Machine Learning Models Prediction Medication Nonadherence Risk in Type 2 Diabetes: A Systematic Review

Victor Trismanjaya Hulu[1*], Yusuf Panserito Hulu[2], Kharis Meiwan K, Telaumbanua[3], Reni Aprinawaty Sirait[4], Arianus Zebua[5], Rumiris Simatupang[6]
[1]Department of Epidemiology, Faculty of Health Science, Universitas Prima Indonesia, Medan
[2,5]Sekolah Tinggi Ilmu Kesehatan Arta Kabanjahe
[3]Universitas Deztron Indonesia
[4]Department of Public Health Sciences, Faculty of Public Health, Institut Kesehatan Medistra, Lubuk Pakam
[6]Sekolah Tinggi Ilmu Kesehatan Nauli Husada
Email: victortrismanjayahulu@unprimdn.ac.id; nerspanser@gmail.com; kharismeiwan@udi.ac.id; reniaprinawaty@medistra.ac.id; arianuszebua1@gmail.com; rumirissimatupang@gmail.com*

ABSTRACT

The prediction of medication nonadherence among patients with T2DM can be improved in accuracy and speed using machine learning (ML). This study aimed to develop an ML model to predict the risk of medication nonadherence among patients with T2DM. Methods, inclusion criteria comprised English-language, open-access journal articles published between 2020 and 2025 that developed and validated ML–based prediction models, including ensemble methods, gradient-boosting models, SVMs, and neural networks. Exclusion criteria included review articles, non-English papers, studies published before 2020, studies lacking prediction model development or validation, and studies using only traditional statistical methods, such as logistic regression. The article search was conducted in PubMed, Scopus, ScienceDirect, and Google Scholar. Prediction Model Risk of Bias Assessment Tool (PROBAST) to assess the methodological quality and usefulness of the qualified studies. This narrative synthesis examines the characteristics of ML-based prediction models, their performance, and the factors that predict adherence among patients with T2DM. The papers were sourced from various scientific journal databases. The results show that cross-sectional and cohort studies were among the research designs used in the five papers reviewed. The AUROC of the internal test was 0.782, and the AUROC of the external test was 0.771. The learned-feature classification model achieved an average accuracy of 79.7%. Among these algorithms, the AUC of the best-performing algorithm was 0.866 ± 0.082. The SVM classifier outperformed the others, achieving a recall of 0.9979 and an AUC of 0.9998. The conclusion indicates that predictive capacity is influenced by clinical metrics and the number of prescribed medications.

Keywords: **Diabetes mellitus, medication non-adherence, diabetes medication consumption, machine learning**

INTRODUCTION

Type 2 diabetes mellitus (T2DM) constitutes a major global public health challenge, with an escalating burden particularly in emerging countries (Yan et al., 2022). T2DM is a major global health concern that calls for a reliable predictive method to enable earlier diagnosis and the introduction of more specific treatments (Kiran et al., 2025). Developing data-driven predictive models to identify the risk of medication non-adherence is crucial for enabling proactive,

targeted interventions. This is because medication nonadherence is a common and preventable cause of poor clinical outcomes (Rhudy et al., 2025). Long-term T2DM and unadjusted hypoglycemic therapy are big risk factors for diabetic complications. The number of hypoglycemic drugs is an important factor in glycemic control among T2DM patients who do not adhere to their medication regimen (Fan et al., 2021).

Globally, in 2022, 14% of persons aged 18 and older worldwide were diagnosed with diabetes, an increase from 7% in 1990. In that year, 59% of adults aged 30 and older with diabetes indicated they were not adhering to their diabetic medication schedule. The minimal treatment coverage was observed in low- and middle-income nations (WHO, 2024). T2DM will affect one in eight adults, or 853 million, by 2050, a 46% rise. Approximately 11.1% of 20–79-year-old adults have DM, with over 40% undiagnosed. T2DM is impacted by genetics, environment, socioeconomic status, and demographics, and it accounts for more than 90% of all DM (IDF, 2025). The sharp rise in T2DM cases over the last ten years, along with predictions that the number will continue to rise, makes it even more important to create ML forecast models that can find out how likely it is that a disease will start or get worse in different groups of people (Kiran et al., 2025; Deberneh & Kim, 2021; Ahmed et al., 2022).

Several studies have utilized statistical and ML models for predicting the likelihood of medication non-adherence in T2DM. These methods include logistic regression analysis, mixed ML models, and logistic least absolute shrinkage and selection operator (LASSO) models (Li et al., 2022; QiMuge et al., 2022; Wang et al., 2021). Thus, by combining routine medical record data (clinical parameters like HbA1c, fasting glucose, body mass index, and history of complications) with sociodemographic, economic, and behavioral factors of T2DM patients, an ML model to predict the risk of medication non-adherence in T2DM patients in developing nations like Indonesia can make a significant contribution to the development of a predictive model for medication non-adherence in T2DM patients. Therefore, this study aimed to develop an ML model to predict the risk of medication nonadherence among patients with T2DM.

METHODS

Approaches method

The procedures and materials used in this investigation were meticulously crafted to ensure a rigorous, open-minded approach to locating, filtering, and categorizing literature pertinent to the investigation. In accordance with the Preferred Reporting Items for Systematic Reviews

and Meta-Analyses (PRISMA) 2020 framework, this systematic review was carried out. This framework provides up-to-date guidance on improving reproducibility, transparency, and methodological rigour throughout the systematic review process.

Settings and study design of included studies

This systematic review looked at cohort and cross-sectional studies.

Eligibility criteria

To determine whether papers were eligible, we examined their titles, abstracts, or full texts. For this investigation, we used the following inclusion criteria: (1) journal papers published between 2020 and 2025, (2) papers published in English that are freely accessible online, and (3) research into the creation and testing of prediction models based on ML techniques (e.g., ensemble, gradient boosting/XGBoost, support vector machine, neural network). The exclusion criteria: (1) journal articles in the form of reviews and not in English, (2) research that does not construct or validate prediction models, as well as articles published prior to 2020, and (3) logistic regression and other classic pure statistical models that omit ML.

Data sources and search strategy

The initial strategy for searching articles and the procedure for formulating research questions use the PICO framework, namely P (Population): adult patients with T2DM who are prescribed drug therapy (oral antidiabetics and/or insulin). I (Intervention): development/validation of an ML model to predict the risk of medication non-adherence (random forest, gradient boosting/XGBoost, SVM, neural network, ensemble). C (Comparison): Non-ML/traditional model (logistic regression). O (Outcome): status of medication non-adherence, model prediction performance. Based on the PICO framework, this research question is: How well does the ML model predict the risk of medication nonadherence in patients with T2DM? The article search was conducted in PubMed, Scopus, ScienceDirect, and Google Scholar using the keywords listed in Table 1.

Table 1. Keyword Search Used in The Screening Process

Database	Keyword used
PubMed	("Diabetes Mellitus, Type 2"[Mesh] OR "type 2 diabetes" OR T2D OR T2DM) AND ("medication nonadherence" OR "medication non-adherence" OR nonadherence OR "non-compliance" OR "medication adherence" OR MPR OR "medication possession ratio" OR PDC OR "proportion of days covered") AND ("machine learning" OR "artificial intelligence" OR "random forest" OR XGBoost OR "support vector machine" OR SVM OR "neural network" OR "ensemble model" OR “deep learning) AND ("prediction model" OR "predictive model" OR "risk prediction" OR "risk model" OR validation OR "external validation")

Database	Keyword used
Scopus	TITLE-ABS-KEY (("type 2 diabetes" OR T2D OR T2DM) AND ("medication nonadherence" OR "medication non-adherence" OR nonadherence OR "non-compliance" OR MPR OR PDC OR "medication possession ratio" OR "proportion of days covered") AND ("machine learning" OR "artificial intelligence" OR "random forest" OR xgboost OR "support vector machine" OR "neural network" OR "ensemble") AND ("prediction model" OR "risk prediction" OR "risk model" OR validation))
ScienceDirect	("type 2 diabetes" OR "type 2 diabetes mellitus" OR T2D OR T2DM) AND ("medication nonadherence" OR "medication non-adherence" OR onadherence OR "non-compliance" OR "refill adherence") AND ("machine learning" OR "artificial intelligence" OR "random forest" OR XGBoost OR "support vector machine" OR SVM OR "neural network") AND ("prediction model" OR "risk prediction" OR "risk model" OR validation)
Google Scholar	("type 2 diabetes" OR T2D OR T2DM) ("medication nonadherence" OR "medication non-adherence") ("prediction model" OR "risk model" OR "model development" OR validation OR "external validation") ("machine learning" OR "artificial intelligence" OR "ensemble")

Study selection

After importing all identified research into Mendeley, duplicate journal articles were deleted. Two authors (VTH and YPH) independently assessed all paper titles and abstracts for eligibility. VTH, KMKT, and RAS discussed differences. All full-text screening was done by VTH, AR, and RS.

Quality assessment of studies

We used the Prediction Model Risk of Bias Assessment Tool (PROBAST) to assess the methodological quality and usefulness of the qualified studies. The PROBAST tool is made to evaluate the quality of clinical prediction model studies that have already been published (Wolff et al., 2019).

Data collection process

Four databases, Google Scholar, PubMed, Scopus, and ScienceDirect, are used to find journal papers. 511 duplicate journal articles were detected in 1.155 Mendeley searches. After screening for duplicates, 644 journal articles were selected, excluding 369 non-full-text articles. 203 of 275 reports were not retrieved. Next, 72 full-text journal papers were evaluated for eligibility; 67 were excluded for being design, acceptance, or review articles. Moreover, 5 of the articles met the criteria for review. Figure 1 illustrates article selection.

Data extraction

Six people extracted data from five eligible articles. Data extraction begins with the initial author's name, nation, study population, sample size, study design, data sources, follow-up length, AI modeling type, predictors, and model performance metrics. According to the Critical

Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling (CHARMS) framework, a standardized data collection form was used to extract data from the included studies (Moons et al., 2014).

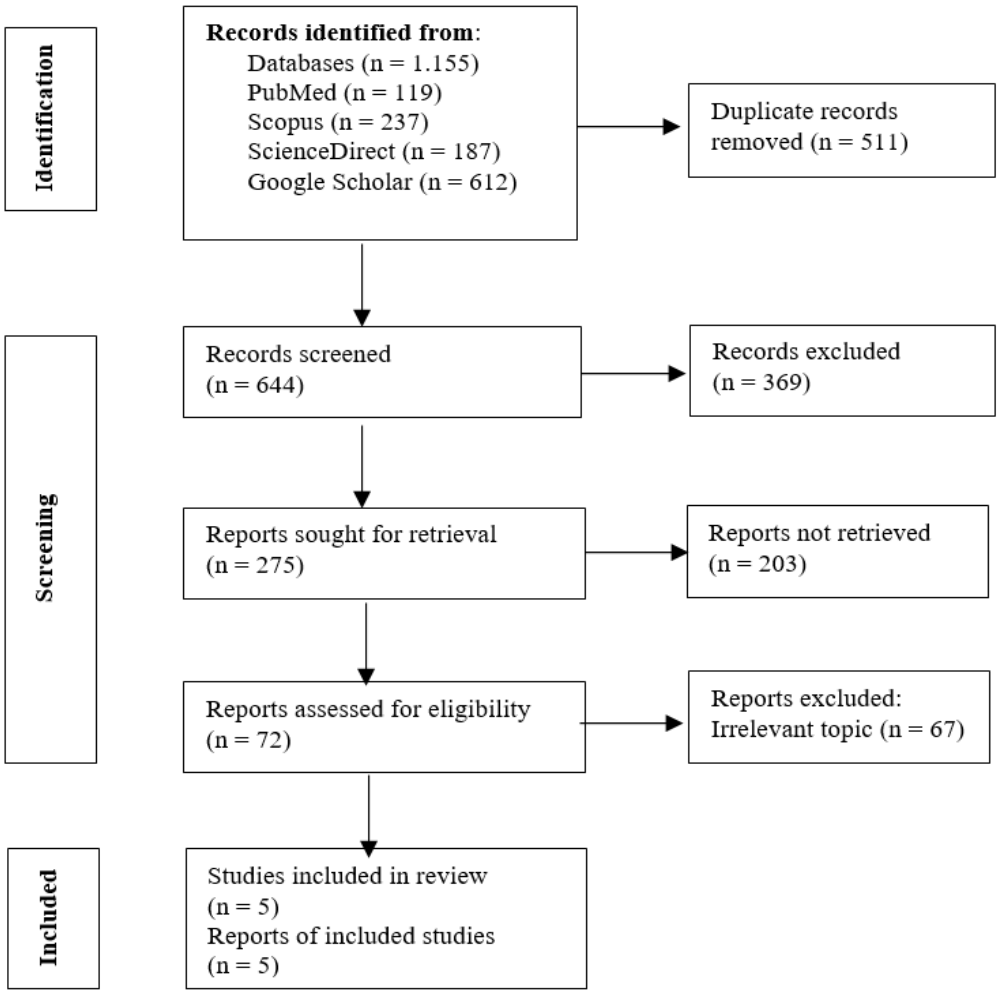

Figure 1. Article Selection Procedure (PRISMA)

Data analysis and synthesis

A qualitative synthesis of the data was conducted by analyzing five journal papers. This narrative synthesis examines the characteristics of ML-based prediction models, their performance, and factors that predict medication adherence among patients with T2DM.

RESULT

This study summarized all papers that met the inclusion criteria, as shown in Table 2. The papers were sourced from various scientific journal databases. Cross-sectional and cohort studies were among the research designs used in the five papers reviewed.

Table 2. Data Extraction

Author and Year	Country	Participant	Study Design	Data Source	Length Follow-up
(Chen et al., 2024)	Taiwan	4.134	Cohort	Taipei Medical University Clinical Research Database	90 days
(Wu et al., 2020)	China	401	Cross-sectional survey	Electronic medical records (EMR) and face-to-face questionnaires at Sichuan Provincial People's Hospital.	April 1, 2018, to March 30, 2019.
(Kassaw et al., 2025)	Ethiopia	403	Cross-sectional	Structured interview-based questionnaire and medical records at the University of Gondar Comprehensive Specialized Hospital (UoGCSH).	February to May 2023.
(Li et al., 2022)	China	980	Cross-sectional	EMR and questionnaire at Sichuan Provincial People's Hospital.	April 2018 to December 2019.
(Thyde et al., 2021)	Denmark	Not available	Cohort	CGM (Continuous Glucose Monitoring) data generated through a modified MVP (Medtronic virtual patient) model for T2DM.	Not available

The attributes of the included studies were classified by year of publication and country of origin. Three journal articles used a cross-sectional design, whilst two used a cohort research methodology. The study's attributes are illustrated in Figure 2 below.

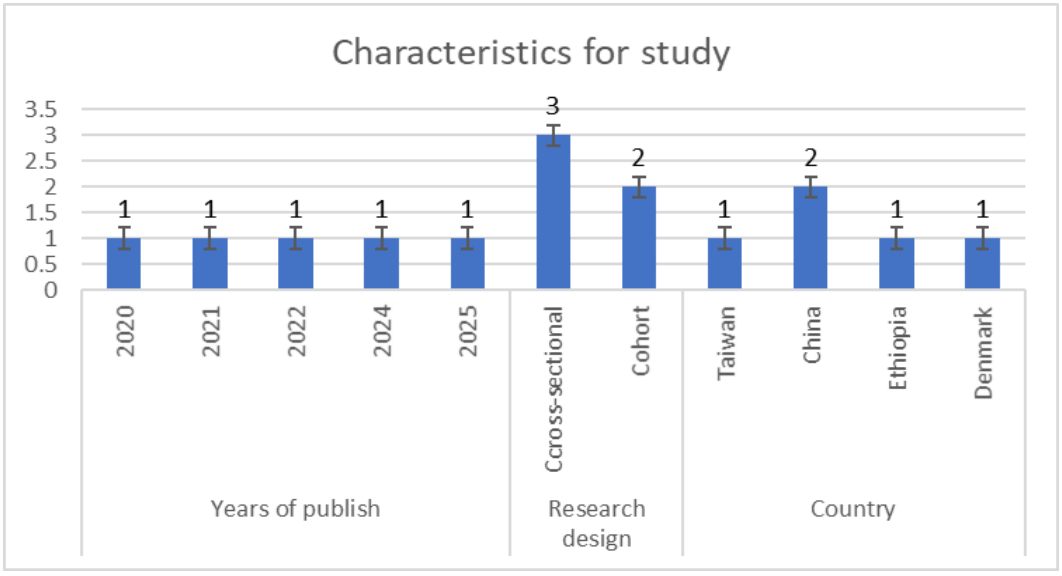

Figure 2. Characteristics for Study Selection (n=5)

Table 3. Predictor Information, AI Modeling Type, and Performance Metrics

Author and Year	AI Modelling type	Predictors	Model performance metrics
(Chen et al., 2024)	XgBoost	Demographics, baseline comorbidities, baseline hypertension medications, baseline laboratory data, dyslipidemia, T2DM, index insulin, baseline healthcare resource utilisation (e.g., inpatient visit counts), and concomitant non-insulin T2DM medications.	Validation AUROC (Area Under the curve of the Receiver Operating Characteristic) Internal 1: 0.791 Internal 2: 0,783 Internal 3: 0,785 Internal 4: 0,777 Internal 5: 0,773 External data: 0.771
(Wu et al., 2020)	Ensemble models (C 5.0 model, logistic regression model, Bayesian network, discriminant model, KNN (K-Nearest Neighbor) algorithm, Random Forest, Support Vector Machine, Tree-AS, CHAID (Chi-squared Automatic Interaction Detection, Quest, C&R Tree Neural. Net, support vector machine)	Nine variables were used to build this model, which included age, gender, whether the prior fasting blood glucose was under control, duration of the current treatment regimen, diet adjustment, daily medication cost, fasting blood glucose value, hyperlipidemia, and BMI (Body Mass Index).	Ensemble with AUC (Area Under the Receiver Operating Characteristic Curve): 0.866±0.082; Precision: 0.824±0.043 Recall: 0.732±0.061; F1 score: 0.773±0.032.
(Kassaw et al., 2025)	Ensemble model (Logistic Regression (LR), Support Vector Machine (SVM), K Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Gradient Boost Classifier (GBC), Multilayer Perceptron (MLP), and 1D Convolutional Neural Network (1DCNN))	Patient behaviour, medication pill or injection burden, medication cost and payment	Support Vector Machine (SVM). Accuracy: 0.9935 Precision: 0,9903 Recall: 0.9969 F1-score: 0.99357 AUC ROC: 0.9998
(Li et al., 2022)	(AdaBoost, Extreme Gradient Boosting (XGBoost), gradient boosting, Bagging, Bernoulli Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes, decision tree, extra tree, K-nearest neighbor (KNN), linear discriminant analysis (LDA), quadratic discriminant analysis	1. Basic characteristics include: age, gender, waistline (cm), weight (Kg), occupational status, education level, family history of diabetes mellitus, BMI (kg/m2), and health status scores (%). 2. Clinical information includes: Course of diabetes (in months), medicare status, frequency of FBG measurements, interval of measurement (in days), previous HbA1c values, present HbA1c values, present FBG level, present FBG values (mmoL/L), present RBG values (mmoL/L), present	Model 1 Ensemble includes AUC: 0.8369 Accuracy: 0.7092 Precision: 0.9474 Recall: 0.6792 F1 Score: 0.7912 AUPRC: 0.9574

Author and Year	AI Modelling type	Predictors	Model performance metrics
	(QDA), logistic regression, passive-aggressive, random forest, Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), and an ensemble algorithm.	PBG values (mmoL/L), type of operation or other communicable deseases, number of comorbid diseases, hypertension, hyperlipidemia 3. Exercise, diet, and mental state include: intensity of exercise, exercise session (mins/day), had a ration and reasonable eating, sleep duration, psychological status, EQ-5D scores. 4. Treatment regimen and medication adherence includes: duration of treatment regimen (in months), type of insulin used, use of insulin, times of insulin use, dose of basal insulin (U), dose of non-basal insulin in morning (U), dose of non-basal insulin in afternoon (U), number of oral drugs, use of other types of drugs, use of metformin, dose of metformin, type of manufacturers of metformin, α-Glucosidase inhibitors, sulfonylureas, DPP-4 inhibitors.	
(Thyde et al., 2021)	CNN (convolutional neural networks), MLP (Multi Layer Perceptron and LR (Logistic Regression)	Minimum PG measure of the interval, maximum PG measure of the interval, mean of entire interval, SD of the interval, percent of interval with PG above 90 mg/dL (5 mmol/L), percent of interval with PG above 108 mg/dL (6 mmol/L), percent of interval with PG above 126 mg/dL (7 mmol/L), percent of interval with PG above 144 mg/dL (8 mmol/L), area under the PG measures in the interval, lowest mean hour of the interval, lowest mean hour between 6:00 am and 9:00 am	META (Mean Ensemble Test Accuracy: A0: 78.6% ± 0.6% A1: 78.2% ± 0.8% A2: 78.3% ± 1.1% A3: 79.7% ± 0.4% A4: 79.7% ± 0.8% A5: 79.8% ± 0.5%

DISCUSSION

The majority of people who have T2DM do not take their medications as prescribed, according to others research. Ensuring drug adherence is greatly influenced by factors such as family support, medication affordability, and the quality of communication between patients and healthcare professionals (Waari et al., 2018). Patients' care for themselves, and awareness of the significance of prescribed medication, is shown in the low number of diabetic patients in Bulgaria who do not follow their treatment plans (Dinkova et al., 2023). Therefore, it is critical to use ML to predict when T2DM patients will not take their medication as prescribed. Prior research indicated that the built ML model demonstrated strong predictive power in identifying

patients who did not attend follow-up visits for diabetes management following a screening program (Okada et al., 2022).

In a study by Chen et al. (2024), an ML model based on the XGBoost algorithm can accurately predict when to start insulin therapy in patients with T2DM. The AUROC values of 0.782 for internal measures and 0.771 for external validation indicate the model's performance. Adherence to injectable medications, like insulin, is reported to be only about 60%, which is still below the clinically acceptable adherence threshold of 80%, despite the fact that diabetes is a chronic condition requiring long-term management. Demographics, baseline therapy, baseline comorbidities, baseline laboratory parameters, healthcare utilization, insulin index, and the use of non-insulin antidiabetic drugs concurrently with therapy are among the seven predictor categories used in this study (Chen et al., 2024).

The study's findings Thyde et al. (2021) demonstrate that adherence to daily basal insulin injections in patients with T2DM can be effectively detected using an ML algorithm based on Continuous Glucose Monitoring (CGM) simulation data, with accuracy improving as additional daily data are acquired. The expert-engineered feature-based models averaged 78.2%–78.6%, while the learnt models averaged 79.7%. The hybrid model, which combined both types of characteristics, achieved the highest accuracy (79.8%) 16 hours after insulin injection.

Furthermore, the study's findings Wu et al. (2020) demonstrated that the use of ML models could accurately forecast the likelihood that patients with T2DM would not comply with their treatment plans. The Ensemble model, which performed best, achieved an AUC of 0.866 ± 0.082. The survey found that 21.20 per cent of the 401 patients at Sichuan Provincial People's Hospital were not adhering to their medication regimens. Nine variables were used to construct this predictive model: gender, age, duration of current treatment, presence of hyperlipidemia, fasting blood glucose levels, dietary adjustments, daily treatment costs, BMI, and past fasting blood glucose control status (Wu et al., 2020).

Based on findings Kassaw et al. (2025), 77.45% (95% CI: 70.1–83.8) of T2DM patients did not adhere to their medications. However, ML was highly effective at detecting and classifying compliance levels. The SVM performed best of the eight algorithms, with a recall of 0.9969 and an AUC of 0.9998. According to Li et al. (2022), ML models, particularly an ensemble approach combining a modified random forest for data imputation, random under-sampling, and Boruta feature selection, were best at predicting medication non-adherence among patients

with T2DM. The model had an AUC of 0.8369 and an Area Under the Precision–Recall Curve (AUPRC) of 0.9574. Out of 980 patients, 18.8% did not adhere to treatment. Age, current fasting blood glucose, current HbA1c, current random blood glucose, and BMI are the biggest predictors of medication adherence risk.

T2DM patients' medication, diet, and exercise adherence are poor. By identifying adherence variables, diabetes management can be improved, and messages tailored to enhance glycemic control (Mirahmadizadeh et al., 2020). Living with DM for less than three years (adjusted OR (AOR) 3.37, 95% CI 1.91 to 5.95), residing in a rural area (AOR 2.67, 95% CI 1.49 to 4.79), having comorbidities (AOR 2.99, 95% CI 1.67 to 5.34), and lacking formal education (AOR 3.26, 95% CI 1.49 to 7.00) were all factors that were significantly linked to non-adherence. Important variables strongly linked to non-adherence included living in a rural area, having comorbid conditions, having less education, and having a shorter time since diagnosis (Kassaw et al., 2025). According to prior research, 58.6% of people with T2DM did not take their medication as prescribed (95% CI: 54.7-62.4). The independent predictors of non-adherence to medication included individuals experiencing major depressive disorder (AOR= 2.3; 95% CI: 1.1, 5.8), having one or more complications (AOR= 3.3; 95% CI: 1.9, 9.0), and having an average income of more than 1000 birr (AOR= 0.4; 95% CI: 0.1, 0.9) (Kusa et al., 2019).

CONCLUSION

The effectiveness of artificial intelligence in this context is evident through the use of ensemble and deep learning models, notably convolutional neural networks, which outperform traditional statistical methods in detecting diabetes status and predicting medication non-adherence. The predictive capacity of these models is largely driven by clinical measures such as HbA1c, blood glucose, and lipid profiles, in combination with indicators of healthcare utilization and the number of prescribed medications.

Future research should place greater emphasis on incorporating patient-reported outcome measures, such as health literacy, disease awareness, and social support, as these factors often play a critical role in medication nonadherence but are not routinely or automatically captured in electronic medical record systems.

LIMITATION

This study is subject to several limitations, including the use of simulated data that may not fully capture real-world patient variability, limited generalizability due to single-center study designs, and potential recall bias in questionnaire data.

REFERENCES

Ahmed, U., Issa, G. F., Khan, M. A., Aftab, S., Khan, M. F., Said, R. A. T., Ghazal, T. M., & Ahmad, M. (2022). Prediction of diabetes empowered with fused machine learning. *IEEE Access*, *10*, 8529–8538. https://doi.org/10.1109/ACCESS.2022.3142097

Chen, Y.-L., Nguyen, P.-A., Chien, C.-H., Hsu, M.-H., Liou, D.-M., & Yang, H.-C. (2024). Machine learning-based prediction of medication refill adherence among first-time insulin users with type 2 diabetes. *Diabetes Research and Clinical Practice*, *207*, 111033. https://doi.org/10.1016/j.diabres.2023.111033

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. https://doi.org/10.3390/ijerph18063317

Dinkova, R., Marinov, L., Doneva, M., & Kamusheva, M. (2023). Medication Adherence among patients with diabetes mellitus and its related factors-a real-world pilot study in Bulgaria. *Medicina (Kaunas, Lithuania), 59*(7), 1205. https://doi.org/10.3390/medicina 59071205

Fan, Y., Long, E., Cai, L., Cao, Q., Wu, X., & Tong, R. (2021). Machine learning approaches to predict risks of diabetic complications and poor glycemic control in nonadherent type 2 diabetes. *Frontiers in Pharmacology*, *12*, 665951. https://doi.org/10.3389/fphar. 2021.665951

IDF. (2025). Diabetes Facts & figure*s*. https://idf.org/about-diabetes/diabetes-facts-figures/

Kassaw, A. T., Tarekegn, T. B., Derbie, A., Ashagrie, G., Girmaw, F., & Mengesha, A. (2025). Non-adherence to antidiabetic medications and associated factors among adult type 2 diabetes mellitus patients in Northeast Ethiopia: Institutional based cross-sectional study. *BMJ Open*, *15*(4), e093413. https://doi.org/10.1136/bmjopen-2024-093413

Kassaw, E. A., Sendekie, A. K., Enyew, B. M., & Abate, B. B. (2025). Machine learning applications to classify and monitor medication adherence in patients with type 2 diabetes in Ethiopia. *Frontiers in Endocrinology*, *16,* 1486350*.* https://doi.org/10.3389/fendo. 2025.1486350

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*, https://doi.org/10.3389/fdgth.2025.1557467

Kusa, W., Tolessa, D., & Abdeta, T. (2019). Type II DM medication non-adherence in Adama Hospital Medical College, Central Ethiopia. *East African Journal of Health and Biomedical Sciences*, *3*(1), 31–38. https://www.researchgate.net/publication/336231945_ Type_II_DM_Medication_Non-Adherence_in_Adama_Hospital_Medical_College_ Central_Ethiopia

Li, M., Lu, X., Yang, H., Yuan, R., Yang, Y., Tong, R., & Wu, X. (2022). Development and assessment of novel machine learning models to predict medication non-adherence risks in type 2 diabetics. *Frontiers in Public Health*, https://doi.org/ 10.3389/fpubh.2022.1000622

Mirahmadizadeh, A., Khorshidsavar, H., Seif, M., & Sharifi, M. H. (2020). Adherence to medication, diet and physical activity and the associated factors amongst patients with type 2 diabetes. *Diabetes Therapy*, *11*(2), 479–494. https://doi.org/10.1007/s13300-019-00750-8

Moons, K. G. M., de Groot, J. A. H., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., Reitsma, J. B., & Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Medicine*, *11*(10), e1001744. https://doi.org/10.1371/journal.pmed.1001744

Okada, A., Hashimoto, Y., Goto, T., Yamaguchi, S., Ono, S., Ikeda Kurakawa, K., Nangaku, M., Yamauchi, T., Yasunaga, H., & Kadowaki, T. (2022). A machine learning–based predictive model to identify patients who failed to attend a follow-up visit for diabetes care after recommendations from a national screening program. *Diabetes Care*, *45*(6), 1346–1354. https://doi.org/10.2337/dc21-1841

QiMuge, N., Fang, X., Chang, B., Li, D. M., & Li, Y. (2022). Predicting population: development and validation of a new predictive nomogram for evaluating medication nonadherence risk in a type 2 diabetes. *PeerJ*, *10*, e13102. https://doi.org/10.7717/peerj.13102

Rhudy, C., Johnson, J., Perry, C., Bumgardner, C., Wesley, M. J., Fardo, D., Barrett, T., & Talbert, J. (2025). Machine learning approaches to predicting medication nonadherence: a scoping review. *International Journal of Medical Informatics*, 106082. https://doi.org/10.1016/j.ijmedinf.2025.106082 Get rights and content

Thyde, D. N., Mohebbi, A., Bengtsson, H., Jensen, M. L., & Mørup, M. (2021). Machine learning-based adherence detection of type 2 diabetes patients on once-daily basal insulin injections. *Journal of Diabetes Science and Technology*, *15*(1), 98–108. https://doi.org/10.1177/1932296820912411

Waari, G., Mutai, J., & Gikunju, J. (2018). Medication adherence and factors associated with poor adherence among type 2 diabetes mellitus patients on follow-up at Kenyatta National Hospital, Kenya. *Pan African Medical Journal*, *29*(1), 1–15. https://doi.org/10.11604/pamj.2018.29.82.12639

Wang, F.-C., Chang, W., Nie, S.-L., Shen, B.-X., He, C.-Y., Zhao, W.-C., Liu, X.-Y., & Lu, J.-T. (2021). Predicting medication nonadherence risk in the Chinese type 2 diabetes mellitus population–establishment of a new risk nomogram model: A retrospective study. *Journal of International Medical Research*, *49*(9), https://doi.org/10.1177/03000605211042502

WHO. (2024). *Diabetes*. https://www.who.int/news-room/fact-sheets/detail/diabetes

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., Mallett, S., & Group†, P. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, *170*(1), 51–58. https://doi.org/10.7326/M18-1376

Wu, X.-W., Yang, H.-B., Yuan, R., Long, E.-W., & Tong, R.-S. (2020). Predictive models of medication non-adherence risks of patients with T2D based on multiple machine learning algorithms. *BMJ Open Diabetes Research & Care*, *8*(1). https://doi.org/10.1136/bmjdrc-2019-001055

Yan, Y., Wu, T., Zhang, M., Li, C., Liu, Q., & Li, F. (2022). Prevalence, awareness and control of type 2 diabetes mellitus and risk factors in Chinese elderly population. *BMC Public Health*, *22*(1), 1–6. https://doi.org/10.1186/s12889-022-13759-9