initial model, at 95.15%. These findings illustrate that translating machine learning results into a format that is easy to use by health workers or the public can improve the effectiveness of risk screening. This approach is very important in the development of predictive models for uncontrolled blood sugar control because model implementability is one of the crucial factors in its effectiveness.

In Indonesia, a number of studies have shown that machine learning methods have great potential for application in the early detection of diabetes and glycemic control. Research by Gunawan et al. (2025) found that the logistic regression algorithm provided an accuracy of 82.51%, precision of 88.04%, recall of 89.76%, and AUC of 83.12%. The most influential variables were age, diet (low consumption of fruits and vegetables), and diastolic blood pressure. This shows that simple and non-invasive variables can provide strong predictive performance, so that similar models can be easily applied in primary health facilities.

The study by Sidiq et al. (2025) shows how imbalanced data is often a major challenge in predictive modeling. The application of AdaBoost and SMOTE successfully improved model

performance from 81.87% accuracy to 82.83%, with an increase in ROC-AUC from 0.9031 to 0.9058. The improvement in performance in the minority class is a very important finding, considering that patients with uncontrolled diabetes are often a minority group in the clinical data population. A model that can predict between the majority and minority classes in a balanced manner is urgently needed in the context of blood sugar control prediction because high-risk groups are the subpopulations most in need of early intervention.

The study by Hanafi et al. (2024) provides an example of how even a simple linear regression model can become an applicable predictive tool when packaged in a web-based system. This model can predict blood sugar levels based on previous glucose history, which is very useful for monitoring short-term glucose changes. This shows that predictive models do not always require complex algorithms to benefit patients, as long as the model is able to consistently process historical data patterns.

The findings in Fitria et al. (2024) study using KNN reveal that even simple algorithms can be used for diabetes classification, but accuracy is highly dependent on dataset size. The challenge in selecting the optimal K value shows that certain algorithms are not suitable for small datasets or data with uneven distribution. This is relevant to the Indonesian context, where many health facilities still have limited access to comprehensive databases.

The research by Irfannandhy et al. (2024) confirms that the Random Forest and CatBoost algorithms have competitive performance with accuracies of 82% and 81%, respectively. Glucose, BMI, and age variables are the main predictors, showing consistency between studies that anthropometric factors and metabolic parameters remain important components in predicting diabetes risk and glycemic control.

The study by Usabili & Indahyanti (2024) shows that ensemble learning provides excellent accuracy (87.30%) and is categorized as excellent classification, although the SMOTE technique is needed to overcome data imbalance. A similar finding was observed in the study by Safitri et al. (2024), which found that logistic regression had the highest accuracy (75%), while random forest had the highest AUC value (0.82). A comparison of the three models shows that each algorithm has unique characteristics, and model selection must consider the population context and prediction objectives.

The study by Susanto & Cahyana (2025), which used optimized XGBoost, showed outstanding performance with an accuracy of 96.33%, recall of 97.16%, precision of 93.4%, and ROC-AUC of 0.99. This model produced only 6 false negatives, which is very important in a clinical context because detecting positive cases is more important than avoiding false positives. These results show that the boosting algorithm has the potential to be the main approach in developing a prediction model for uncontrolled blood sugar risk.

In summary, the overall findings of the study indicate that the most consistent risk predictor variables consist of clinical variables such as glucose, HbA1c, blood pressure, lipids, creatinine, and electrolytes; demographic variables such as age, gender, and BMI; lifestyle variables such as sleep duration, diet, and physical activity; and health behavior variables such as frequency of health checkups and medication adherence. This combination of variables provides

an opportunity for the development of a more accurate and relevant risk prediction model for uncontrolled blood sugar, especially for developing countries such as Indonesia.

The results of the study also revealed a number of challenges in developing prediction models, such as dependence on expensive laboratory data, the risk of overfitting in some models, limitations in generalization because the data often comes from a limited population, and the lack of integration of behavioral and social variables. In addition, the problem of data imbalance is a recurring challenge, which requires the use of oversampling techniques such as SMOTE.

Based on these findings, it can be concluded that there are significant opportunities for developing predictive models for uncontrolled blood sugar risk in type 2 diabetes patients, especially by utilizing machine learning algorithms and non-laboratory data. The resulting predictive models are expected to be not only accurate but also easy to apply in clinical practice in Indonesia. The integration of prediction models into primary health care systems, including community health centers and non-communicable disease prevention posts, can have a major impact on early risk detection, optimization of type 2 DM management, and prevention of long-term complications. All previous research results provide a strong foundation for developing more adaptive, sustainable, and applicable prediction models in Indonesia.

4. CONCLUSION

Based on the overall review in this document, it can be concluded that the development of diabetes risk prediction and uncontrolled glycemic control models is showing rapid progress in line with the development of modern analytical methods, particularly machine learning, which is capable of processing clinical, behavioral, demographic, and digital biomarker variables more comprehensively than traditional approaches. Various algorithms such as logistic regression, random forest, XGBoost, gradient boosting, neural networks, and ensemble learning have been proven to provide high predictive performance with competitive accuracy, sensitivity, and specificity in various international and national studies. These findings confirm that blood sugar control is influenced by a combination of multidimensional factors, including lifestyle, nutritional status, blood pressure, laboratory parameters, physical activity levels, and socioeconomic variables. In Indonesia, the predictive models developed show great potential for application in primary health care, especially since they can utilize non-laboratory data that is more easily accessible. Overall, these developments indicate that machine learning-based prediction models play an important role in early detection, risk stratification, and improving the effectiveness of diabetes management, thereby contributing to efforts to prevent long-term complications and improve patients' quality of life.

REFERENCES

- Abegaz, T. M., Ahmed, M., Sherbeny, F., Diaby, V., Chi, H., & Ali, A. A. (2023). Application of Machine Learning Algorithms to Predict Uncontrolled Diabetes Using the All of Us Research Program Data. *Healthcare*, 11(8), 1138. <https://doi.org/10.3390/healthcare11081138>

- Alfaqeeh, M., Alfian, S. D., & Abdullah, R. (2024). Factors Associated with Diabetes Mellitus among Adults: Findings from the Indonesian Family Life Survey-5. *Endocrine and Metabolic Science*, 14, 100161. <https://doi.org/10.1016/j.endmts.2024.100161>
- Dubale, M., Gizaw, K., & Dessalegn, D. (2023). Magnitude and Predictors of Poor Glycemic Control in Patients with Diabetes at Jimma Medical Center, Ethiopia. *Scientific Reports*, 13(1), 15952. <https://doi.org/10.1038/s41598-023-42774-y>
- ElSayed, N. A., Aleppo, G., Aroda, V. R., Bannuru, R. R., Brown, F. M., Bruemmer, D., Collins, B. S., Hilliard, M. E., Isaacs, D., Johnson, E. L., Kahan, S., Khunti, K., Leon, J., Lyons, S. K., Perry, M. Lou, Prahalad, P., Pratley, R. E., Seley, J. J., Stanton, R. C., & Gabbay, R. A. (2023). 6. Glycemic Targets: Standards of Care in Diabetes—2023. *Diabetes Care*, 46(Supplement_1), S97–S110. <https://doi.org/10.2337/dc23-S006>
- Fitria, R. N., Sugianto, W., & Nur'aidha, A. C. (2024). Prediksi Penyakit Diabetes Mellitus Tipe I and Tipe II Menggunakan Metode KNN di Klinik Dharma Husada. *Antigen: Jurnal Kesehatan Masyarakat and Ilmu Gizi*, 2(3), 39–53. <https://doi.org/10.57213/antigen.v2i3.303>
- Goldner, D. R., Osborn, C. Y., Sears, L. E., Huddleston, B., & Dachis, J. (2018). A Machine-Learning Model Accurately Predicts Projected Blood Glucose. *Diabetes*, 67(Supplement_1). <https://doi.org/10.2337/db18-46-LB>
- Gunawan, S., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). Prediksi Diabetes Mellitus Tipe 2 dengan Algoritma Logistic Regression untuk Pendeteksian Dini. *Jurnal Informatika and Teknik Elektro Terapan*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5747>
- Hanafi, H., Alfian, G., Widodo, T., & Syafrudin, M. (2024). Pemodelan Prediksi Kadar Gula Darah Pada Pasien Diabetes Menggunakan Metode Regresi Linear. *Journal of Internet and Software Engineering*, 5(1), 1–8. <https://doi.org/10.22146/jise.v5i1.8480>
- Hossain, M. J., Al-Mamun, M., & Islam, M. R. (2024). Diabetes Mellitus, the Fastest Growing Global Public Health Concern: Early Detection Should be Focused. *Health Science Reports*, 7(3). <https://doi.org/10.1002/hsr2.2004>
- IDF. (2025). *IDF Diabetes Atlas 11th Edition*. <https://diabetesatlas.org/resources/idf-diabetes-atlas-2025/#>
- Irfannandhy, R., Handoko, L. B., & Ariyanto, N. (2024). Analisis Performa Model Random Forest and CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 714–723. <https://doi.org/10.29408/edumatic.v8i2.27990>
- Jiang, L., Xia, Z., Zhu, R., Gong, H., Wang, J., Li, J., & Wang, L. (2023). Diabetes Risk Prediction Model Based on Community Follow-up Data Using Machine Learning. *Preventive Medicine Reports*, 35, 102358. <https://doi.org/10.1016/j.pmedr.2023.102358>
- Khan, M. A. B., Hashim, M. J., King, J. K., Govender, R. D., Mustafa, H., & Al Kaabi, J. (2019). Epidemiology of Type 2 Diabetes – Global Burden of Disease and Forecasted Trends. *Journal of Epidemiology and Global Health*, 10(1), 107. <https://doi.org/10.2991/jegh.k.191028.001>
- Lertwanichwattana, T., Suwannahitatorn, P., Mungthin, M., & Rangsin, R. (2024). Prognostic Nomogram for Uncontrolled Type 2 Diabetes Using Thailand Nation-wide Cross-sectional Studies. *PLOS ONE*, 19(4), e0298010. <https://doi.org/10.1371/journal.pone.0298010>
- Maciag, M. C., Sheehan, W. J., Bartnikas, L. M., Lai, P. S., Petty, C. R., Filep, S., Chapman, M. D., & Phipatanakul, W. (2021). Detection of Food Allergens in School and Home Environments

- of Elementary Students. *The Journal of Allergy and Clinical Immunology: In Practice*, 9(10), 3735–3743. <https://doi.org/10.1016/j.jaip.2021.06.014>
- Ojurongbe, T. A., Afolabi, H. A., Oyekale, A., Bashiru, K. A., Ayelagbe, O., Ojurongbe, O., Abbasi, S. A., & Adegoke, N. A. (2024). Predictive Model for Early Detection of Type 2 Diabetes Using Patients' Clinical Symptoms, Demographic Features, and Knowledge of Diabetes. *Health Science Reports*, 7(1). <https://doi.org/10.1002/hsr2.1834>
- Ong, K. L., Stafford, L. K., McLaughlin, S. A., Boyko, E. J., Vollset, S. E., Smith, A. E., Dalton, B. E., Duprey, J., Cruz, J. A., Hagins, H., Lindstedt, P. A., Aali, A., Abate, Y. H., Abate, M. D., Abbasian, M., Abbasi-Kangevari, Z., Abbasi-Kangevari, M., Abd ElHafeez, S., Abd-Rabu, R., ... Vos, T. (2023). Global, Regional, and National Burden of Diabetes from 1990 to 2021, with Projections of Prevalence to 2050: A Systematic Analysis for the Global Burden of Disease Study 2021. *The Lancet*, 402(10397), 203–234. [https://doi.org/10.1016/S0140-6736\(23\)01301-6](https://doi.org/10.1016/S0140-6736(23)01301-6)
- Reddy, N., Verma, N., & Dungan, K. (2023). *Monitoring Technologies- Continuous Glucose Monitoring, Mobile Technology, Biomarkers of Glycemic Control*. South Dartmouth (MA): MDText.com, Inc.; 2000. <https://www.ncbi.nlm.nih.gov/books/NBK279046/>
- Safitri, E., Rofianto, D., Purwati, N., Kurniawan, H., & Karnila, S. (2024). Prediksi Penyakit Diabetes Melitus Menggunakan Algoritma Machine Learning Diabetes Mellitus Disease Prediction using Machine Learning Algorithms. *JUSTIN: Jurnal Sistem and Teknologi Informasi*, 12(4), 760–766. <https://doi.org/10.26418/justin.v12i4.84620>
- Sherwani, S. I., Khan, H. A., Ekhzaimy, A., Masood, A., & Sakharkar, M. K. (2016). Significance of HbA1c Test in Diagnosis and Prognosis of Diabetic Patients. *Biomarker Insights*, 11, BMI.S38440. <https://doi.org/10.4137/BMI.S38440>
- Sidiq, S., Alfian, & Maburur, N. S. (2025). Pengembangan Model Prediksi Risiko Diabetes Menggunakan Pendekatan AdaBoost and Teknik Oversampling SMOTE. *Jurnal Ilmiah Informatika and Ilmu Komputer (JIMA-ILKOM)*, 4(1), 13–23. <https://doi.org/10.58602/jima-ilkom.v4i1.41>
- Susanto, E. R., & Cahyana, A. (2025). Penerapan Algoritma XGBoost untuk Prediksi Diabetes: Analisis Confusion Matrix and ROC Curve. *Fountain of Informatics Journal*, 10(1), 40–50. <https://doi.org/10.21111/fij.v10i1.14311>
- Usabili, S., & Indahyanti, U. (2024). Pemodelan Deteksi Dini Diabetes Mellitus menggunakan Pendekatan Ensemble Learning. *Manajemen Pelayanan Kesehatan*, 1(1), 12. <https://doi.org/10.47134/mpk.v1i1.2924>
- WHO. (2024, November 14). Diabetes. *Who.Int*, Home/Newsroom/Fact sheets/Detail/Diabetes. <http://who.int/news-room/fact-sheets/detail/diabetes>
- 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, 190109. <https://doi.org/10.5888/pcd16.190109>
- 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. <https://doi.org/10.1038/s41598-020-61123-x>