Supplementary Table 1 TRIPOD Checklist for reporting prediction model development and validation

Section/topic	Item	Checklist item	Location in the manuscript (section, sub-section, sub-section paragraphs)
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	Title
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Abstract
Introduction			

Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	Introduction, paragraphs 1-7
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	Introduction, paragraph 8
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	Methods, Study Design and Data Source, paragraph 1

	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	Methods, Inclusion and Exclusion Criteria, paragraph 1
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centers.	Methods, Inclusion and Exclusion Criteria, paragraph 1
	5b	Describe eligibility criteria for participants.	Methods, Inclusion and Exclusion Criteria, paragraph 1
	5c	Give details of treatments received, if relevant.	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	Methods, Outcomes definitions, paragraph 1
	6b	Report any actions to blind assessment of the outcome to be predicted.	NA

Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	Table 1
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	NA
Sample size	8	Explain how the study size was arrived at.	Results, General Cohort, paragraph 1
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	Methods, Data Pre-processing and Feature Engineering, paragraph 1; Table 1
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	Methods, Data Pre-processing and Feature Engineering, paragraph 1

	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	Methods, Pre-processing and Feature Engineering, paragraph 1; Methods, Machine Learning Approach and Algorithm Definition, paragraph 1
	10d	Compare all measures used to assess model performance and, if relevant, to specify multiple models.	Methods, Performance Assessment, paragraphs 1-3
Risk groups	11	Provide details on how risk groups were created, if done.	NA
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Results, General Cohort, paragraph 1; Figure 1

	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Results, General Cohort, paragraph 1; Table 1
Model development	14a	Specify the number of participants and outcome events in each analysis.	Results, General Cohort
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Supplementary Materials, Public Repository

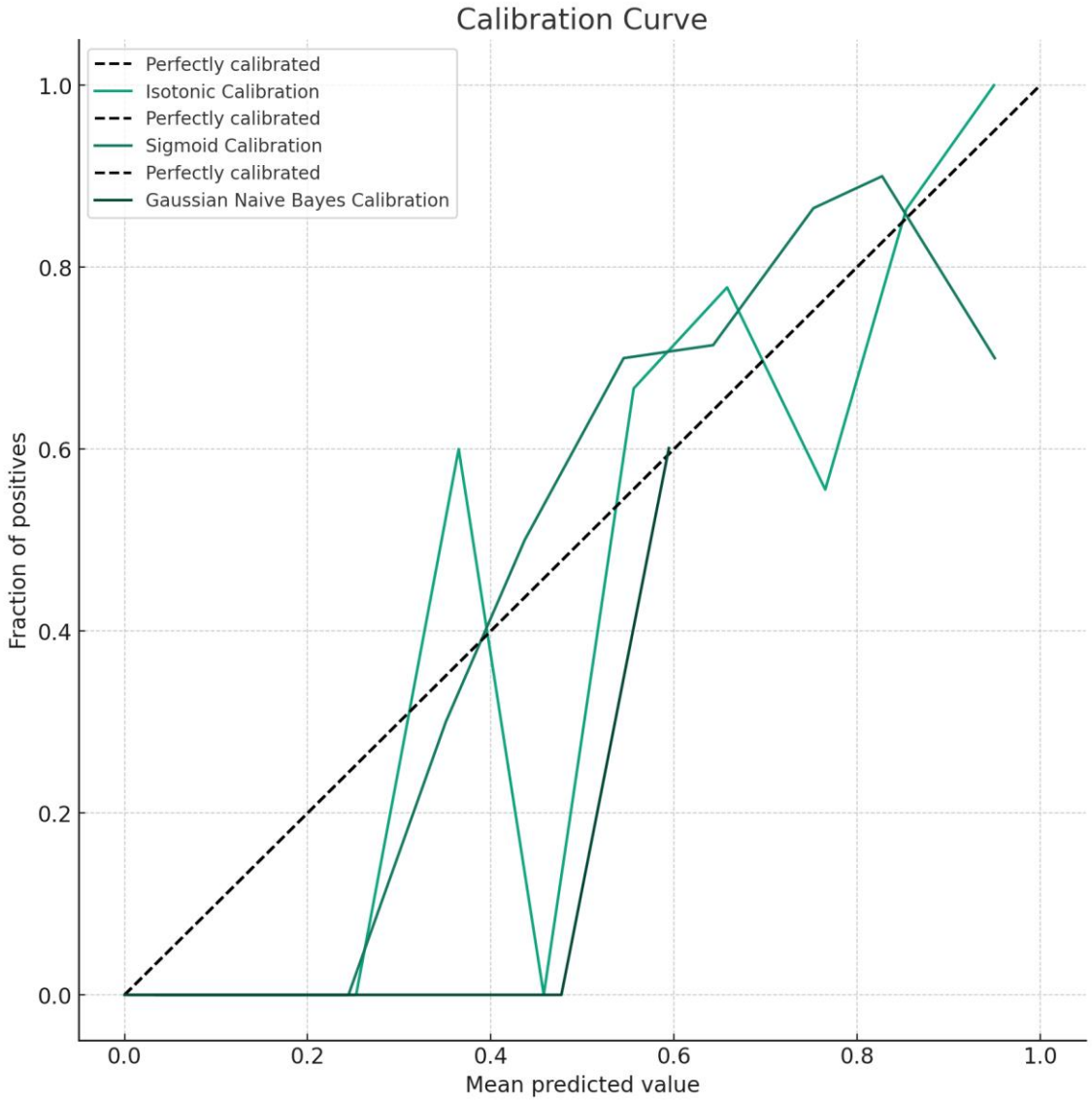
	15b	Explain how to use the prediction model.	Methods, *Code availability and web deployment*
Model performance	16	Report performance measures (with CIs) for the prediction model.	Results, Model Performance; Figure 2
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	Discussion, paragraph 14
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	Discussion, paragraphs 1-6
Implications	20	Discuss the potential clinical use of the model and implications for future research.	Discussion, paragraphs 7-13

Other information			Supplementary Materials
Supplementary	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Methods, Supplementary Material and Web Deployment
Funding	22	Give the source of funding and the role of the funders for the present study.	Footnote Page

This checklist, adapted from the TRIPOD Statement, outlines the reporting of fundamental aspects for the development and validation of a multivariable machine learning prediction model in this study. The TRIPOD statement is made available by its authors at https://www.tripod-statement.org/.

Supplementary Table 2 Additional information on the XGBoost hyperparameters optimized for the outcome prediction process as well as a general overview of their role in the algorithm functionality

Hyperparameter	Value	Description
Objective	binary:logistic	Specifies binary logistic regression as the objective for binary classification.
Booster	gbtree	Uses a gradient boosting decision tree as the base learner.
Lambda	1.94e-06	L2 regularization to prevent overfitting.
Alpha	0.462	L1 regularization to prevent overfitting.
Max_depth	7	Maximum depth of each decision tree.
Eta	0.785	Learning rate that shrinks the contribution of each tree.
Gamma	4.63e-07	Minimum loss reduction required for a split.
Grow_policy	lossguide	Choose split points to optimize loss reduction.



Supplementary Figure 1 Calibration curves visualizing the relationship between the mean predicted probabilities for various methods of calibration and the true outcomes. The ideal "perfectly calibrated line serves as a benchmark, with deviations from this line by the isotonic, sigmoid, and Gaussian naive Bayes calibration curves indicating their respective calibration performances.