KMANPUB | Health Nexus Vol. 4 No. 4 (2026): 1-17

Article history:
Received 14 July 2026
Revised 17 September 2026
Accepted 24 September 2026
Published online 01 October 2026

Interpretable Machine Learning-Based Multiclass Prediction of Vascular Complication Profiles in Type 2 Diabetes: A Comparative Study

Hiva Nargesi Khoramabadi[1*], Sedigheh Vaezifar[1]

1. Department of Medical Engineering, Payame Noor University, Tehran 19395-3697, Iran

*** Corresponding author email address:** Nargesi.hiva@gmail.com

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Nargesi Khoramabadi, H. & Vaezifar, S. (2026). Interpretable Machine Learning-Based Multiclass Prediction of Vascular Complication Profiles in Type 2 Diabetes: A Comparative Study. *Health Nexus, 4*(4), 1-17.
https://doi.org/10.61838/kman.hn.6657

© 2026 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

A B S T R A C T

Vascular complications are common and heterogeneous in type 2 diabetes, yet multiclass classification is challenging because of substantial class imbalance. This study evaluated interpretable machine-learning approaches for characterizing microvascular and macrovascular complication profiles. After removing duplicated clinical profiles, 1,068 profiles were classified into four categories: no complications, microvascular only, macrovascular only, and both. Multiple machine-learning models and imbalance-handling strategies were compared using stratified five-fold cross-validation and an independent test set. Performance was assessed using macro-F1, balanced accuracy, class-specific metrics, bootstrap confidence intervals, and SHAP-based feature interpretation. SMOTENC achieved the strongest overall balance between minority-class performance and discrimination, with a cross-validation macro-F1 of 0.273 ± 0.011 and balanced accuracy of 0.294 ± 0.026. On the independent test set, macro-F1 was 0.277 (95% CI, 0.213–0.343) and balanced accuracy was 0.299 (95% CI, 0.220–0.385). The macrovascular-only class remained particularly difficult to distinguish. Medication adherence, smoking, physical activity, sex, and medication use were among the most influential features. Machine-learning classification of vascular complication profiles remains challenging in this imbalanced dataset. Nevertheless, explicit imbalance handling and class-balanced evaluation provided a more informative assessment of model performance than accuracy alone. These findings support the use of interpretable, imbalance-aware frameworks for future studies and highlight the need for larger, longitudinal, and externally validated datasets before clinical application.

Keywords: *Vascular Complication, Type 2 Diabetes, Machine Learning Analysis, Healthcare.*

1. Introduction

Type 2 diabetes mellitus (T2DM) is a chronic metabolic disorder associated with a broad spectrum of vascular complications that substantially increase morbidity, disability, healthcare utilization, and premature mortality. Although persistent hyperglycemia is central to the pathophysiology of diabetes, the clinical consequences of T2DM extend well beyond glucose dysregulation and involve progressive damage to multiple vascular beds and organ systems. Diabetes-related complications are commonly categorized into microvascular and macrovascular manifestations. Microvascular complications principally include diabetic retinopathy, nephropathy, and peripheral neuropathy, whereas macrovascular complications encompass cardiovascular disease, cerebrovascular disease, and peripheral vascular disease. These conditions frequently coexist, interact, and accumulate over the course of the disease, producing heterogeneous clinical profiles that cannot always be adequately represented by a single binary outcome. The interrelationship among retinopathy, nephropathy, and neuropathy also reflects shared metabolic, inflammatory, endothelial, and microcirculatory mechanisms, highlighting the systemic nature of diabetic vascular injury (1). Consequently, identifying clinically meaningful patterns of complication burden is essential for more precise risk stratification, surveillance, and management.

The vascular burden associated with T2DM is influenced by several demographic and disease-related characteristics. Age at diabetes onset, for example, may have important implications for long-term vascular exposure, disease duration, and cumulative cardiovascular risk. Evidence from systematic review and dose-response meta-analysis indicates that an earlier age at onset of T2DM is associated with a greater burden of vascular disease and heart failure, suggesting that the temporal characteristics of diabetes may contribute meaningfully to subsequent complication profiles (2). At the same time, vascular complications do not occur in isolation. Patients may exhibit no recorded complications, develop only microvascular abnormalities, present predominantly with macrovascular disease, or simultaneously experience both forms of vascular involvement. This heterogeneity presents a clinically relevant classification problem because patients who share the same diagnosis of T2DM may nevertheless differ substantially in complication patterns, prognosis, therapeutic needs, and preventive priorities.

Traditional statistical methods have played an essential role in identifying risk factors for diabetic complications; however, the increasing availability of multidimensional clinical data has stimulated interest in machine learning (ML) as a complementary analytical framework. ML algorithms can examine complex relationships among demographic, metabolic, behavioral, treatment-related, and disease-history variables without requiring every interaction or nonlinear association to be prespecified. Such flexibility may be particularly useful in diabetes, where vascular outcomes result from the interaction of multiple risk domains rather than a single determinant. Studies applying ML to T2DM have demonstrated the feasibility of predicting several complications simultaneously using routinely collected clinical information. Nicolucci et al., for example, showed that ML approaches could be used to estimate the occurrence of multiple T2DM complications from clinical data, supporting the potential value of data-driven methods for multidimensional complication assessment (3). Subsequent work has increasingly extended these approaches toward both complication-specific and multi-outcome prediction.

The literature on diabetic complications now includes ML applications across several major vascular outcomes. Diabetic peripheral neuropathy has received substantial attention because its occurrence depends on multiple interacting metabolic and clinical factors. A recent systematic review demonstrated growing interest in ML-based prediction of diabetic peripheral neuropathy and documented substantial methodological diversity in predictors, algorithms, validation procedures, and model performance (4). Similar developments have occurred in diabetic nephropathy, where ML techniques have been investigated for identifying individuals at elevated risk based on clinical, laboratory, and demographic characteristics. A review of the literature on diabetic nephropathy prediction concluded that ML offers considerable potential for integrating heterogeneous risk information, although differences in study design, data sources, feature selection, and validation continue to limit direct comparison between models (5). These findings illustrate both the promise and

Health Nexus
E-ISSN: 2981-2569

methodological complexity of applying ML to diabetes-related microvascular disease.

Artificial intelligence has also been increasingly applied to macrovascular outcomes among people with T2DM. A systematic review and meta-analysis evaluating AI models for stroke, cardiovascular disease, and peripheral vascular disease demonstrated the growing use of computational approaches to estimate major vascular risks in this population (6). At a broader level, Schallmoser et al. evaluated ML for predicting both microvascular and macrovascular complications among individuals with prediabetes or diabetes, demonstrating that routinely available clinical information can support automated complication-risk modeling across different vascular domains (7). Likewise, interpretable ML models have been developed to assess risks across several diabetes-related complications, with increasing attention to the need not only to achieve discrimination but also to explain how individual predictors contribute to model output (8). These studies support a transition from single-complication modeling toward more comprehensive representations of diabetic vascular burden.

Recent advances have also broadened the range of information that can be incorporated into diabetes-related artificial intelligence systems. Contemporary models increasingly integrate conventional clinical variables with imaging, physiological signals, and molecular data. Image-based approaches have been developed for identifying diabetes-related pathological changes, including applications to diabetic foot assessment. For example, ML methods using RGB and thermal imaging have been investigated for classification of diabetic plantar foot conditions, illustrating the potential of computer vision to identify complications that may not be adequately captured through conventional tabular clinical variables alone (9). Other work has demonstrated that less conventional biomarkers may also provide clinically relevant information. Voice characteristics, for instance, have been investigated as potential biomarkers for detecting hypoglycemia using ML, indicating that physiological consequences of glucose disturbances may be reflected in complex signal patterns accessible to computational analysis (10). These developments indicate that diabetes-related AI is expanding from traditional structured clinical datasets toward multimodal prediction systems.

At the most advanced end of this trajectory, artificial intelligence is increasingly being applied across the continuum from routine clinical variables to molecular and omics information. Xie et al. described the emerging role of AI in predicting diabetic complications using data ranging from clinical characteristics to molecular omics, emphasizing the potential for algorithms to integrate heterogeneous biological and clinical information into individualized risk assessment (11). This direction is consistent with the broader movement toward precision medicine, in which prediction models are expected to identify clinically meaningful heterogeneity rather than merely estimate average population-level associations. Nevertheless, increasing model complexity also creates important challenges related to transparency, reproducibility, bias, validation, and clinical interpretability.

The expansion of ML applications in diabetes has therefore been accompanied by growing scrutiny of whether algorithmic methods actually outperform conventional statistical approaches. Evidence remains mixed. A recent systematic review comparing ML with traditional statistical methods for predicting diabetes-related complications found that ML can provide predictive advantages in some settings, particularly when complex nonlinear relationships are present, but the magnitude and consistency of improvement depend strongly on study design, dataset characteristics, model development procedures, and validation quality (12). Earlier systematic evaluation similarly showed substantial heterogeneity across ML models developed for diabetes complications and identified methodological limitations that could restrict their clinical translation (13). Thus, the use of sophisticated algorithms alone is insufficient to establish clinical utility. Appropriate outcome definition, transparent preprocessing, robust validation, clinically relevant performance metrics, and careful interpretation remain fundamental.

One of the most important methodological challenges in diabetes complication modeling is class imbalance. In clinical datasets, severe or specific complication phenotypes may be substantially less common than other outcome categories. Standard learning algorithms optimized for overall classification performance may therefore

Health Nexus
E-ISSN: 2981-2569

preferentially predict majority classes while performing poorly for clinically important minority groups. This problem becomes especially consequential when the research objective involves multiclass classification rather than a simple binary distinction. A model may achieve apparently acceptable overall accuracy while having little ability to identify rare complication profiles. Simulation evidence has further demonstrated that class-imbalance corrections are not universally beneficial and may introduce harms or distortions depending on the data-generating process, modeling method, and evaluation strategy (14). Accordingly, imbalance handling should be treated as an empirical modeling question rather than as an automatic preprocessing step.

The choice of performance measure is equally important in imbalanced multiclass settings. Overall accuracy may be dominated by large outcome classes and can therefore provide an overly optimistic view of model performance when minority-class recognition is poor. Metrics such as macro-F1, balanced accuracy, class-specific recall, precision, and F1-score provide more direct information about whether predictive performance is distributed across all clinically relevant classes. This issue is particularly important when the categories correspond to distinct vascular complication phenotypes, because poor performance in a small but clinically consequential group may be obscured by acceptable performance in common categories. Consequently, studies of imbalanced clinical outcomes require evaluation strategies that explicitly account for class prevalence and report class-specific error structures rather than relying on a single global metric.

Another critical requirement is model interpretability. Complex ML models may identify predictive patterns that are difficult to translate into clinically understandable reasoning. This limitation has contributed to growing interest in explanation methods such as SHapley Additive exPlanations (SHAP), which estimate the contribution of individual features to a model's predictions. SHAP analysis has become increasingly prominent in biomedical ML because it provides both global information regarding influential predictors and local explanations for individual predictions. Practical guidance on SHAP emphasizes its utility for interpreting supervised ML models while also cautioning that feature attribution does not itself establish causality (15). In diabetes research, this distinction is particularly important because features such as medication adherence, smoking, physical activity, glycemic measures, disease duration, and treatment patterns may be associated with both disease severity and clinical management. Interpretability methods can therefore clarify how a fitted model uses available information, but they should not be interpreted as proving independent causal effects.

The increasing clinical use of AI has also resulted in stronger methodological and reporting standards. The TRIPOD+AI statement extends reporting guidance for clinical prediction models using regression or ML methods and emphasizes transparent description of data sources, participants, outcomes, predictors, sample size, missing data, model development, performance assessment, and validation (16). Complementing this reporting framework, PROBAST+AI provides updated guidance for assessing risk of bias, quality, and applicability in prediction models developed using regression or artificial intelligence approaches (17). Together, these frameworks reinforce the principle that predictive performance should not be interpreted independently of the methodological conditions under which it was obtained. For diabetes complication models, this means that data provenance, outcome timing, class distribution, resampling procedures, internal validation, external validation, and uncertainty must be explicitly considered before claims of clinical applicability are made.

These concerns are particularly relevant because recent reviews of ML for diabetes complications continue to report limitations in validation and generalizability. Although ML has been used for diagnosis and classification across diabetic retinopathy, nephropathy, neuropathy, diabetic foot disease, and macrovascular complications, the quality of evidence varies considerably across applications (18). Many studies remain retrospective, employ relatively small samples, use heterogeneous definitions of complications, or lack independent external validation. These limitations can produce optimistic performance estimates and make it difficult to determine whether a model will generalize across populations, institutions, or clinical settings. Accordingly, contemporary research increasingly emphasizes not only predictive performance but also robustness, transparency, fairness, uncertainty, and clinical plausibility.

E-ISSN: 2981-2569

The distinction between static classification and dynamic prediction is also important. Many datasets contain information on complications measured at the same time as the predictor variables and therefore support classification of observed clinical phenotypes rather than prediction of future incident outcomes. By contrast, dynamic longitudinal models can update predicted risks as new patient information becomes available. McCoy et al. recently developed and validated dynamic models for predicting acute and chronic diabetes complications among adults with newly diagnosed diabetes, demonstrating the potential of longitudinally updated clinical information for individualized risk estimation (19). Such approaches represent an important direction for future clinical decision support, but they also highlight the need to distinguish clearly between contemporaneous phenotype classification and prospective risk forecasting.

The availability of publicly accessible datasets has contributed substantially to methodological development in this field by allowing researchers to compare algorithms, preprocessing strategies, and evaluation frameworks under reproducible conditions. The dataset released by Vamsi and Bhattacharyya includes clinical, metabolic, behavioral, treatment-related, microvascular, and macrovascular variables relevant to T2DM and therefore offers an opportunity to examine vascular complication patterns using supervised learning (20). However, secondary analysis of publicly available clinical data also requires careful attention to data provenance, repeated records, variable coding, class structure, and the limitations imposed by the original dataset design. These issues are particularly important when observed complication indicators are combined to generate multiclass phenotypes.

Taken together, the current literature demonstrates substantial progress in AI-based assessment of diabetic complications but also identifies several unresolved methodological issues. Previous studies have frequently focused on individual complications, binary outcomes, or aggregate risk prediction, whereas the simultaneous differentiation of clinically distinct microvascular and macrovascular complication profiles poses a more difficult multiclass problem. In such settings, substantial class imbalance can obscure poor recognition of minority phenotypes, conventional accuracy can misrepresent performance, and highly flexible models may remain difficult to interpret. Furthermore, although resampling and class-weighting approaches may improve minority-class representation, their effects must be evaluated on untouched data rather than assumed to improve clinical discrimination. Interpretable modeling and class-specific performance analysis are therefore essential for establishing what an algorithm can and cannot distinguish.

Accordingly, this study aimed to compare interpretable machine-learning approaches and class-imbalance handling strategies for multiclass classification of four observed vascular complication profiles—no complications, microvascular complications only, macrovascular complications only, and combined microvascular and macrovascular complications—among adults with type 2 diabetes, using class-balanced performance measures, independent test-set evaluation, and SHAP-based interpretation.

2. Methods and Materials

2.1. Study design and data source

This retrospective cross-sectional machine-learning study evaluated the classification of vascular complication profiles among adults with type 2 diabetes using a publicly available dataset deposited in Mendeley Data by Bandi Vamsi and Debnath Bhattacharyya. The dataset contains demographic, anthropometric, clinical, metabolic, behavioral, treatment-related and complication variables, including age, body mass index, systolic and diastolic blood pressure, HbA1c, fasting and postprandial plasma glucose, family history, age at diabetes onset, smoking, physical activity, medication use and medication adherence, together with nephropathy, neuropathy, retinopathy, cardiovascular and peripheral vascular indicators.

The original file contained 3,068 rows and 22 columns. SL.NO and NAME were treated as identifiers and were excluded from model predictors. An audit of the remaining clinical and outcome fields identified 2,000 exact repeated profiles, leaving 1,068 distinct observed clinical profiles. Because the dataset does not provide a complete, independently verifiable patient identifier or provenance statement establishing whether repeated profiles represent duplicate records or different individuals with identical

Health Nexus
E-ISSN: 2981-2569

recorded values, the primary analysis refers to these observations as unique clinical profiles rather than verified unique patients.

Table 1

Predictor variables included in the machine-learning analysis.

Variable	Abbreviation	Variable category	Description/role in analysis
Age	AGE	Demographic	Patient age
Sex	SEX	Demographic	Biological sex
Body mass index	BMI	Anthropometric	Body mass index
Systolic pressure	SP	Clinical/cardiovascular	Systolic blood pressure or pressure-related measurement, as defined in the dataset
Blood pressure	BP	Clinical/cardiovascular	Blood pressure measurement, as defined in the dataset
Glycated hemoglobin	HbA1c	Metabolic	Indicator of long-term glycemic control
Fasting plasma sugar	FPS	Metabolic	Fasting glucose measurement
Postprandial plasma sugar	PPS	Metabolic	Postprandial glucose measurement
Family history	FAMILY H/O	Clinical/familial	Family history of diabetes
Age at diabetes onset	ONSET AGE	Clinical	Age at onset of diabetes
Diabetes life/duration	DIA LIFE	Disease history	Diabetes duration variable
Smoking	SMOKING	Behavioral	Smoking status
Physical activity	PHY ACT	Behavioral	Physical activity status
Medication use	MED USE	Treatment-related	Medication utilization
Medication adherence	MED ADH	Treatment-related	Treatment adherence

2.2. Outcome construction

Microvascular involvement was defined as the presence of at least one positive indicator among NEP, NEU or RET. Macrovascular involvement was defined as the presence of at least one positive indicator among CV or PER VAS. Combining these two binary dimensions yielded four mutually exclusive classes: no recorded vascular complication, microvascular complications only, macrovascular complications only, and both microvascular and macrovascular complications.

The resulting distribution was 69 profiles (6.46%) with neither complication type, 491 (45.97%) with microvascular complications only, 63 (5.90%) with macrovascular complications only, and 445 (41.67%) with both complication types. Thus, the largest-to-smallest class ratio was approximately 7.8:1, indicating substantial multiclass imbalance.

2.3. Predictor preprocessing

Numerical predictors were median-imputed and categorical predictors were mode-imputed within the modelling pipeline. Variables coded categorically, including SEX, FAMILY H/O, SMOKING, PHY ACT, MED USE and MED ADH, were treated as categorical rather than ordinal in the absence of a verified codebook establishing ordinal meaning. DIA LIFE was parsed as years for numeric entries and explicit month expressions were converted to fractions of a year; because the original data dictionary was not supplied, this coding assumption requires confirmation from the data provider and was designated for sensitivity analysis.

Health Nexus
E-ISSN: 2981-2569

2.4. *Data partitioning and validation*

The 1,068-profile analytic dataset was divided using a stratified 80:20 development/test split with a fixed random seed, yielding 854 training profiles and 214 independent test profiles. The test set was locked before model comparison and was not used for model selection, hyperparameter tuning or resampling. Within the training data, stratified five-fold cross-validation was used for comparative evaluation. The primary model-selection metric was macro-F1 because it weights the four classes equally and therefore prevents the two large classes from dominating the model-selection criterion. Balanced accuracy, macro-precision, macro-recall, weighted F1 and accuracy were retained as secondary measures.

2.5. *Class-imbalance strategies*

Five imbalance configurations were compared: (1) no resampling, (2) inverse-frequency class weighting, (3) random oversampling of minority classes, (4) SMOTENC, and (5) Balanced Random Forest. All resampling operations were restricted to the training portion of each cross-validation fold and were never applied to validation or test observations. SMOTENC was selected rather than conventional SMOTE because the predictor set contains both numerical and categorical variables; SMOTE-NC was originally developed to extend synthetic oversampling to mixed nominal and continuous data. The independent test set retained the observed class distribution so that evaluation reflected performance on non-synthetic observations.

2.6. *Candidate algorithms and interpretability*

Candidate models included a majority-class Dummy classifier, multinomial logistic regression, RBF support vector machine, random forest, ExtraTrees, Balanced Random Forest and LightGBM. ExtraTrees was used as the principal tree-ensemble framework for the imbalance-strategy comparison. Model interpretation was performed using SHAP (SHapley Additive exPlanations), which assigns feature-level attribution values to model predictions. SHAP values were interpreted as model-specific predictive attributions rather than causal effects.

2.7. *Performance measures and uncertainty*

For the independent test set, accuracy, balanced accuracy, macro-precision, macro-recall, macro-F1, weighted F1, one-vs-rest macro ROC-AUC and macro average precision were calculated. Class-specific precision, recall, F1-score, ROC-AUC and average precision were also reported. Ninety-five percent bootstrap confidence intervals for macro-F1 and balanced accuracy were estimated from the test predictions. Because the two minority classes contained only 13 and 14 test observations, their class-specific estimates were interpreted as imprecise and were not treated as definitive evidence of clinical discrimination.

3. Findings and Results

3.1. *Class distribution and analytic sample*

The final analytic cohort comprised 1,068 distinct observed clinical profiles after exclusion of 2,000 exact repeated profiles from the original 3,068-row file. The distribution was strongly asymmetric, with microvascular-only and both-complication profiles accounting for 45.97% and 41.67% of profiles, respectively, whereas no-complication and macrovascular-only profiles accounted for only 6.46% and 5.90%. This distribution justified explicit evaluation of imbalance-aware learning strategies and the use of class-balanced performance measures.

Health Nexus
E-ISSN: 2981-2569

Figure 1

Analytical workflow and data handling.

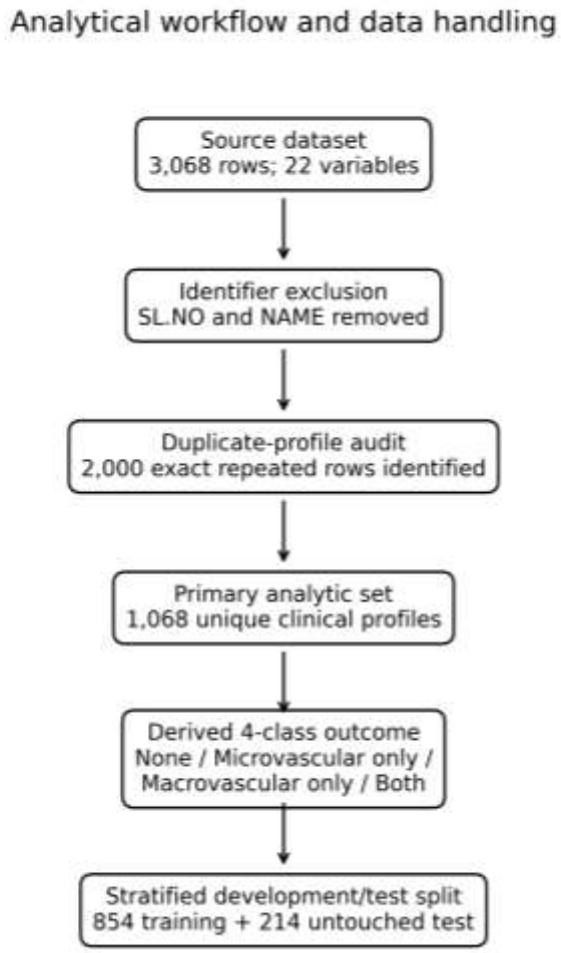
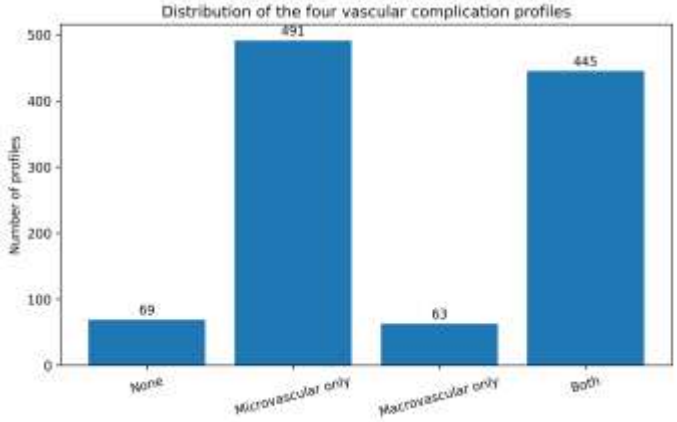


Figure 2

Distribution of the four observed vascular complication profiles.

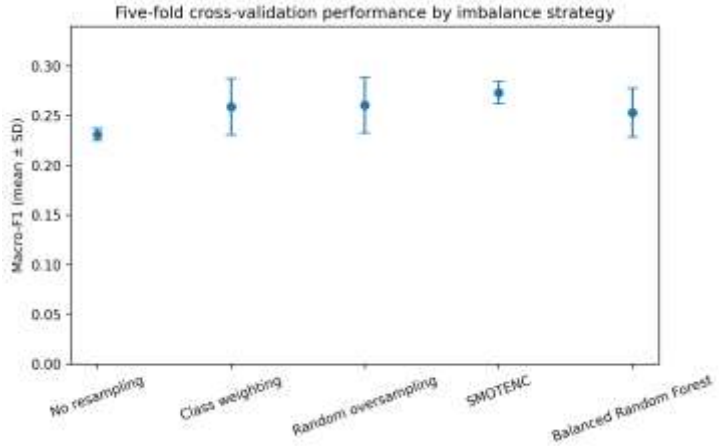


Health Nexus
E-ISSN: 2981-2569

3.2. Cross-validation comparison

Figure 3

Five-fold cross-validation Macro-F1 according to imbalance-handling strategy.



In five-fold cross-validation, SMOTENC produced the highest mean macro-F1 (0.273 ± 0.011) and the highest mean balanced accuracy (0.294 ± 0.026) among the five imbalance configurations. Random oversampling and class weighting produced smaller improvements in macro-F1 (0.260 ± 0.028 and 0.259 ± 0.028, respectively), while Balanced Random Forest achieved a mean macro-F1 of 0.253 ± 0.024. The no-resampling configuration produced the lowest macro-F1 (0.231 ± 0.006). These results support the use of imbalance-aware learning for further evaluation, but the relatively modest absolute scores indicate that resampling did not transform the problem into a high-discrimination classification task.

Table 2

Class distribution in the analytic dataset

Class	Total n	Training n	Test n
None	69	55	14
Microvascular only	491	393	98
Macrovascular only	63	50	13
Both	445	356	89

Interpretation: the two minority classes were retained in the independent test set rather than artificially equalized. This is essential because a resampled test set would no longer represent the observed distribution and could produce misleading performance estimates.

Table 3

Cross-validation comparison of imbalance-handling strategies

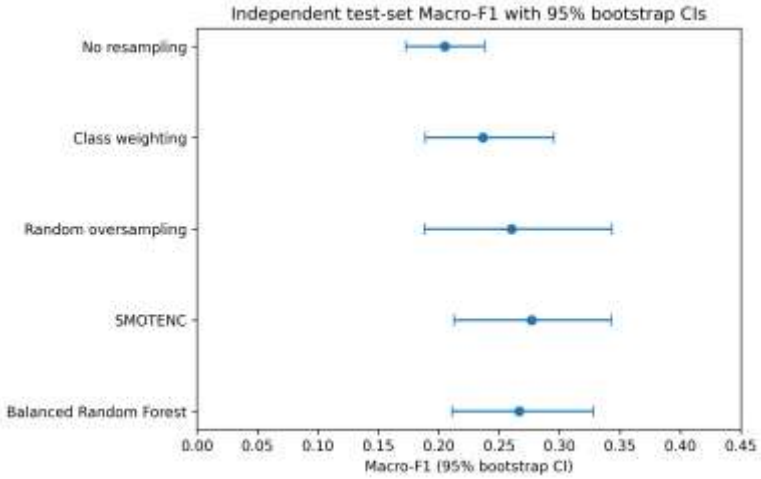
Strategy	Macro-F1	Balanced accuracy	Accuracy	Weighted F1
No resampling	0.231 ± 0.006	0.249 ± 0.006	0.440 ± 0.010	0.408 ± 0.011
Class weighting	0.259 ± 0.028	0.258 ± 0.019	0.405 ± 0.008	0.393 ± 0.006
Random oversampling	0.260 ± 0.028	0.260 ± 0.021	0.422 ± 0.020	0.405 ± 0.018
SMOTENC	0.273 ± 0.011	0.294 ± 0.026	0.372 ± 0.021	0.391 ± 0.019
Balanced Random Forest	0.253 ± 0.024	0.270 ± 0.029	0.334 ± 0.035	0.358 ± 0.039

Health Nexus
E-ISSN: 2981-2569

3.3. Independent test-set performance

Figure 4

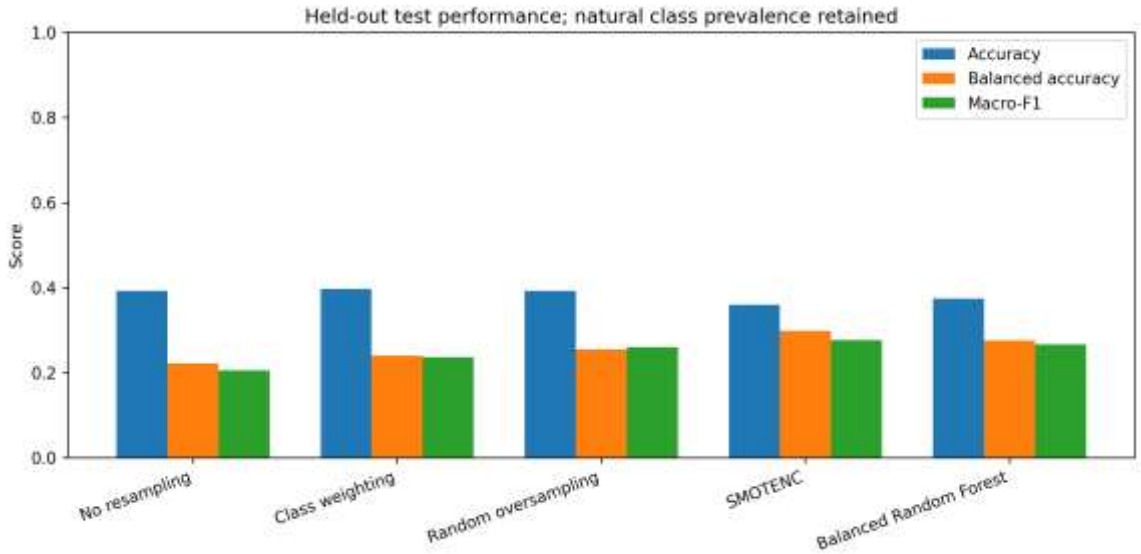
Macro-F1 estimates on the untouched test set with 95% bootstrap confidence intervals.



On the untouched test set, SMOTENC achieved the highest macro-F1 (0.277; 95% bootstrap CI 0.213–0.343) and balanced accuracy (0.299; 95% CI 0.220–0.385). Balanced Random Forest yielded a slightly lower macro-F1 of 0.267 (95% CI 0.211–0.328) and balanced accuracy of 0.276 (95% CI 0.208–0.359), while random oversampling achieved macro-F1 0.260 (95% CI 0.188–0.343). Class weighting produced macro-F1 0.237 (95% CI 0.189–0.295), and the no-resampling model produced 0.205 (95% CI 0.173–0.238). The confidence intervals overlap substantially, so the results do not establish a statistically definitive superiority of SMOTENC over the other imbalance-aware strategies.

Figure 5

Held-out test performance



Health Nexus
E-ISSN: 2981-2569

Table 4

Independent test-set performance

Strategy	Accuracy	Balanced accuracy	Macro precision	Macro recall	Macro-F1	Weighted F1	Macro OvR AUC
No resampling	0.393	0.222	0.193	0.222	0.205	0.361	0.547
Class weighting	0.397	0.240	0.238	0.240	0.237	0.384	0.545
Random oversampling	0.393	0.255	0.295	0.255	0.260	0.379	0.552
SMOTENC	0.360	0.299	0.274	0.299	0.277	0.373	0.552
Balanced Random Forest	0.374	0.276	0.271	0.276	0.267	0.392	0.579

Interpretation: the imbalance-aware strategies improved macro-level performance relative to the unbalanced baseline, but this came with a reduction in overall accuracy for SMOTENC. This is expected when the classifier is encouraged to allocate more decision capacity to minority classes. Accordingly, accuracy alone would have favoured class weighting, whereas macro-F1 and balanced accuracy identify SMOTENC as the strongest compromise for class-balanced discrimination in this analysis.

3.4. Class-specific performance and error structure

Figure 6

Confusion matrix for SMOTENC on the untouched test set.

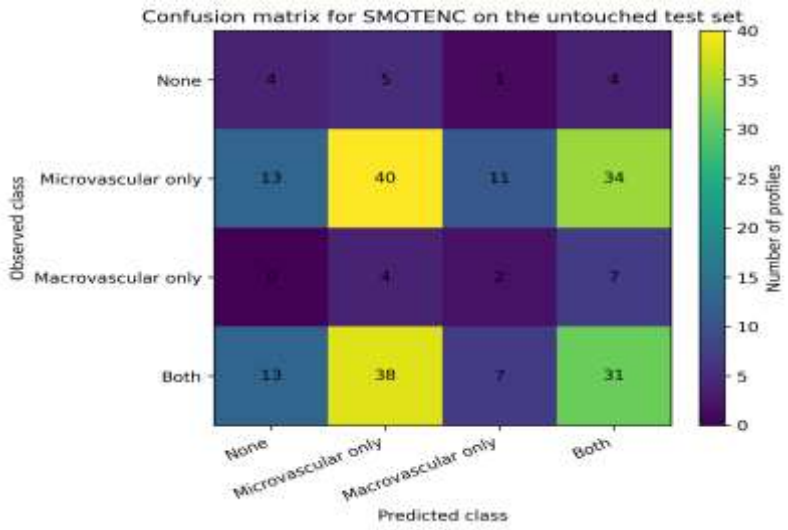


Table 5

Class-specific performance and error structure

Class	Support	Precision	Recall	F1	ROC-AUC	Average precision
None	14	0.133	0.286	0.182	0.745	0.130
Microvascular only	98	0.460	0.408	0.432	0.495	0.473
Macrovascular only	13	0.095	0.154	0.118	0.484	0.137
Both	89	0.408	0.348	0.376	0.485	0.396

Class-specific evaluation demonstrated that the apparent improvement in macro-level performance was not uniform across the four classes. For SMOTENC, recall was 0.286 for profiles with no recorded complications, 0.408 for microvascular-only profiles, 0.154 for macrovascular-only profiles and 0.348 for profiles with both complication types.

Health Nexus
E-ISSN: 2981-2569

The macrovascular-only class therefore remained the principal failure mode, with only 2 of 13 test observations correctly identified. The confusion matrix showed substantial overlap between macrovascular-only and both-complication profiles, suggesting that the available predictors do not provide sufficiently stable separation of these phenotypes.

Table 6

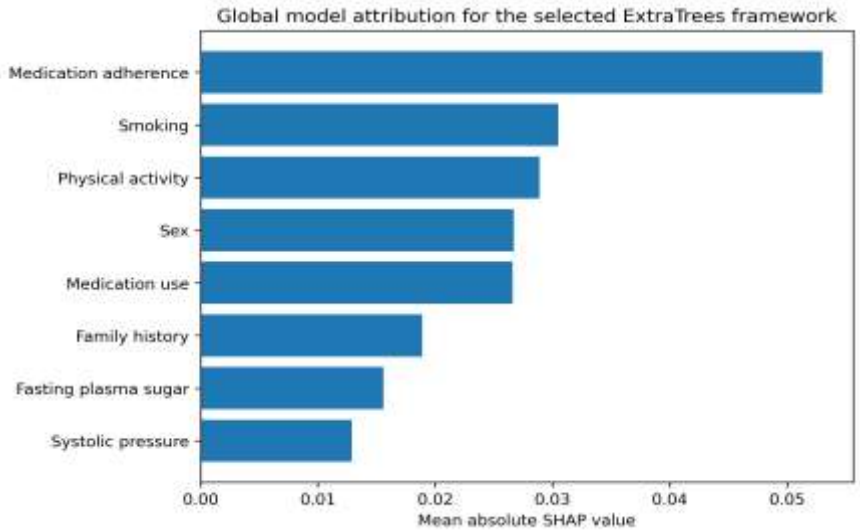
Bootstrap uncertainty for macro-level performance

Strategy	Macro-F1	95% CI	Balanced accuracy	95% CI
No resampling	0.205	0.173–0.238	0.222	0.189–0.258
Class weighting	0.237	0.189–0.295	0.240	0.196–0.296
Random oversampling	0.260	0.188–0.343	0.255	0.197–0.324
SMOTENC	0.277	0.213–0.343	0.299	0.220–0.385
Balanced Random Forest	0.267	0.211–0.328	0.276	0.208–0.359

3.5. Interpretability analysis

Figure 7

Global mean absolute SHAP attribution for the selected tree-ensemble framework.



The global SHAP analysis identified medication adherence (mean absolute SHAP 0.0530), smoking (0.0305), physical activity (0.0289), sex (0.0267), medication use (0.0266), family history (0.0189), fasting plasma sugar (0.0156) and systolic pressure (0.0129) among the leading model attributions. These values quantify the contribution of features to model output within the fitted model; they do not establish causal effects, independent risk factors, or clinical directionality. Given the modest overall discrimination and the small minority classes, the SHAP results should therefore be considered exploratory.

4. Discussion and Conclusion

Type 2 diabetes mellitus (T2DM) is a chronic metabolic disorder associated with a broad spectrum of vascular complications that substantially increase morbidity, disability, healthcare utilization, and premature mortality. Although persistent hyperglycemia is central to the pathophysiology of diabetes, the clinical consequences of T2DM extend well beyond glucose dysregulation and involve progressive damage to multiple vascular beds and organ systems. Diabetes-related complications are commonly categorized into microvascular and macrovascular manifestations. Microvascular complications

Health Nexus
E-ISSN: 2981-2569

principally include diabetic retinopathy, nephropathy, and peripheral neuropathy, whereas macrovascular complications encompass cardiovascular disease, cerebrovascular disease, and peripheral vascular disease. These conditions frequently coexist, interact, and accumulate over the course of the disease, producing heterogeneous clinical profiles that cannot always be adequately represented by a single binary outcome. The interrelationship among retinopathy, nephropathy, and neuropathy also reflects shared metabolic, inflammatory, endothelial, and microcirculatory mechanisms, highlighting the systemic nature of diabetic vascular injury (1). Consequently, identifying clinically meaningful patterns of complication burden is essential for more precise risk stratification, surveillance, and management.

The vascular burden associated with T2DM is influenced by several demographic and disease-related characteristics. Age at diabetes onset, for example, may have important implications for long-term vascular exposure, disease duration, and cumulative cardiovascular risk. Evidence from systematic review and dose-response meta-analysis indicates that an earlier age at onset of T2DM is associated with a greater burden of vascular disease and heart failure, suggesting that the temporal characteristics of diabetes may contribute meaningfully to subsequent complication profiles (2). At the same time, vascular complications do not occur in isolation. Patients may exhibit no recorded complications, develop only microvascular abnormalities, present predominantly with macrovascular disease, or simultaneously experience both forms of vascular involvement. This heterogeneity presents a clinically relevant classification problem because patients who share the same diagnosis of T2DM may nevertheless differ substantially in complication patterns, prognosis, therapeutic needs, and preventive priorities.

Traditional statistical methods have played an essential role in identifying risk factors for diabetic complications; however, the increasing availability of multidimensional clinical data has stimulated interest in machine learning (ML) as a complementary analytical framework. ML algorithms can examine complex relationships among demographic, metabolic, behavioral, treatment-related, and disease-history variables without requiring every interaction or nonlinear association to be prespecified. Such flexibility may be particularly useful in diabetes, where vascular outcomes result from the interaction of multiple risk domains rather than a single determinant. Studies applying ML to T2DM have demonstrated the feasibility of predicting several complications simultaneously using routinely collected clinical information. Nicolucci et al., for example, showed that ML approaches could be used to estimate the occurrence of multiple T2DM complications from clinical data, supporting the potential value of data-driven methods for multidimensional complication assessment (3). Subsequent work has increasingly extended these approaches toward both complication-specific and multi-outcome prediction.

The literature on diabetic complications now includes ML applications across several major vascular outcomes. Diabetic peripheral neuropathy has received substantial attention because its occurrence depends on multiple interacting metabolic and clinical factors. A recent systematic review demonstrated growing interest in ML-based prediction of diabetic peripheral neuropathy and documented substantial methodological diversity in predictors, algorithms, validation procedures, and model performance (4). Similar developments have occurred in diabetic nephropathy, where ML techniques have been investigated for identifying individuals at elevated risk based on clinical, laboratory, and demographic characteristics. A review of the literature on diabetic nephropathy prediction concluded that ML offers considerable potential for integrating heterogeneous risk information, although differences in study design, data sources, feature selection, and validation continue to limit direct comparison between models (5). These findings illustrate both the promise and methodological complexity of applying ML to diabetes-related microvascular disease.

Artificial intelligence has also been increasingly applied to macrovascular outcomes among people with T2DM. A systematic review and meta-analysis evaluating AI models for stroke, cardiovascular disease, and peripheral vascular disease demonstrated the growing use of computational approaches to estimate major vascular risks in this population (6). At a broader level, Schallmoser et al. evaluated ML for predicting both microvascular and macrovascular complications among individuals with prediabetes or diabetes, demonstrating that routinely

E-ISSN: 2981-2569

available clinical information can support automated complication-risk modeling across different vascular domains (7). Likewise, interpretable ML models have been developed to assess risks across several diabetes-related complications, with increasing attention to the need not only to achieve discrimination but also to explain how individual predictors contribute to model output (8). These studies support a transition from single-complication modeling toward more comprehensive representations of diabetic vascular burden.

Recent advances have also broadened the range of information that can be incorporated into diabetes-related artificial intelligence systems. Contemporary models increasingly integrate conventional clinical variables with imaging, physiological signals, and molecular data. Image-based approaches have been developed for identifying diabetes-related pathological changes, including applications to diabetic foot assessment. For example, ML methods using RGB and thermal imaging have been investigated for classification of diabetic plantar foot conditions, illustrating the potential of computer vision to identify complications that may not be adequately captured through conventional tabular clinical variables alone (9). Other work has demonstrated that less conventional biomarkers may also provide clinically relevant information. Voice characteristics, for instance, have been investigated as potential biomarkers for detecting hypoglycemia using ML, indicating that physiological consequences of glucose disturbances may be reflected in complex signal patterns accessible to computational analysis (10). These developments indicate that diabetes-related AI is expanding from traditional structured clinical datasets toward multimodal prediction systems.

At the most advanced end of this trajectory, artificial intelligence is increasingly being applied across the continuum from routine clinical variables to molecular and omics information. Xie et al. described the emerging role of AI in predicting diabetic complications using data ranging from clinical characteristics to molecular omics, emphasizing the potential for algorithms to integrate heterogeneous biological and clinical information into individualized risk assessment (11). This direction is consistent with the broader movement toward precision medicine, in which prediction models are expected to identify clinically meaningful heterogeneity rather than merely estimate average population-level associations. Nevertheless, increasing model complexity also creates important challenges related to transparency, reproducibility, bias, validation, and clinical interpretability.

The expansion of ML applications in diabetes has therefore been accompanied by growing scrutiny of whether algorithmic methods actually outperform conventional statistical approaches. Evidence remains mixed. A recent systematic review comparing ML with traditional statistical methods for predicting diabetes-related complications found that ML can provide predictive advantages in some settings, particularly when complex nonlinear relationships are present, but the magnitude and consistency of improvement depend strongly on study design, dataset characteristics, model development procedures, and validation quality (12). Earlier systematic evaluation similarly showed substantial heterogeneity across ML models developed for diabetes complications and identified methodological limitations that could restrict their clinical translation (13). Thus, the use of sophisticated algorithms alone is insufficient to establish clinical utility. Appropriate outcome definition, transparent preprocessing, robust validation, clinically relevant performance metrics, and careful interpretation remain fundamental.

One of the most important methodological challenges in diabetes complication modeling is class imbalance. In clinical datasets, severe or specific complication phenotypes may be substantially less common than other outcome categories. Standard learning algorithms optimized for overall classification performance may therefore preferentially predict majority classes while performing poorly for clinically important minority groups. This problem becomes especially consequential when the research objective involves multiclass classification rather than a simple binary distinction. A model may achieve apparently acceptable overall accuracy while having little ability to identify rare complication profiles. Simulation evidence has further demonstrated that class-imbalance corrections are not universally beneficial and may introduce harms or distortions depending on the data-generating process, modeling method, and evaluation strategy (14). Accordingly, imbalance handling should be treated as an

Health Nexus
E-ISSN: 2981-2569

empirical modeling question rather than as an automatic preprocessing step.

The choice of performance measure is equally important in imbalanced multiclass settings. Overall accuracy may be dominated by large outcome classes and can therefore provide an overly optimistic view of model performance when minority-class recognition is poor. Metrics such as macro-F1, balanced accuracy, class-specific recall, precision, and F1-score provide more direct information about whether predictive performance is distributed across all clinically relevant classes. This issue is particularly important when the categories correspond to distinct vascular complication phenotypes, because poor performance in a small but clinically consequential group may be obscured by acceptable performance in common categories. Consequently, studies of imbalanced clinical outcomes require evaluation strategies that explicitly account for class prevalence and report class-specific error structures rather than relying on a single global metric.

Another critical requirement is model interpretability. Complex ML models may identify predictive patterns that are difficult to translate into clinically understandable reasoning. This limitation has contributed to growing interest in explanation methods such as SHapley Additive exPlanations (SHAP), which estimate the contribution of individual features to a model's predictions. SHAP analysis has become increasingly prominent in biomedical ML because it provides both global information regarding influential predictors and local explanations for individual predictions. Practical guidance on SHAP emphasizes its utility for interpreting supervised ML models while also cautioning that feature attribution does not itself establish causality (15). In diabetes research, this distinction is particularly important because features such as medication adherence, smoking, physical activity, glycemic measures, disease duration, and treatment patterns may be associated with both disease severity and clinical management. Interpretability methods can therefore clarify how a fitted model uses available information, but they should not be interpreted as proving independent causal effects.

The increasing clinical use of AI has also resulted in stronger methodological and reporting standards. The TRIPOD+AI statement extends reporting guidance for clinical prediction models using regression or ML methods and emphasizes transparent description of data sources, participants, outcomes, predictors, sample size, missing data, model development, performance assessment, and validation (16). Complementing this reporting framework, PROBAST+AI provides updated guidance for assessing risk of bias, quality, and applicability in prediction models developed using regression or artificial intelligence approaches (17). Together, these frameworks reinforce the principle that predictive performance should not be interpreted independently of the methodological conditions under which it was obtained. For diabetes complication models, this means that data provenance, outcome timing, class distribution, resampling procedures, internal validation, external validation, and uncertainty must be explicitly considered before claims of clinical applicability are made.

These concerns are particularly relevant because recent reviews of ML for diabetes complications continue to report limitations in validation and generalizability. Although ML has been used for diagnosis and classification across diabetic retinopathy, nephropathy, neuropathy, diabetic foot disease, and macrovascular complications, the quality of evidence varies considerably across applications (18). Many studies remain retrospective, employ relatively small samples, use heterogeneous definitions of complications, or lack independent external validation. These limitations can produce optimistic performance estimates and make it difficult to determine whether a model will generalize across populations, institutions, or clinical settings. Accordingly, contemporary research increasingly emphasizes not only predictive performance but also robustness, transparency, fairness, uncertainty, and clinical plausibility.

The distinction between static classification and dynamic prediction is also important. Many datasets contain information on complications measured at the same time as the predictor variables and therefore support classification of observed clinical phenotypes rather than prediction of future incident outcomes. By contrast, dynamic longitudinal models can update predicted risks as new patient information becomes available. McCoy et al. recently developed and validated dynamic models for predicting acute and chronic diabetes complications among adults with newly diagnosed diabetes, demonstrating the potential of longitudinally updated clinical information for individualized risk

Health Nexus
E-ISSN: 2981-2569

estimation (19). Such approaches represent an important direction for future clinical decision support, but they also highlight the need to distinguish clearly between contemporaneous phenotype classification and prospective risk forecasting.

The availability of publicly accessible datasets has contributed substantially to methodological development in this field by allowing researchers to compare algorithms, preprocessing strategies, and evaluation frameworks under reproducible conditions. The dataset released by Vamsi and Bhattacharyya includes clinical, metabolic, behavioral, treatment-related, microvascular, and macrovascular variables relevant to T2DM and therefore offers an opportunity to examine vascular complication patterns using supervised learning (20). However, secondary analysis of publicly available clinical data also requires careful attention to data provenance, repeated records, variable coding, class structure, and the limitations imposed by the original dataset design. These issues are particularly important when observed complication indicators are combined to generate multiclass phenotypes.

Taken together, the current literature demonstrates substantial progress in AI-based assessment of diabetic complications but also identifies several unresolved methodological issues. Previous studies have frequently focused on individual complications, binary outcomes, or aggregate risk prediction, whereas the simultaneous differentiation of clinically distinct microvascular and macrovascular complication profiles poses a more difficult multiclass problem. In such settings, substantial class imbalance can obscure poor recognition of minority phenotypes, conventional accuracy can misrepresent performance, and highly flexible models may remain difficult to interpret. Furthermore, although resampling and class-weighting approaches may improve minority-class representation, their effects must be evaluated on untouched data rather than assumed to improve clinical discrimination. Interpretable modeling and class-specific performance analysis are therefore essential for establishing what an algorithm can and cannot distinguish.

Accordingly, this study aimed to compare interpretable machine-learning approaches and class-imbalance handling strategies for multiclass classification of four observed vascular complication profiles—no complications, microvascular complications only, macrovascular complications only, and combined microvascular and macrovascular complications—among adults with type 2 diabetes, using class-balanced performance measures, independent test-set evaluation, and SHAP-based interpretation.

Authors' Contributions

All authors contributed substantially to the study and to manuscript development, and all approved the final version.

Declaration

The authors declare that artificial intelligence tools were used only to assist with language editing, translation, and improvement of the manuscript's readability. All conceptualization, study design, data collection, data analysis, interpretation of findings, and final approval of the manuscript were performed by the authors. The authors take full responsibility for the accuracy, integrity, and originality of the content.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

The authors thank the participating centers, clinicians, adolescents, and guardians who made the study possible.

Declaration of Interest

The authors report no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Considerations

The study placed a high emphasis on ethical considerations. Informed consent obtained from all participants, ensuring they are fully aware of the nature of the study and their role in it. Confidentiality strictly maintained, with data anonymized to protect individual

privacy. The study adhered to the ethical guidelines for research with human subjects as outlined in the Declaration of Helsinki.

References

1. Kulkarni A, Thool AR, Daigavane S. Understanding The Clinical Relationship Between Diabetic Retinopathy, Nephropathy, And Neuropathy: A Comprehensive Review. Cureus. 2024;16(3):e56674. [DOI]
2. Goldney J. Age At Onset Of Type 2 Diabetes And Prevalence Of Vascular Disease And Heart Failure: Systematic Review And Dose-Response Meta-Analysis. Journal of Diabetes and Its Complications. 2024. [PMID: 39213715] [DOI]
3. Nicolucci A, Romeo L, Bernardini M, Vespasiani M, Rossi MC, Petrelli M, et al. Prediction Of Complications Of Type 2 Diabetes: A Machine Learning Approach. Diabetes Research and Clinical Practice. 2022;190:110013. [PMID: 35870573] [DOI]
4. Ma Y, Wang Z, Yao Z, Lu B, He Y. Machine Learning In The Prediction Of Diabetic Peripheral Neuropathy: A Systematic Review. BMC Medical Informatics and Decision Making. 2025;25(1):344. [PMID: 41024051] [PMCID: PMC12481961] [DOI]
5. Mesquita F, Bernardino J, Henriques J, Raposo JF, Ribeiro RT, Paredes S. Machine Learning Techniques To Predict The Risk Of Developing Diabetic Nephropathy: A Literature Review. Journal of Diabetes & Metabolic Disorders. 2024;23:825-39. [PMID: 38932857] [PMCID: PMC11196462] [DOI]
6. Nur AN. Predicting The Risks Of Stroke, Cardiovascular Disease, And Peripheral Vascular Disease Among People With Type 2 Diabetes With Artificial Intelligence Models: A Systematic Review And Meta-Analysis. Narra J. 2025;5(1):e2116. [PMID: 40352176] [PMCID: PMC12059823] [DOI]
7. Schallmoser S, Zueger T, Kraus M, Saar-Tsechansky M, Stettler C, Feuerriegel S. Machine Learning For Predicting Micro- And Macrovascular Complications In Individuals With Prediabetes OR Diabetes: Retrospective Cohort Study. Journal of Medical Internet Research. 2023;25:e42181. [PMID: 36848190] [PMCID: PMC10012007] [DOI]
8. Ye S, Ye J, Ye X, Ni X. Interpretable Prediction Model For Assessing Diabetes Complication Risks In Chinese Sufferers. Diabetes Research and Clinical Practice. 2024;209:111560. [PMID: 38316188] [DOI]
9. Muscillo L, Otesteanu CF, Menon C. Machine Learning Approaches For Diabetic Plantar Foot Classification From Rgb And Thermal Images. Discover Artificial Intelligence. 2026. [DOI]
10. Lehmann V, Hilpert M, Mostaani Z, Hovsepyan S, Wallace E, Verzat C, et al. Listening To Hypoglycemia: Voice As A Biomarker For Detection Of A Medical Emergency Using Machine Learning. Diabetes Care. 2026;49(1):105-10. [PMID: 41251597] [DOI]
11. Xie X, Wu C, Huang Z, Zhou Y, Huang J, Dao F. Artificial Intelligence-Powered Prediction Of Diabetic Complications: From Clinical Data To Molecular Omics. Briefings in Bioinformatics. 2026;27(1):bbag083. [PMID: 41744227] [PMCID: PMC12936790] [DOI]
12. Ghafoor B, Zaccardi F, Khunti K, Shabnam S. Comparative Predictive Accuracy Of Machine Learning Versus Traditional Statistical Methods For Diabetes-Related Complications: A Systematic Review And Analysis. Diabetic Medicine. 2026:e70398. [PMID: 42316875] [DOI]
13. Tan KR, Seng JJB, Kwan YH. Evaluation Of Machine Learning Methods Developed For Prediction Of Diabetes Complications: A Systematic Review. Journal of Diabetes Science and Technology. 2023;17(2):474-89. [PMID: 34727783] [PMCID: PMC10012374] [DOI]
14. Carriero A, Luijken K, De Hond A, Moons KGM, Van Calster B, Van Smeden M. The Harms Of Class Imbalance Corrections For Machine Learning Based Prediction Models: A Simulation Study. Statistical Medicine. 2025;44(3–4):e10320. [PMID: 39865585] [PMCID: PMC11771573] [DOI]
15. Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical Guide To Shap Analysis: Explaining Supervised Machine Learning Model Predictions In Drug Development. Clinical and Translational Science. 2024;17(11):e70056. [PMID: 39463176] [PMCID: PMC11513550] [DOI]
16. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B. Tripod+Ai Statement: Updated Guidance For Reporting Clinical Prediction Models That Use Regression OR Machine Learning Methods. BMJ. 2024;385:e078378. [PMID: 38626948] [DOI]
17. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P. Probast+Ai: An Updated Quality, Risk Of Bias, And Applicability Assessment Tool For Prediction Models Using Regression OR Artificial Intelligence Methods. BMJ. 2025;388:e082505. [PMID: 40127903] [PMCID: PMC11931409] [DOI]
18. Scheideman AF, Shao MM, Zelada H, Cuadros J, Foreman J, Sarder P. Machine Learning To Diagnose Complications Of Diabetes. Journal of Diabetes Science and Technology. 2025;19(6):1650-70. [PMID: 40932163] [PMCID: PMC12425951] [DOI]
19. Mccoy RG, Patel S, Faust L, Heien HC, Zhang N, Caffo B, et al. Predicting Acute And Chronic Diabetes Complications Among Adults Newly Diagnosed With Diabetes: Development And Validation Of Dynamic Risk Prediction Models. Nature Communications. 2026. [PMID: 42778552] [PMCID: PMC13601617] [DOI]
20. Vamsi B, Bhattacharyya D. Micro And Macro Vascular Complications In Type_Ii Diabetes. Version 1 ed: Mendeley Data; 2021. [DOI]