American Journal of Biomedical Science & Research
www.biomedgrid.com
ISSN: 2642-1747

Mini Review

Machine Learning-Based Prediction of Type 2 Diabetes Mellitus Using Routine Laboratory Parameters

Danish Arshad[1], Amna Maqsood[2], Shahzad Akram[3], Ihsan Ali[4], Laraib Akhtar Khan[5], Aiman Farooq[4], Muhammad Shoaib rafiq[6], Muhammad Yasir Malik[2-7]*

[1]*Omsk State Agrarian University named after P.A. Stolypin, Russia*

[2]*Virology Lab, Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture Faisalabad Pakistan*

[3]*Faculty of Veterinary Medicine, Chiang Mai University, Chiang Mai, 50100, Thailand*

[4]*Institute of Molecular Biology and Biotechnology (IMBB), Bahauddin Zakariya University (BZU), Multan*

[5]*Department of Botany, University of Agriculture Faisalabad*

[6]*Department of Pathology, The Islamia University of Bahawalpur, Pakistan*

[7]*Department of Agriculture and Forestry, Southwest University of Science and Technology (SWUST), Mianyang, Sichuan, china,*

ORCID: 0000-0001-5336-6637

***Corresponding author:** Muhammad Yasir Malik, Department of Agriculture and Forestry, Southwest University of Science and Technology (SWUST), Mianyang, Sichuan, china, **ORCID:** 0000-0001-5336-6637

To Cite This article: *Danish Arshad, Amna Maqsood, Shahzad Akram, Ihsan Ali, Laraib Akhtar Khan, Aiman Farooq, Muhammad Shoaib rafiq, Muhammad Yasir Malik*, Machine Learning-Based Prediction of Type 2 Diabetes Mellitus Using Routine Laboratory Parameters. Am J Biomed Sci & Res. 2026 32(2) AJBSR.MS.ID.004143,* **DOI:** 10.34297/AJBSR.2026.32.004143

Received: September 01, 2026 **Published:** September 16, 2026

Abstract

Type 2 Diabetes Mellitus (T2DM) is a common metabolic disorder that often develops insidiously and may remain undiagnosed for many years. Early identification of individuals at high risk is essential for timely intervention and prevention of diabetes-related complications. Routine laboratory parameters, including fasting blood glucose, glycated haemoglobin (HbA1c), lipid profile, liver and kidney function markers, and haematological indices, provide valuable information on metabolic health and disease risk. However, these parameters are often assessed individually in conventional clinical practice and may not fully capture complex interactions among multiple biomarkers. Machine learning (ML) offers a promising approach for integrating routine laboratory data and identifying patterns associated with T2DM risk. Several algorithms, including logistic regression, decision trees, random forests, support vector machines, and gradient boosting methods, have demonstrated potential for diabetes prediction and risk stratification. In this mini-review, we summarize the application of ML techniques for T2DM prediction using commonly available laboratory parameters, with particular emphasis on biomarker selection, model performance, clinical applicability, and methodological challenges. Although ML-based models have demonstrated promising predictive performance, challenges related to data quality, over fitting, population heterogeneity, model interpretability, and limited external validation remain. Integrating reliable laboratory data with interpretable and externally validated ML models may improve early T2DM risk assessment and support clinical decision-making. Future research should emphasize standardized laboratory data, prospective studies, and robust external validation to facilitate the clinical translation of ML-based models into routine laboratory practice.

‘Keywords: Type 2 Diabetes Mellitus, Machine Learning, Laboratory Parameters, Biomarkers, HbA1c, Glucose, Early Diagnosis, Clinical Laboratory.

Introduction

Type 2 Diabetes Mellitus (T2DM) is a persistent metabolic disease caused mainly by the insulin resistance and gradual decrease of β-cell function. It is becoming more common and is a significant burden on health care systems around the world with many people not diagnosed until when metabolic abnormalities start to become apparent. Thus, timely identification of persons at high risk is critical to early intervention and prevention of diabetes-related complications [1,18]. Current diagnostic tests are primarily based on Glycated Hemoglobin (HbA1c), Fasting Plasma Glucose (FPG), 2-h Plasma Glucose (PG) from an Oral Glucose Tolerance Test (OGTT), or random plasma glucose (PG) in people with typical symptoms (American Diabetes Association Professional Practice Committee, 2026). But these individual measurements may not capture the complex relationships between metabolic and hematological traits of T2DM. In recent years, Machine Learning (ML) has shown great promise as a tool to leverage multiple clinical and lab parameters to uncover patterns that are linked to diabetes risk. ML algorithms can detect non-linear relationships and interactions between several variables, which are usually hard to detect in conventional statistical models that assume predefined relationships between predictors and outcomes [6,12]. Earlier, systematic reviews have reported positive predictive performance of ML models in the prediction of T2DM, but methodological limitations and lack of external validation of the models are significant obstacles to their clinical application [5,6].

Routine Laboratory Parameters for Diabetes Prediction

A major reason that the routine laboratory investigations are attractive for diabetes prediction using ML is that they are readily available in most clinical laboratories, also they are relatively cheap and are produced routinely. FPG and HbAIC are the most clinically proven biochemistry markers of diabetes. The American Diabetes Association defines diabetes by an HbA1c ≥6.5%, FPG ≥126 mg/dL, 2-h plasma glucose ≥200 mg/dL after a 75-g OGTT, or random plasma glucose ≥200 mg/dL with classic symptoms or hyperglycemic crisis (Professional Practice Committee of the American Diabetes Association, 2026) [1]. In addition to glucose parameters, lipid parameters like triglycerides and high-density lipoprotein cholesterol can also give clues to insulin resistance and a state of metabolic dysfunction. Other biochemical parameters that are regularly assessed might also help in predicting diabetes, such as liver and kidney function markers. Several studies have shown that routine clinical and laboratory parameters can be used to build diabetes prediction ML models using combinations of these parameters [15,9]. The hematological parameters are also increasingly being considered as potential predictors. In an approximately 9,000 adult cohort study, often measured haematological markers were examined for their link with T2DM, highlighting the potential of modestly priced blood-based markers in predicting diabetes [10,11]. More recently, Li, et al., [8] employed routine blood parameters to create ML models and compared XGBoost with random forest, support vector machine, and elastic-net methods. Their XGBoost model showed high prediction accuracy, and the SHAP analysis revealed that many hematological parameters were included as significant predictors [8] (Figure 1).

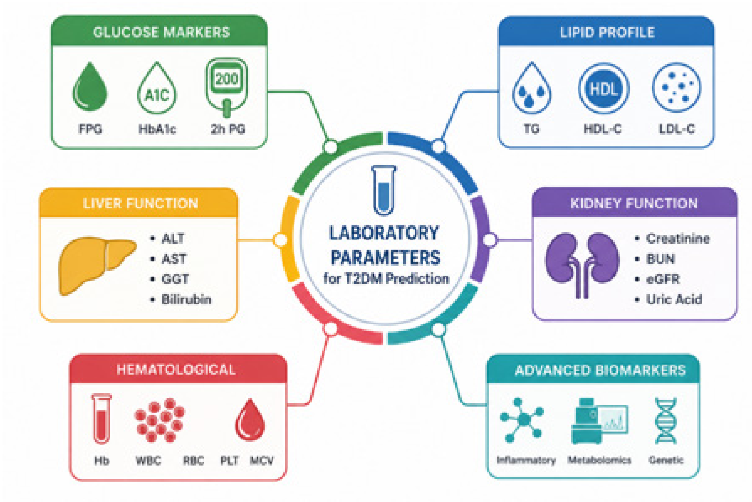

Figure 1: Routine laboratory parameters used for machine-learning-based prediction of T2DM.

Machine Learning Approaches

A few supervised ML algorithms have been explored for diabetes prediction, such as logistic regression, decision trees, random forests, Support Vector Machines (SVM), gradient boosting, XGBoost and artificial neural networks. These algorithms vary in their capacity to model the nonlinear relationships and interactions between the laboratory variables [12,6]. A systematic review and meta-analysis of 23 studies with 40 prediction models indicated that the pooled c-index value for predicting T2DM using ML was ~0.812. But no one of the studies presented was externally validated, which indicates a discrepancy between the development of a model and clinical implementation [5]. In another meta-analysis, Olusanya et al. [13] found that decision tree models exhibited the highest accuracy with over 136,000 participants in 34 studies, reporting a mean accuracy of 86%. Researches have also shown that using typical health-screening variables, diabetes prediction is also possible with ML. Shin, et al., [15,16] created prediction models using the information which is easily collected from health examination and assessed the performance of gradient boosting and random forest algorithms. Also, Olisah et al., [12] showed that the performance of diabetes prediction models could change significantly based on the preprocessing, feature selection, handling of missing values, and hyperparameters tuning. Explainable ML has been highlighted in more recent studies. Using SHAP analysis, Li et al., (2024) [7,8] found individual routine blood parameters that contribute to predicting diabetes. Explainability is significant since clinicians and laboratory professionals want to comprehend which of these variables are most crucial to a prediction, instead of assuming an unexplained risk score that they do not understand [8,9] (Figure 2).

Figure 2: Machine-learning workflow for T2DM prediction using routine laboratory data.

Clinical Potential and Relevance to Medical Laboratory Technology

Combining ML with standard lab testing might be another avenue of determining who may need more diabetes evaluation. Importantly, the prediction of ML should not supplant, but be used in addition to, existing diagnostic laboratory tests. Current guidelines continue to recommend standardized HbA1c and plasma glucose measurements for diagnosis (American Diabetes Association Professional Practice Committee, 2026) [1]. Many structured data are produced in routine laboratory medicine that could be used for analyzing with ML algorithms. There were reviews of ML applications in laboratory medicine focusing on the possibility of automated interpretation, prediction, anomaly detection and clinical decision support [14,3,4]. In the context of diabetes, EHRs with regular laboratory metrics have also been employed to anticipate glycemic abnormalities as well as to facilitate proactive clinical interventions [17]. The importance of reliability of laboratory data is especially significant for Medical Laboratory Technology (MLT) workers. Errors in the preanalytical process, analytical variation, differences in assays or laboratory methods, absence of values and reference interval discrepancies may lead to a low performance of the models. Therefore, data of good quality and standardized laboratory procedures are vital for establishing reliable prediction models [14,12] (Table 1).

Table 1: Clinical relevance of machine learning and laboratory data for T2DM prediction.

Aspect	Clinical/Laboratory Relevance	Implication for MLT
Routine Testing	HbA1c, glucose, lipid profile, and other parameters feed ML risk prediction models.	Accurate, consistent testing ensures reliable ML input data.
Diagnostic Support	ML flags individuals needing further diabetes evaluation.	ML complements, not replaces, standardized diagnostic testing.

Hba1c & Glucose	Core diagnostic markers per current clinical guidelines.	Requires strict accuracy, QC, and standardized procedures.
Automated Interpretation	ML aids pattern recognition in lab results.	MLT professionals validate and implement decision-support systems.
Prediction & Early Detection	Lab data predicts glycemic abnormalities and risk.	Reliable records enable earlier clinical intervention.
Data Quality	Poor/inconsistent data weakens model performance.	Rigorous QA/QC is essential throughout testing.
Preanalytical Errors	Sample handling issues add variability to ML inputs.	Standardized preanalytical procedures are critical.
Analytical Variation	Measurement differences affect predictions.	Regular calibration and method verification needed.
Assay/method Differences	Platforms may yield systematically different results.	Harmonization improves model generalizability.
Missing Values	Gaps in data reduce accuracy and introduce bias.	Complete, well-documented records are essential.
Reference Intervals	Inter-lab differences affect interpretation.	Documentation and harmonization needed.
Clinical Decision Support	ML aids risk stratification and evaluation decisions.	MLT contributes to quality and clinical integration.
Model Reliability	Needs high-quality, standardized, representative data.	Lab standardization is key to useful ML prediction.

Challenges and Future Perspectives

However, several hurdles prevent the clinical utilization of diabetes prediction using ML. There are many studies with relatively small, single-center datasets, with the potential of overfitting and low generalizability. Age, sex, ethnicity, prevalence of disease, laboratory technique and population characteristics can also affect model performance [5,6].

One of the critical steps before the deployment of ML models in clinical practice is external validation. Independent data sets of models and preferably prospective clinical data sets should be used to evaluate models. In addition, if predictors, such as HbA1c or glucose, are included in the definition of diabetes, they need to be interpreted with caution as this can lead to target leakage and artificially high model performance [8].

Therefore, future studies need to concentrate on large multicenter datasets, standard values measured in the labs, proper selection of features, external validation and explainable algorithms. Advanced biomarkers or metabolomics data can be used in combination with routine lab parameters to further stratify risk. Recent metabolomic studies coupled with ML have revealed potential metabolomic panels for T2DM, highlighting the future potential of the integration of the laboratory medicine with computational methods [2] (Figure 3).

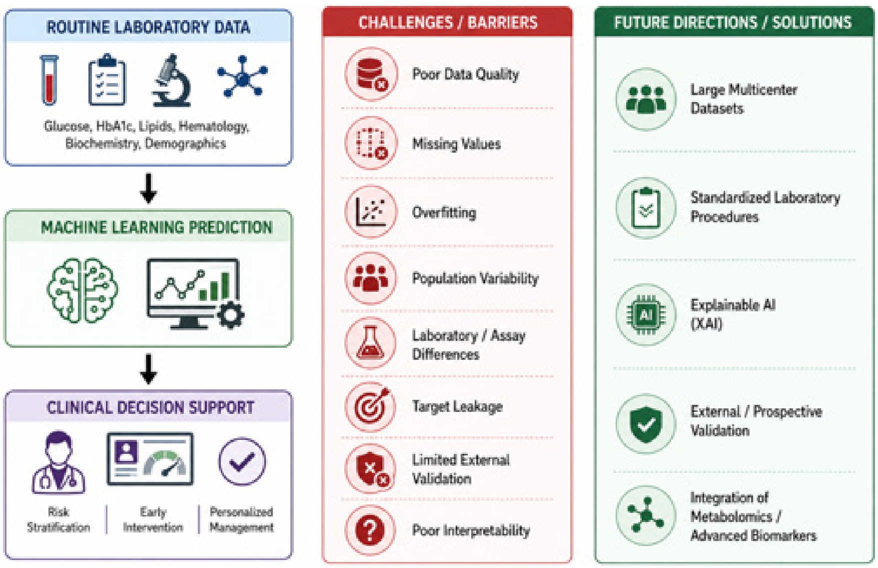

Figure 3: Clinical translation, challenges and future direction of ML-based T2DM prediction.

Conclusion

Machine learning offers a potential remedy to combine routine laboratory parameters and enhance prediction and early diagnosis of T2DM. Other conventional measures (HbA1c, glucose) can be supplemented with lipid, biochemical and hematological measures to reveal complex patterns associated with diabetes risk. Evidence thus far indicates good predictive accuracy for a variety of ML algorithms, but there are certain restrictions on the quality of the data, overfitting, explainability, and external validation of these models. In order to practice ML, a successful implementation of ML will rely on accurate laboratory testing, standardized data generation and clinically interpretable models, and rigorously validate these externally. If ML tools are operated correctly and are validated, they could potentially be used to "augment" traditional lab testing methods, and thus help identify an individual at higher risk of developing T2DM at an earlier stage.

Acknowledgement

None.

Conflict of Interest

None.

References

1. American Diabetes Association Professional Practice Committee for Diabetes (2026) Diagnosis and classification of diabetes: Standards of Care in Diabetes-2026. Diabetes Care1; 49(Supplement_1): S27-S49.
2. Arslan A K, Yagin F H, Algarni A, Karaaslan E, Al Hashem F, et al. (2024) Enhancing type 2 diabetes mellitus prediction by integrating metabolomics and tree-based boosting approaches. Front Endocrinol 15: 1444282.
3. Cardozo G, Tirloni S F, Pereira Moro A R, Marques J LB (2022) Use of artificial intelligence in the search for new information through routine laboratory tests: systematic review. JMIR bioinform biotechnol 23; 3(1): e40473.
4. Daniels J, Herrero P, Georgiou P (2022) A multitask learning approach to personalized blood glucose prediction. IEEE J Biomed Health Inform 26(1): 436-445.
5. Desilva K L K, Wai Kit Lee, Andrew Forbes, Ryan T Demmer, Christopher Barton (2020) Use and performance of machine learning models for type 2 diabetes prediction in community settings: a systematic review and meta-analysis. Int J Med Inform 143: 104-268.
6. Fregoso Aparicio L, Noguez J, Montesinos L, Garcia Garcia JA (2021) Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. Diabetol Metab syndr 13(1): 148.
7. Jabara M, Kose O, Perlman G, Simon Corcos, Marc Antoine Pelletier, et al. (2024) Artificial intelligence-based digital biomarkers for type 2 diabetes: A review. Can J Cardiol 40(10): 1922-1933.
8. Li H, Su D, Zhang X, He Y, Luo X, et al. (2024) Machine learning-based prediction of diabetic patients using blood routine data. Methods 229: 156-162.
9. Lv X, Luo J, Huang W, Hui Guo, Xue Bai, et al. (2024) Identifying diagnostic indicators for type 2 diabetes mellitus from physical examination using interpretable machine learning approach. Front Endocrinol 15: 1376220.
10. Mansoori A, Sahranavard T, Hosseini Z S, Soflaei S S, Emrani N, et al. (2023) Prediction of type 2 diabetes mellitus using hematological factors based on machine learning approaches: a cohort study analysis. Sci Rep 13(1): 663.
11. Mohsen F, Al Absi H R, Yousri N A, El Hajj N, Shah Z (2023) A scoping review of artificial intelligence-based methods for diabetes risk prediction. NPJ Digital Medicine 25; 6(1): 197.
12. Olisah C C, Smith L, Smith M (2022) Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. Comp Methods Programs Biomed 220: 106-773.
13. Olusanya MO, Ogunsakin RE, Ghai M, Adeleke M A (2022) Accuracy of machine learning classification models for the prediction of type 2 diabetes mellitus: A systematic survey and meta-analysis approach. Int J Environ Res public health 1; 19(21): 14280.
14. Rabbani N, Kim G Y E, Suarez C J, Chen J H (2022) Applications of machine learning in routine laboratory medicine: Current state and future directions. Clin Biochem 103: 1-7.
15. Shin J, Kim J, Lee C, Yoon J Y, Kim S, et al. (2022) Development of various diabetes prediction models using machine learning techniques. Diabetes Metabol J 46(4): 650-657.
16. Wadghiri M Z, Idri A, El Idrissi T, Hakkoum H (2022) Ensemble blood glucose prediction in diabetes mellitus: A review. Comput Biol Med 147: 105674.
17. Witte H, Nakas C, Bally L, Leichtle A B (2022) Machine learning prediction of hypoglycemia and hyperglycemia from electronic health records: algorithm development and validation. JMIR formative research 6(7): e36176.
18. Zheng Y, Ley S H, Hu F B (2018) Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 14(2): 88-98.