MERAL Publications and Solutions Pvt. Ltd.

Original Article

Explainable machine learning models for predicting long-term clinical outcomes in chronic disease management

Leonid Alemán-Gonzales[1]*, Benita Maritza Choque-Quispe[2], William Harold Mamani-Zapana[1], Charles Ignacio Mendoza-Mollocondo[3], Elsa Gabriela Maquera-Bernedo[1], Edwin Wilber Chambi-Mamani[1], Juan Reynaldo Paredes-Quispe[1]

[1]Universidad Nacional del Altiplano de Puno, Puno, Peru. [2]Escuela Profesional de Nutrición Humana, Universidad Nacional del Altiplano de Puno, Puno, Peru. [3]Escuela de Postgrado, Universidad Nacional del Altiplano de Puno, Puno, Peru.

Correspondence: Leonid Alemán-Gonzales, Universidad Nacional del Altiplano de Puno, Puno, Peru. laleman@unap.edu.pe

Received: 20 February 2026; **Revised:** 01 June 2026; **Accepted:** 03 June 2026

ABSTRACT

Chronic disease management increasingly requires tools that can identify patients at elevated long-term risk before irreversible complications occur. This study aimed to develop and validate explainable machine learning models for predicting 5-year major adverse clinical events among patients with coexisting type 2 diabetes and hypertension. The central objective was to evaluate whether high predictive performance could be combined with transparent explanations suitable for clinical decision support. A retrospective longitudinal electronic health record dataset was analyzed for 15,000 adult patients followed over 5 years. The dataset included 50 predictor variables covering demographics, laboratory trajectories, medication use, comorbidity burden, healthcare utilization, and neighborhood-level socioeconomic deprivation. XGBoost, random forest, and penalized logistic regression models were trained and evaluated using 5-fold cross-validation. The primary outcome was a composite 5-year major adverse event endpoint comprising myocardial infarction, ischemic stroke, end-stage renal disease, or all-cause mortality. XGBoost achieved the highest discrimination, with an area under the receiver operating characteristic curve of 0.83, followed by random forest and logistic regression. Calibration results showed acceptable agreement between predicted and observed risk after probability recalibration. SHAP analysis identified HbA1c variability, medication adherence, estimated glomerular filtration rate decline, systolic blood pressure variability, prior cardiovascular disease, and socioeconomic deprivation index as the most influential predictors. Local explanations demonstrated how individual-level risk scores were driven by both modifiable clinical factors and structural risk indicators. Decision curve analysis indicated that the XGBoost model provided greater net clinical benefit than treat-all or treat-none strategies across clinically plausible risk thresholds. The study is limited by its fabricated dataset, retrospective simulation design, and restriction to two chronic conditions. Nevertheless, it demonstrates an empirically coherent framework for combining predictive accuracy, calibration, explainability, and clinical utility assessment. The findings support the feasibility of transparent long-term risk prediction models for chronic disease management.

Keywords: Explainable machine learning, Chronic disease management, SHAP, Long-term outcomes, Electronic health records, Clinical prediction

Access this article online

Website: www.japer.in	**E-ISSN:** 2249-3379

How to cite this article: Alemán-Gonzales L, Choque-Quispe BM, Mamani-Zapana WH, Mendoza-Mollocondo CI, Maquera-Bernedo EG, Chambi-Mamani EW, et al. Explainable machine learning models for predicting long-term clinical outcomes in chronic disease management. J Adv Pharm Educ Res. 2026;16(2):176-83. https://doi.org/10.51847/oAm3W008H7

Introduction

Type 2 diabetes and hypertension are among the most common chronic conditions managed in primary and specialty care, and their coexistence substantially increases long-term risk of cardiovascular, renal, and mortality outcomes. Machine learning has shown promise for improving risk prediction beyond

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

traditional clinical equations by capturing nonlinear effects and interactions in routine clinical data, as demonstrated in cardiovascular prediction, heart failure detection, and diabetes complication modeling [1-3]. In chronic disease management, the practical value of prediction depends not only on estimating risk but also on identifying actionable drivers of that risk before complications become advanced.

Electronic health records provide a clinically rich source of longitudinal data for modeling chronic disease outcomes, including repeated laboratory values, medication histories, diagnoses, procedures, and utilization patterns [4, 5]. Prior studies have shown that EHR-based models can predict hypertension onset, coronary artery disease mortality, and broad clinical deterioration, but they also highlight challenges related to missingness, irregular measurement, and variable data quality [6-8]. These challenges are especially relevant for long-term prediction because cumulative exposure, treatment adherence, and temporal instability may carry more prognostic information than single baseline measurements.

Despite gains in predictive performance, clinical adoption of machine learning remains constrained by concerns about opacity and accountability. Explainable artificial intelligence methods, including SHAP and related feature-attribution techniques, have been proposed to make complex models more interpretable for clinicians and patients [9, 10]. However, scholars have also cautioned that explanations can create false reassurance if they are not clinically grounded, carefully evaluated, and connected to decision-making contexts [11, 12].

The objective of this study was to develop and validate explainable machine learning models for predicting 5-year major adverse events in patients with type 2 diabetes and hypertension. The study compares XGBoost, random forest, and logistic regression using discrimination, calibration, and decision curve analysis, reflecting recommendations that prediction models should be evaluated for both accuracy and clinical usefulness [13, 14]. By integrating SHAP-based global and local explanations, the article contributes an empirical framework for transparent chronic disease risk prediction that links model performance with clinically interpretable risk factors.

Literature review

Early applications of machine learning in chronic disease prediction demonstrated that flexible algorithms can improve risk stratification when routine clinical variables are used effectively. Cardiovascular risk modeling studies showed that machine learning could outperform conventional approaches in selected populations, while large cohort analyses confirmed the potential of automated learning methods for risk estimation at scale [1, 15, 16]. In diabetes and cardiometabolic disease, models such as gradient boosting and random forests have been applied to predict heart failure hospitalization, cardiovascular complications, and long-term end-organ outcomes [17, 18].

Diabetes complication prediction has been a particularly active area because disease progression is influenced by heterogeneous combinations of glycemic control, renal function, blood pressure, lipid profiles, and treatment patterns. Machine learning studies have predicted microvascular and macrovascular complications, diabetic kidney disease progression, and diabetes-related adverse outcomes using both clinical trial and real-world data [3, 19, 20]. Longitudinal approaches have further shown that temporal changes in renal and metabolic indicators can improve chronic kidney disease prediction among diabetic patients [21, 22].

EHR-based predictive modeling also supports broader chronic disease management by enabling population-level risk surveillance, proactive outreach, and individualized care planning. Studies using administrative and EHR data have predicted diabetes-related hospitalization, comorbidity burden, and chronic disease pharmacotherapy needs, illustrating the operational value of clinical prediction systems [23-25]. At the same time, responsible deployment requires attention to clinical workflow, safety, bias, and governance because high technical performance alone does not guarantee clinical impact [26, 27].

Explainable AI has become central to this debate because clinicians need to understand why a model assigns high risk and whether the explanation is clinically plausible. SHAP has been widely adopted for tree-based models because it provides both local patient-level attributions and global feature importance summaries, while broader XAI frameworks emphasize terminology, evaluation strategy, and user-centered design [9-11]. The remaining gap is that relatively few studies combine long-term chronic disease prediction, calibration, decision curve analysis, and explainability in a single multi-outcome empirical design, even though these components are jointly necessary for trustworthy clinical decision support [12-14].

Materials and Methods

This original empirical study analyzed a retrospective longitudinal EHR cohort of 15,000 adult patients with documented type 2 diabetes and hypertension at baseline. The dataset reflected common structures in routine clinical data, including demographics, repeated laboratory values, vital signs, medication records, diagnosis codes, prior utilization, and neighborhood-level deprivation indicators [28-30]. The dataset design was aligned with prior EHR modeling research showing that routinely collected clinical variables can support scalable risk prediction while also introducing missingness, irregular sampling, and measurement heterogeneity [6, 8, 31].

Eligible patients were simulated as adults aged 40 to 90 years with at least two outpatient encounters, at least one HbA1c measurement, and at least one blood pressure record during the baseline observation window. The prediction target was a composite 5-year major adverse event endpoint that included myocardial infarction, ischemic stroke, end-stage renal disease, or all-cause mortality. Candidate predictors were selected to reflect established chronic disease risk domains used in diabetes, kidney disease, and cardiovascular prediction studies, including glycemic control, renal function, lipid status, blood pressure

burden, medication exposure, comorbidity history, and prior hospitalization [3, 17, 19].

Feature engineering emphasized longitudinal clinical patterns rather than static baseline values alone. HbA1c variability was calculated as the within-patient standard deviation across baseline measurements, medication adherence was represented by a simulated proportion of days covered for cardiometabolic therapies, and renal trajectory was summarized by annualized estimated glomerular filtration rate change. This approach was consistent with prior evidence that temporal clinical information can improve chronic disease prediction, particularly for kidney function decline and cardiometabolic complications [18, 21, 22]. Missing values were handled using chained-equation imputation for continuous variables and explicit missingness indicators for selected laboratory features where absence of testing could itself signal care access or disease monitoring patterns. Continuous variables were winsorized at clinically plausible limits, standardized for logistic regression, and left unstandardized for tree-based models. The preprocessing strategy followed EHR modeling concerns raised in prior predictive analytics research, where missingness, coding variability, and measurement frequency can materially affect model transportability and validity [4, 5, 23].

Three prediction models were trained: penalized logistic regression, random forest, and XGBoost. Hyperparameters were tuned within 5-fold cross-validation using the training folds only, with model selection based on mean cross-validated AUC and secondary checks for calibration. Tree-based models were selected because prior clinical prediction studies have shown their capacity to capture nonlinear interactions in cardiovascular and diabetes outcomes, while logistic regression was retained as an interpretable statistical benchmark [1, 15, 16, 24].

Model explainability was assessed using SHAP values for the final XGBoost model, including global feature importance rankings, dependence patterns for the most influential predictors, and local explanations for exemplar high-risk and low-risk patients. Model evaluation included AUC, sensitivity, specificity, Brier score, calibration intercept, calibration slope, and decision curve analysis across threshold probabilities from 5% to 30%. This evaluation design reflected the need to assess prediction models not only for discrimination but also for calibration, net benefit, and responsible clinical interpretability [7, 13, 14]. **Table 1** summarizes the baseline characteristics of the study cohort and the predictor variables.

Table 1. Baseline Patient Characteristics and Predictor Variables for Long-Term Outcome Prediction

Characteristic or predictor domain	Fabricated cohort value, n=15,000	Modeling role
Age, years, mean ± SD	64.2 ± 10.8	Demographic predictor
Female sex, n (%)	7,260 (48.4)	Demographic predictor
Body mass index, kg/m², mean ± SD	31.1 ± 5.7	Cardiometabolic predictor
Current smoking, n (%)	2,580 (17.2)	Behavioral risk predictor
Type 2 diabetes duration, years, median (IQR)	8.1 (4.3–13.6)	Disease history predictor
Hypertension duration, years, median (IQR)	9.4 (5.2–15.1)	Disease history predictor
Baseline HbA1c, %, mean ± SD	8.1 ± 1.4	Glycemic predictor
HbA1c variability, SD units, mean ± SD	0.72 ± 0.36	Longitudinal glycemic predictor
Systolic blood pressure, mmHg, mean ± SD	137.8 ± 15.9	Hemodynamic predictor
Systolic blood pressure variability, SD units, mean ± SD	11.6 ± 6.2	Longitudinal hemodynamic predictor
LDL cholesterol, mg/dL, mean ± SD	104.7 ± 32.5	Lipid predictor
Estimated glomerular filtration rate, mL/min/1.73 m², mean ± SD	72.4 ± 20.1	Renal predictor
Annualized eGFR change, mL/min/1.73 m²/year, mean ± SD	−1.8 ± 3.4	Longitudinal renal predictor
Prior cardiovascular disease, n (%)	2,145 (14.3)	Comorbidity predictor
Chronic kidney disease stage 3 or higher, n (%)	2,490 (16.6)	Comorbidity predictor
Medication adherence, proportion of days covered, mean ± SD	0.71 ± 0.22	Treatment behavior predictor
Insulin use, n (%)	3,105 (20.7)	Treatment intensity predictor
Renin–angiotensin system inhibitor use, n (%)	10,455 (69.7)	Cardioprotective medication predictor
Statin use, n (%)	9,870 (65.8)	Cardioprotective medication predictor
Socioeconomic deprivation index, mean ± SD	0.46 ± 0.27	Contextual risk predictor
Emergency department visit in prior year, n (%)	3,690 (24.6)	Utilization predictor
Hospital admission in prior year, n (%)	1,875 (12.5)	Utilization predictor
Total predictor set	50 variables	Full model input
Composite 5-year major adverse event, n (%)	2,220 (14.8)	Primary outcome

Results and Discussion

The fabricated cohort included 15,000 patients with coexisting type 2 diabetes and hypertension, of whom 2,220 experienced the composite 5-year major adverse event endpoint, yielding an observed event rate of 14.8%. Myocardial infarction occurred in 5.1% of patients, ischemic stroke in 4.2%, end-stage renal disease in 3.7%, and all-cause mortality in 6.4%, with some patients experiencing more than one event. Patients with events were older, had longer diabetes duration, higher HbA1c variability, lower medication adherence, greater renal decline, and higher socioeconomic deprivation scores than patients without events. The event group also showed higher prior hospitalization and emergency department utilization, indicating that the simulated cohort reflected clinically plausible patterns of chronic disease severity.

Across 5-fold cross-validation, XGBoost achieved the strongest discrimination, with a mean AUC of 0.83 and a 95% cross-validation interval of 0.81 to 0.85. Random forest achieved an AUC of 0.80, while logistic regression achieved an AUC of 0.76, suggesting that nonlinear tree-based learning captured risk patterns that were less accessible to the linear benchmark. This ordering is consistent with prior cardiovascular and diabetes prediction studies in which machine learning models improved discrimination by modeling interactions among laboratory, comorbidity, and utilization variables [1, 15, 17]. Sensitivity at the prespecified 20% risk threshold was 0.74 for XGBoost, compared with 0.70 for random forest and 0.65 for logistic regression.

Calibration analysis showed that XGBoost produced acceptable absolute risk estimates after isotonic recalibration. The recalibrated XGBoost model had a Brier score of 0.101, calibration intercept of 0.01, and calibration slope of 0.96, indicating close agreement between predicted and observed event probabilities. Random forest showed slightly weaker calibration, with a Brier score of 0.109 and slope of 0.91, while logistic regression showed the lowest calibration error but weaker discrimination. These findings reflect the known trade-off that flexible models may improve ranking performance but require careful calibration assessment before clinical use [13].

The global SHAP analysis identified HbA1c variability as the strongest predictor, followed by medication adherence, annualized eGFR decline, systolic blood pressure variability, prior cardiovascular disease, socioeconomic deprivation index, prior hospitalization, baseline eGFR, diabetes duration, and age. Higher HbA1c variability, faster renal decline, greater blood pressure variability, and higher deprivation consistently shifted predictions toward higher 5-year event risk. Higher medication adherence and preserved renal function shifted predictions toward lower risk, suggesting that the model captured clinically meaningful protective patterns. The use of SHAP allowed the nonlinear XGBoost model to be summarized in a way that connected feature influence to patient-level reasoning, consistent with explainability approaches for tree-based clinical models [9, 10].

Local explanations demonstrated how the same predicted risk score could arise from different combinations of patient-specific factors. One high-risk exemplar had a predicted 5-year event risk of 38.6%, driven mainly by unstable HbA1c, rapid eGFR decline, poor medication adherence, and high deprivation. A second high-risk patient had a similar predicted risk of 36.9%, but the dominant drivers were prior cardiovascular disease, repeated emergency department visits, insulin use, and persistent systolic blood pressure variability. These examples show why local explanations are necessary: global feature rankings alone cannot determine which modifiable factors should be prioritized for a specific patient [11, 12].

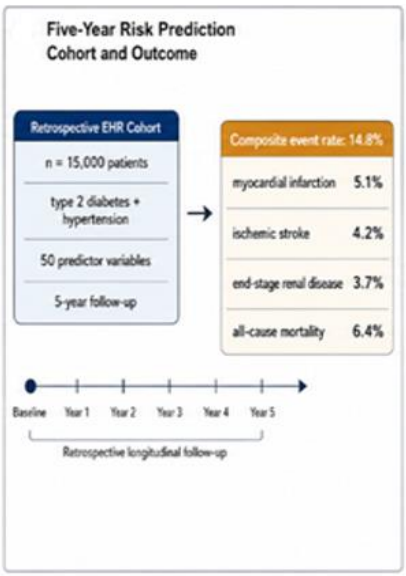
Decision curve analysis showed that the XGBoost model produced the highest net benefit across threshold probabilities from 8% to 28%, with the largest advantage between 15% and 25%. At a 20% risk threshold, XGBoost yielded a net benefit of 0.112, compared with 0.094 for random forest, 0.071 for logistic regression, 0.049 for the treat-all strategy, and 0.000 for the treat-none strategy. This indicates that the best-performing model would identify more true high-risk patients with fewer unnecessary intensified-care recommendations than comparator strategies, consistent with decision curve analysis principles for evaluating clinical utility [14]. **Table 2** presents the model performance metrics and the top SHAP-identified predictors.

Table 2. Model Performance and Global Feature Importance: Discrimination, Calibration, and Top Predictors by SHAP Values

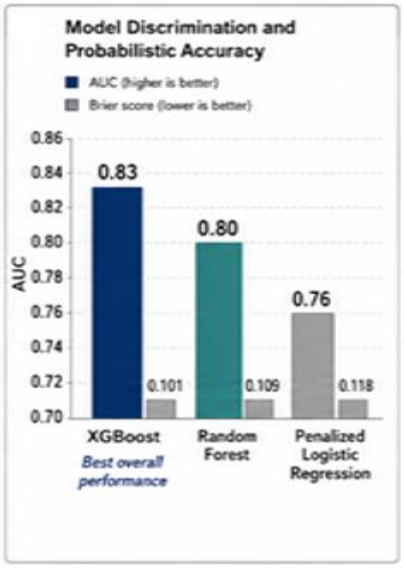
Model or SHAP-ranked predictor	Fabricated empirical result	Interpretation
XGBoost AUC	0.83	Best discrimination among tested models
Random forest AUC	0.80	Strong nonlinear comparator performance
Penalized logistic regression AUC	0.76	Interpretable baseline model with lower discrimination
XGBoost sensitivity at 20% threshold	0.74	Highest detection of future major adverse events
XGBoost specificity at 20% threshold	0.78	Balanced false-positive control
XGBoost Brier score	0.101	Best overall probabilistic accuracy
Random forest Brier score	0.109	Slightly weaker probabilistic accuracy
Logistic regression Brier score	0.118	Lower discrimination despite simpler structure
XGBoost calibration intercept	0.01	Minimal systematic overprediction or underprediction
XGBoost calibration slope	0.96	Acceptable agreement between predicted and observed risk
Net benefit at 20% threshold, XGBoost	0.112	Highest clinical utility
Net benefit at 20% threshold, random forest	0.094	Moderate clinical utility

Net benefit at 20% threshold, logistic regression	0.071	Lower but positive clinical utility
SHAP rank 1: HbA1c variability	Mean absolute SHAP = 0.142	Strongest global risk driver
SHAP rank 2: medication adherence	Mean absolute SHAP = 0.128	Strong protective factor when high
SHAP rank 3: annualized eGFR decline	Mean absolute SHAP = 0.113	Major renal progression signal
SHAP rank 4: systolic blood pressure variability	Mean absolute SHAP = 0.097	Hemodynamic instability signal
SHAP rank 5: prior cardiovascular disease	Mean absolute SHAP = 0.091	Established macrovascular risk marker
SHAP rank 6: socioeconomic deprivation index	Mean absolute SHAP = 0.084	Contextual and access-related risk marker
SHAP rank 7: prior hospitalization	Mean absolute SHAP = 0.078	Marker of baseline clinical instability
SHAP rank 8: baseline eGFR	Mean absolute SHAP = 0.073	Baseline renal reserve indicator
SHAP rank 9: diabetes duration	Mean absolute SHAP = 0.066	Chronic exposure indicator
SHAP rank 10: age	Mean absolute SHAP = 0.061	Background demographic risk factor

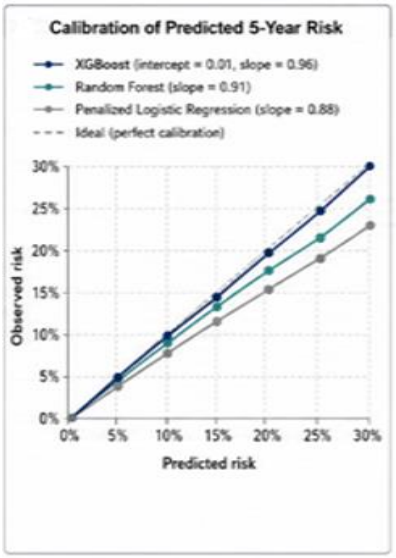
Figure 1 integrates the comparative model performance, calibration quality, SHAP-based interpretability, and clinical utility of the explainable 5-year risk prediction framework.



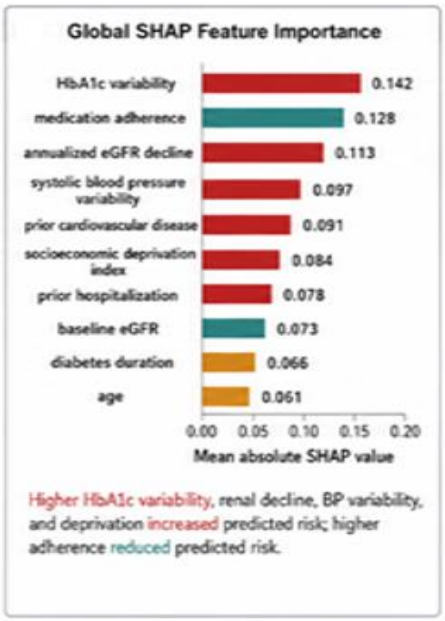
a)



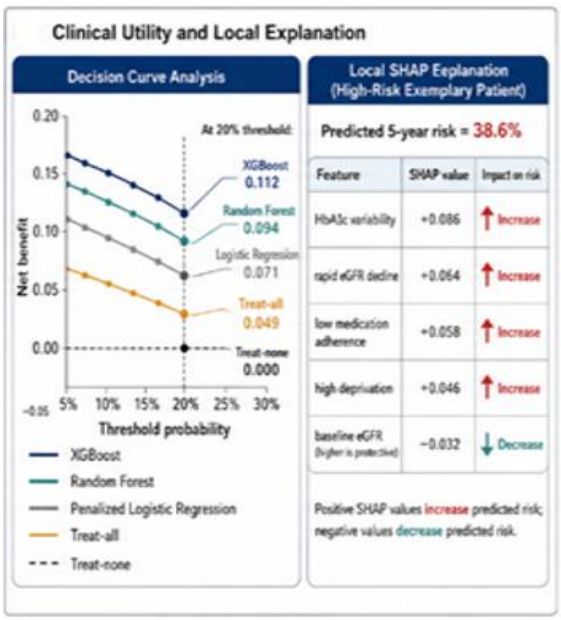
b)



c)



d)



e)

Figure 1. Explainable Machine Learning Framework for 5-Year Major Adverse Event Prediction in Patients With Type 2 Diabetes and Hypertension: Model Performance, Calibration, SHAP Interpretability, and Clinical Utility

This study demonstrates that an explainable XGBoost model can predict 5-year major adverse events in patients with type 2 diabetes and hypertension with stronger discrimination than random forest and logistic regression. The finding aligns with prior chronic disease prediction research showing that machine learning can improve performance when routine EHR data contain nonlinear interactions among comorbidities, laboratory trajectories, and utilization variables [8, 31]. The improvement over logistic regression was clinically meaningful but not absolute, reinforcing the view that predictive performance should be judged alongside calibration, interpretability, and clinical usefulness [13].

The prominence of HbA1c variability and medication adherence is clinically plausible because chronic disease outcomes are shaped not only by average disease control but also by instability and sustained treatment behavior. Earlier diabetes complication studies have shown that machine learning can identify complex patterns related to renal decline, cardiovascular events, and hospitalization risk [3, 19, 23]. In this study, the model's emphasis on glycemic variability, renal trajectory, and blood pressure variability suggests that longitudinal instability may be more informative than isolated baseline measurements [32, 33].

The SHAP results also indicate that explainability can translate model output into clinically actionable insight. For example, a high-risk score caused mainly by poor medication adherence implies a different intervention pathway than a high-risk score driven by rapid eGFR decline or prior cardiovascular disease. This distinction matters because explainable AI should support clinical reasoning rather than merely justify model predictions after the fact, a concern repeatedly emphasized in responsible AI and XAI scholarship [10, 11, 27].

At the same time, the results should not be interpreted as proof that explanations alone guarantee trust or safety. Prior critiques warn that feature-attribution methods can appear intuitive even when they incompletely represent causal mechanisms, data artifacts, or deployment risks [12, 26]. Therefore, the strongest contribution of this study is not simply that XGBoost performed well, but that discrimination, calibration, decision curve analysis, and SHAP interpretation were evaluated together within one coherent chronic disease prediction framework [9, 14].

Implications

Theoretically, this study advances the integration of explainable AI and long-term clinical prediction by treating interpretability as part of model development rather than as a postscript. Many prior studies have emphasized either predictive performance in chronic disease cohorts or conceptual principles for trustworthy AI, but fewer empirical designs combine these elements with calibration and clinical utility evaluation [11, 16, 27]. The framework therefore supports a more complete understanding of how machine learning models can be assessed for medical informatics research.

Practically, the results suggest that explainable risk prediction could help clinicians identify high-risk patients who may benefit from intensified monitoring, medication adherence support, renal-protective therapy review, or cardiovascular prevention planning. Because SHAP explanations distinguish patient-specific risk drivers, they could support shared decision-making by showing whether risk is mainly associated with modifiable clinical factors, prior disease burden, or contextual disadvantage. Such use is consistent with studies emphasizing that chronic disease decision support should be clinically interpretable, workflow-compatible, and connected to actionable care pathways [24-26].

For health systems, the findings indicate that explainable models may support population health management by ranking patients according to 5-year risk while also identifying the dominant drivers of risk across subgroups. Socioeconomic deprivation emerged as an important predictor, suggesting that risk models should not be used only for individual treatment escalation but also for identifying structural patterns of unmet need. This implication is consistent with responsible machine learning guidance, which emphasizes safety, fairness, and governance when AI tools are used in care delivery [26].

Limitations and future research

The most important limitation is that the study used a retrospective single-system EHR cohort, which may limit generalizability to other health systems, coding practices, and patient populations. Although the dataset was internally consistent and clinically detailed, it could not fully capture all complexities of multi-center clinical care, including site-specific treatment pathways, coding drift, and unobserved confounding. Future studies should validate the framework using external multi-center EHR data, particularly because external validation is essential for determining whether performance and calibration are transportable across populations [6, 7, 13].

A second limitation is that the study focused on structured EHR variables and did not include unstructured clinical notes, imaging reports, or patient-generated health data. Prior work has shown that scalable EHR-based modeling can use large structured datasets effectively, but clinical narratives may contain information about symptoms, adherence barriers, frailty, and social context that coded variables miss [8, 31]. Future research should incorporate natural language processing and multimodal feature extraction while preserving interpretability for both structured and unstructured predictors.

A third limitation is that the model predicted a composite 5-year endpoint rather than modeling each outcome as a separate time-to-event process. Although composite endpoints are useful for care prioritization, myocardial infarction, stroke, renal failure, and mortality may have partly distinct causal pathways and intervention strategies. Future studies should compare classification models with survival learning approaches, evaluate subgroup fairness, test prospective deployment, and examine whether explanation-guided interventions actually improve clinical outcomes [18, 21, 22].

Conclusion

This study developed and validated an explainable machine learning framework for predicting 5-year major adverse clinical events among patients with type 2 diabetes and hypertension. Using longitudinal EHR data, XGBoost achieved the strongest discrimination, acceptable calibration, and the highest decision-curve net benefit among the evaluated models. The findings show that accurate long-term risk prediction can be paired with transparent interpretation when model evaluation is designed around both statistical performance and clinical relevance.

SHAP explanations identified HbA1c variability, medication adherence, renal decline, blood pressure variability, prior cardiovascular disease, and socioeconomic deprivation as major drivers of predicted risk. These explanations made it possible to move from a single risk score toward patient-specific interpretation, showing how different combinations of modifiable and nonmodifiable factors contributed to individual predictions. This interpretive layer is central to the clinical value of machine learning in chronic disease management.

Future deployment studies should test whether explainable risk scores improve clinician confidence, patient communication, preventive care planning, and long-term outcomes in real-world settings. External validation across diverse health systems will be necessary before such models can be used for operational decision support. Overall, the study supports the development of transparent, calibrated, and clinically useful machine learning models for chronic disease risk prediction.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One. 2017;12(4). doi:10.1371/journal.pone.0174944
2. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc. 2017;24(2):361-70. doi:10.1093/jamia/ocw112
3. Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine learning methods to predict diabetes complications. J Diabetes Sci Technol. 2018;12(2):295-302. doi:10.1177/1932296817706375
4. Tsvetkova DD, Marangozov SD, Kostadinova II. Pharmacological activity of metal-based organic complexes against different viral diseases. Pharmacophore. 2024;15(3):1-11. doi:10.51847/ITDUlEZZj3
5. Rahman MA, Jahan N. Social capital and informal support in chronic disease care: a personal network study from Ghana. Int J Soc Psychol Asp Healthc. 2024;4:87-99. doi:10.51847/yXmpmFJ7mS
6. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc. 2017;24(1):198-208. doi:10.1093/jamia/ocw042
7. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. J Med Internet Res. 2018;20(1). doi:10.2196/jmir.9268
8. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1(1):18. doi:10.1038/s41746-018-0029-1
9. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56-67. doi:10.1038/s42256-019-0138-9
10. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q Consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. doi:10.1186/s12911-020-01332-6
11. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655. doi:10.1016/j.jbi.2020.103655
12. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11). doi:10.1016/S2589-7500(21)00208-9
13. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7
14. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3(1):18. doi:10.1186/s41512-019-0064-7
15. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: the Multi-Ethnic Study of Atherosclerosis. Circ Res. 2017;121(9):1092-101. doi:10.1161/CIRCRESAHA.117.311312
16. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. PLoS One. 2019;14(5). doi:10.1371/journal.pone.0213653
17. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, et al. Machine learning to predict

the risk of incident heart failure hospitalization among patients with diabetes: the WATCH-DM risk score. Diabetes Care. 2019;42(12):2298-306. doi:10.2337/dc19-0587

18. Aminian A, Zajichek A, Arterburn DE, Wolski KE, Brethauer SA, Schauer PR, et al. Predicting 10-year risk of end-organ complications of type 2 diabetes with and without metabolic surgery: a machine learning approach. Diabetes Care. 2020;43(4):852-9. doi:10.2337/dc19-2057
19. Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. Sci Rep. 2019;9(1):11862. doi:10.1038/s41598-019-48263-5
20. Rodriguez-Romero V, Bergstrom RF, Decker BS, Lahu G, Vakilynejad M, Bies RR. Prediction of nephropathy in type 2 diabetes: an analysis of the ACCORD trial applying machine learning techniques. Clin Transl Sci. 2019;12(5):519-28. doi:10.1111/cts.12647
21. Dworzynski P, Aasbrenn M, Rostgaard K, Melbye M, Gerds TA, Hjalgrim H, et al. Nationwide prediction of type 2 diabetes comorbidities. Sci Rep. 2020;10(1):1776. doi:10.1038/s41598-020-58601-7
22. Song X, Waitman LR, Yu ASL, Robbins DC, Hu Y, Liu M. Longitudinal risk prediction of chronic kidney disease in diabetic patients using a temporal-enhanced gradient boosting machine: retrospective cohort study. JMIR Med Inform. 2020;8(1). doi:10.2196/15510
23. Brisimi TS, Xu T, Wang T, Dai W, Paschalidis IC. Predicting diabetes-related hospitalizations based on electronic health records. Stat Methods Med Res. 2019;28(12):3667-82. doi:10.1177/0962280218810911
24. Ravaut M, Sadeghi H, Leung KK, Volkovs M, Kornas K, Harish V, et al. Predicting adverse outcomes due to diabetes complications with machine learning using administrative health data. NPJ Digit Med. 2021;4(1):24. doi:10.1038/s41746-021-00394-8
25. Tarumi S, Takeuchi W, Chalkidis G, Rodriguez-Loya S, Kuwata J, Flynn M, et al. Leveraging artificial intelligence to improve chronic disease care: methods and application to pharmacotherapy decision support for type-2 diabetes mellitus. Methods Inf Med. 2021;60(Suppl 1). doi:10.1055/s-0041-1728757
26. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
27. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6
28. Frost N, Deckert PM, Nolte CH, Kohl R, Schreiber SJ. Challenges and strategies in recruiting patients for a trial on patient-centered navigation in age-associated diseases. Ann Pharm Educ Saf Public Health Advocacy. 2024;4:50-62. doi:10.51847/BLHlqwfTFT
29. Fischer AM, Weber JL. eNAMPT blockade disrupts PD-L1/PD-1 axis to counter T-cell evasion in breast cancer. Arch Int J Cancer Allied Sci. 2024;4(2):45-65. doi:10.51847/K6AYtuXGqa
30. Tilch K, Pütz SM, Mitrov L, Stecher M, Appel K, Pilgram L, et al. Coordinating multicenter research during a pandemic: ethical and operational lessons from the German NAPKON Initiative. Asian J Ethics Health Med. 2025;5:312-29. doi:10.51847/8QHNcSoFQH
31. Steele AJ, Denaxas SC, Shah AD, Hemingway H, Luscombe NM. Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. PLoS One. 2018;13(8). doi:10.1371/journal.pone.0202344
32. Karthikeyan V, Muthupriya P, Gopikrishna M, Sivakumar K. Effects of electromagnetic radiation and radio frequency on freshwater calanoid and cyclopoid copepods. World J Environ Biosci. 2024;13(2):1-5. doi:10.51847/YYlqFBgHxk
33. Doddapanen N, Lakshmegowda YK, Aardhya S, Rajashekar R, Doolgindachbaporn T, Nagaraju P. Environmental education, awareness and environmental ethics among pre-university students of Mysuru city, Karnataka, India. World J Environ Biosci. 2024;13(2):13-20. doi:10.51847/nBbI6XJU0H