The American Journal of Cardiology xxx (2025) xxx-xxx

Contents lists available at ScienceDirect

The American Journal of Cardiology

journal homepage: www.elsevier.com/locate/ajc

Comparative Performance of Machine Learning and Traditional Risk Scores in Predicting Adverse Events After Transcatheter Aortic Valve Replacement in Patients With Atrial Fibrillation

Q1 Johny Nicolas, MD, MSc[a], George Dangas, MD, PhD[a*], Amanda Borrow, PhD[b], Rüdiger Smolnik, MD[c], Felix Just, PhD[c], Cathy Chen, MD[b], Krishna Padmanabhan, PhD[d], Eva-Maria Fronk, PhD[c], Christian Hengstenberg, MD[e], Nicolas Van Mieghem, MD, PhD[f], Martin Unverdorben, MD, PhD[b]

[a] *Mount Sinai Fuster Heart Hospital, Icahn School of Medicine at Mount Sinai, New York, New York*
[b] *Daiichi Sankyo, Inc., Basking Ridge, New Jersey*
[c] *Daiichi Sankyo Europe, GmbH, Munich, Germany*
[d] *Cytel, Cambridge, Massachusetts*
[e] *Department of Internal Medicine II, Division of Cardiology, Vienna General Hospital, Medical University, Vienna, Austria*
[f] *Department of Cardiology, Erasmus University Medical Center, Rotterdam, The Netherlands*

ARTICLE INFO

Article History:
Received 27 June 2025
Revised 17 November 2025
Accepted 4 December 2025

Keywords:
artificial intelligence
atrial fibrillation
oral anticoagulation
TAVR
machine learning

Patients with atrial fibrillation (AF) following transcatheter aortic valve replacement (TAVR) remain at risk of ischemic stroke (IS) and bleeding. However, traditional risk scores provide modest predictions of IS and bleeding in these patients. We aimed to develop machine learning (ML) models that predict IS, major gastrointestinal bleeding (MGIB), all clinically relevant bleeding (CRB), and net adverse clinical events (NACE) using data from patients in the ENVISAGE-TAVI AF trial. Ten ML algorithms were trained per outcome using nested cross-validation; the best-performing model (highest F1 score) was validated on a 25% holdout set. Model performance was compared with logistic regression models using CHA_2DS_2-VA or HAS-BLED. Among 1,377 patients, 41 had an IS, 83 had MGIB, 375 had CRB, and 255 experienced NACE. The predictive abilities of a linear discriminant analysis algorithm for IS (F1 score = 0.08) and CHA_2DS_2-VA (F1 score = 0.09) were similarly low, but numerically better than HAS-BLED (F1 score = 0.05). Prediction of MGIB was similarly low for a logistic-lasso algorithm (F1 score = 0.11), CHA_2DS_2-VA (F1 score = 0.09), and HAS-BLED (F1 score = 0.12). For CRB, the predictive performance of a Naïve Bayes algorithm (F1 score = 0.39) was similar to CHA_2DS_2-VA (F1 score = 0.38) and HAS-BLED (F1 score = 0.41). The predictive ability of a logistic regression algorithm for NACE (F1 score = 0.33) was numerically better than CHA_2DS_2-VA (F1 score = 0.22) or HAS-BLED (F1 score = 0.27). In conclusion, ML offered similar predictive ability to established risk scores for thromboembolic and bleeding outcomes among transcatheter aortic valve replacement (TAVR) patients with AF.

Patients with atrial fibrillation (AF) undergoing transcatheter aortic valve replacement (TAVR) are at significant risk of both thromboembolic and bleeding complications.[1] Clinical risk scores are routinely used to risk-stratify patients with AF as recommended by both the European and American guidelines for the management of AF.[2,3] The CHA_2DS_2-VA score estimates stroke risk, whereas the HAS-BLED score assesses the bleeding risk in patients with AF.[4,5] Nonetheless, while simple and widely adopted, the CHA_2DS_2-VA and HAS-BLED scores were derived from general AF populations rather than post-TAVR patients, who tend to be older and have more comorbidities. Machine learning (ML) could provide an alternative way to improve risk prediction by incorporating complex linear and nonlinear interactions in high-dimensional datasets. In this study, we aimed to (1) develop ML models using data from the ENVISAGE-TAVI AF trial to predict adverse thromboembolic and bleeding events in patients with AF after successful TAVR and subsequently (2) compare the performance of these ML models to 2 commonly used risk scores (CHA_2DS_2-VA and HAS-BLED scores) in terms of discrimination, calibration, and reclassification.

Methods

Study design

We used patient-level data from the ENVISAGE-TAVI AF trial (NCT02943785),[6] a multinational, multicenter, prospective, randomized, open-label trial with blinded endpoint adjudication, designed to

This study was sponsored by Daiichi Sankyo.
*Corresponding author:
E-mail address: george.dangas@mountsinai.org (G. Dangas).

https://doi.org/10.1016/j.amjcard.2025.12.002

compare edoxaban with vitamin K antagonist (VKA) therapy in patients with AF following successful TAVR. The trial enrolled patients who were 18 years of age or older, with either pre-existent paroxysmal/persistent AF or new-onset AF after successful TAVR for severe AS. Patients with comorbidities associated with a high-bleeding risk (i.e., history of intracranial hemorrhage, esophageal varices, and unresolved periprocedural complications, etc.) were excluded. A full description of inclusion and exclusion criteria is given in Supplementary Table 1. A total of 1,426 patients were enrolled across international centers and randomized between 12 hours and 5 days after the TAVR procedure.[6,7] All patients were followed for a median of approximately 18 months. Patients were allocated to receive either once-daily oral edoxaban (60 mg or 30 mg based on dose-reduction criteria per local labeling) or a standard VKA regimen targeting an international normalized ratio of 2.0 to 3.0 (or 1.6-2.6 for patients aged ≥70 years in Japan). Concomitant antiplatelet therapy was permitted at the discretion of the treating physician. Randomization was stratified by anticipated edoxaban dose reduction. The maximum treatment duration was up to 36 months.

Clinical outcomes

We examined 4 clinical outcomes of interest: (1) clinically relevant bleeding (CRB), (2) ischemic stroke (IS), (3) major gastrointestinal bleeding (MGIB), and (4) net adverse clinical events (NACE). CRB was classified according to the International Society on Thrombosis and Hemostasis definition and included both major (clinically overt bleeding associated with a reduced hemoglobin level, blood transfusion, symptomatic bleeding at a critical site, or death) and clinically relevant nonmajor bleeding (any sign or symptom of hemorrhage that does not fit the criteria for major bleeding but does meet at least 1 of the following criteria: (1) requires medical intervention by a healthcare professional, (2) leads to hospitalization or increased level of care, or (3) prompts a face-to-face evaluation).[8,9] NACE was defined as the composite of death from any cause, myocardial infarction, IS, systemic thromboembolic event, valve thrombosis, or major bleeding. IS was defined in accordance with the Valve Academic Research Consortium-2 consensus as an acute episode of focal cerebral, spinal, or retinal dysfunction caused by infarction of the central nervous system tissue.[10]

Model development and statistical analysis

This analysis utilized data from the safety cohort, which included patients who received ≥1 dose of the study drug. The dataset included 180 variables, selected by experts for their potential clinical relevance. After basic data transformations and cleaning, including the removal of homogeneous features and highly correlated variables, 155 variables remained in the final dataset. Since this dataset was derived from a randomized controlled trial, baseline characteristics were well balanced across the edoxaban and VKA arms. Therefore, we did not separate model development by treatment group. Instead, treatment allocation was included as a candidate predictor. Ten different ML algorithms (logistic regression, lasso regression, ridge regression, elastic net, linear discriminant analysis, Naïve Bayes, decision tree, random forest, XGBoost, and multilayer perceptron) were evaluated using nested cross-validation on a training dataset comprising 75% of the patients from the full dataset. The nested cross-validation involved 2 loops to prevent overfitting and bias in model selection. In the inner loop, hyperparameter tuning was performed using Bayesian optimization. In the outer loop, the models were scored using the F1 score to determine the best model for each endpoint. The F1 score was selected as the appropriate performance metric for imbalanced classification tasks. Finally, trained models were evaluated on a holdout test set comprising 25% of the patients not included in the training set. We also compared the performance of the best ML model for each endpoint to a logistic regression model that used a traditional risk score (CHA_2DS_2-VA and HAS-BLED) as its only input variable. The CHA_2DS_2-VA and HAS-BLED scores are widely used clinical tools to guide anticoagulation management in patients with AF.[11] The CHA_2DS_2-VA score estimates the risk of stroke or systemic embolism by assigning points for congestive heart failure, hypertension, age ≥75 years (2 points), diabetes mellitus, prior stroke or transient ischemic attack (2 points), vascular disease, and age 65 to 74 years. In contrast, the HAS-BLED score evaluates the risk of major bleeding, with points assigned for hypertension, abnormal renal or liver function, prior stroke, history of bleeding, labile INR, age >65 years, and concomitant drug or alcohol use. Both scores share several overlapping factors (i.e., hypertension, stroke history, and age), but they differ in their primary purpose. Unlike CHA_2DS_2-VA and HAS-BLED, which are based on a small set of predefined clinical risk factors, our ML models incorporated over 150 routinely collected clinical, laboratory, and procedural variables. This broader input allows ML to explore nonlinear and higher-order interactions that are not captured by traditional scores and to identify predictors specific to the TAVR population.

To identify the most impactful variables for the outcomes of interest, SHAP (Shapley Additive exPlanations) was employed. SHAP values utilize a game-theoretical approach to quantify the importance of a variable in a prediction by assessing its contribution across various combinations of variables.[12] The features were ranked using the mean absolute SHAP value from the holdout test set. A detailed explanation of the methods used for the derivation and optimization of the ML models was previously described.[13]

Results

Patient characteristics

Baseline characteristics of the overall safety cohort are listed in Table 1. The study population included 1,377 patients, with a mean age of 82.1 ± 5.4 years, approximately half (52.2%) were men, and the majority (83.2%) identified as White. The mean body mass index was 27.7 ± 5.5 kg/m^2, and renal function was moderately impaired, with a mean creatinine clearance of 58.2 ± 24.1 mL/min. Cardiovascular and metabolic comorbidities were highly prevalent, including hypertension (91.4%), hypercholesterolemia (70.0%), diabetes (36.7%), and advanced heart failure (New York Heart Association class III or IV) in nearly half (44.5%) of the cohort. Baseline thromboembolic and bleeding risks were elevated, reflected by a median CHA_2DS_2-VA score of 4 (mean 4.5 ± 1.3) and a mean HAS-BLED score of 1.6 ± 0.8.

Clinically relevant bleeding

Out of 1,377 patients, CRB was recorded in 375 patients (27.2%). The Naïve Bayes model (Supplementary Table 2) provided the best predictive performance, with an F1 score of 0.39, closely matching HAS-BLED (F1 score = 0.41) and CHA_2DS_2-VA (F1 score = 0.38). Among the ML predictors, the absence of history of previous valve replacement before index procedure (no valve replacement) had the highest SHAP value, indicating its strong influence (Supplementary Figure 1).

Ischemic stroke

IS occurred in 41 (3.0%) patients. Among 10 ML models tested, linear discriminant analysis demonstrated the highest predictive performance (Supplementary Table 2), with an F1 score of 0.08, comparable to the traditional CHA_2DS_2-VA score (F1 score = 0.09; Figure 1). Both these models had a higher F1 score than the HAS-BLED model (F1 score = 0.05). The variables with the greatest impact on the ML

Table 1
Baseline characteristics

Parameter	Overall *N* = 1,377
Age at enrollment, years, mean ± SD	82.1 ± 5.4
Sex, male	719 (52.2%)
Race	
White	1,146 (83.2%)
Ethnicity	
Hispanic or Latino	352 (25.6%)
Smoking status at baseline	
Current	52 (3.8%)
Weight, kg, mean ± SD	75.3 ± 17.6
BMI, kg/m^2, mean ± SD	27.7 ± 5.5
CrCl, mL/min, mean ± SD	58.2 ± 24.1
Hemoglobin, g/L, mean ± SD	115.7 ± 77.8
AF	874 (63.5%)
AF paroxysmal	569 (41.3%)
AF persistent	
>7 days but <1 year	158 (11.5%)
≥1 year	110 (8.0%)
AF permanent	518 (37.6%)
AF flutter	18 (1.3%)
Hypertension	1,258 (91.4%)
Mitral valve disease[a]	599 (43.5%)
Diabetes mellitus	506 (36.7%)
Hypercholesterolemia	964 (70.0%)
COPD	199 (14.5%)
Prior MB or predisposition to bleeding	119 (8.6%)
Heart failure	
NYHA III or IV	613 (44.5%)
History of stroke or TIA	233 (16.9%)
Prior PCI	354 (25.7%)
Prior MI	191 (13.9%)
Prior CABG	124 (9.0%)
HAS-BLED score, mean ± SD	1.6 ± 0.8
CHA_2DS_2-VA score, mean ± SD	4.5 ± 1.3
STS risk score, mean ± SD	4.9 ± 3.8
AC therapy prior to randomization	1,075 (78.1%)
TAVR devices	
Balloon-expandable valve	631 (45.8%)
Self-expandable valve	449 (32.6%)
Other	297 (21.6%)
History of labile INR	108 (7.8%)
Treatment group	
Edoxaban arm	693 (50.3%)
VKA arm	684 (49.7%)
Edoxaban baseline dose	
30 mg	320 (23.2%)
60 mg	370 (26.9%)

Data are presented as n (%) unless otherwise noted. Treatment arm is the actual treatment arm received which, in the safety analysis set, differs to the randomized treatment arm in 1 patient who was randomized to VKA but received edoxaban.

[a] Mitral valve disease is classified based on the investigator response to the question, "Mitral valve disease is present," on the baseline STS questionnaire. If mitral valve disease is present, moderate and severe disease is indicated in response to evidence of mitral valve insufficiency/regurgitation on the baseline STS questionnaire.

[b]Patients who had an indication for dose reduction at randomization met at least 1 of the following parameters: CrCl ≤50 mL/min, body weight ≤60 kg, or PGP-inhibitor use.

AC = anticoagulant; AF = atrial fibrillation; BMI = body mass index; CHA_2DS_2-VA = Congestive heart failure, Hypertension, Age ≥75 years (doubled), Diabetes mellitus, Stroke (doubled), Vascular disease, Age 65-74 years; CABG = coronary artery bypass graft; CrCl = creatinine clearance; COPD = chronic obstructive pulmonary disease; HAS-BLED = Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly, Drugs/alcohol concomitantly; INR = international normalized ratio; MB = major bleeding; MI = myocardial infarction; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; SD = standard deviation; STS = Society of Thoracic Surgeons; TAVR = transcatheter aortic valve replacement; TIA = transient ischemic attack; VKA = vitamin K antagonist.

prediction for IS were patient history of stroke and of heart failure (Supplementary Figure 2).

Major gastrointestinal bleeding

MGIB occurred in 83 (6.0%) patients. The logistic-lasso algorithm (Supplementary Table 2) achieved the highest F1 score (0.11), closely followed by HAS-BLED (F1 score = 0.12) and CHA_2DS_2-VA (F1 score = 0.09). The treatment arm (edoxaban) emerged as the most influential variable affecting MGIB predictions, as evidenced by having the highest SHAP value (Supplementary Figure 3).

Net adverse clinical events

NACE occurred in 255 (18.5%) patients. Logistic regression (Supplementary Table 2) was the top-performing ML model, achieving an F1 score of 0.33, numerically outperforming CHA_2DS_2-VA (F1 score = 0.22) and HAS-BLED (F1 score = 0.27). Treatment arm (edoxaban) and sex were identified as the most influential variables in predicting NACE outcomes (Supplementary Figure 4).

Comparison of predictors: machine learning models versus traditional risk scores

Overlap and divergence of risk factors identified by ML models versus traditional scores (HAS-BLED and CHA_2DS_2-VA) for bleeding and IS are illustrated in Figures 2A and 2B. ML models identified clinical and TAVR procedure specific predictors, such as valve type or mean aortic valve gradient, with only a few shared variables with traditional scores. For NACE, specific ML-derived risk factors are presented in Figure 2C, overlaid with those for bleeding and IS. AF was a common predictor across all 3 outcomes, and having no history of heart failure was shared by IS and MGIB, whereas some clinical and procedural characteristics (e.g., hypertension, Evolut valve type, edoxaban treatment assignment) were specific for NACE.

Discussion

In this analysis of data from the ENVISAGE-TAVI AF trial, we developed and evaluated ML models to predict IS, bleeding outcomes, and NACE among TAVR patients with AF. Our derived ML models demonstrated similar or modestly improved predictive performance compared to traditional risk scores and identified potential clinical and procedural predictors not captured by these tools.

Both ML algorithms and conventional risk scores underperformed for IS prediction, likely due to low event rates. All patients in the trial were on anticoagulation, which reduced their risk of thromboembolic events. Indeed, the conventional scoring systems are primarily intended to stratify stroke and bleeding risk before initiating anticoagulation.[3] Once patients are on therapy, the observed stroke rates are typically below those of untreated cohorts,[14] leading to an overestimation of stroke risk and poor score calibration. Prior studies in anticoagulated AF populations reported only modest discrimination for stroke using ML prediction models and the CHA_2DS_2-VASc and HAS-BLED scores.[15] Similarly, in TAVR patients, the CHA_2DS_2-VASc score was shown to have a strong correlation with in-hospital stroke but with low accuracy.[16]

Consistent with prior studies, both ML algorithms and conventional risk scores have demonstrated only a modest ability to predict bleeding in anticoagulated TAVR patients. The HAS-BLED score, originally developed for bleeding risk assessment in AF, has shown moderate discrimination for severe bleeding (defined as life-threatening and major bleeding according to the Valve Academic Research Consortium-2 criteria) in the TAVR population (area under the curve [AUC] 0.71).[17] Recent studies have identified the limitations of

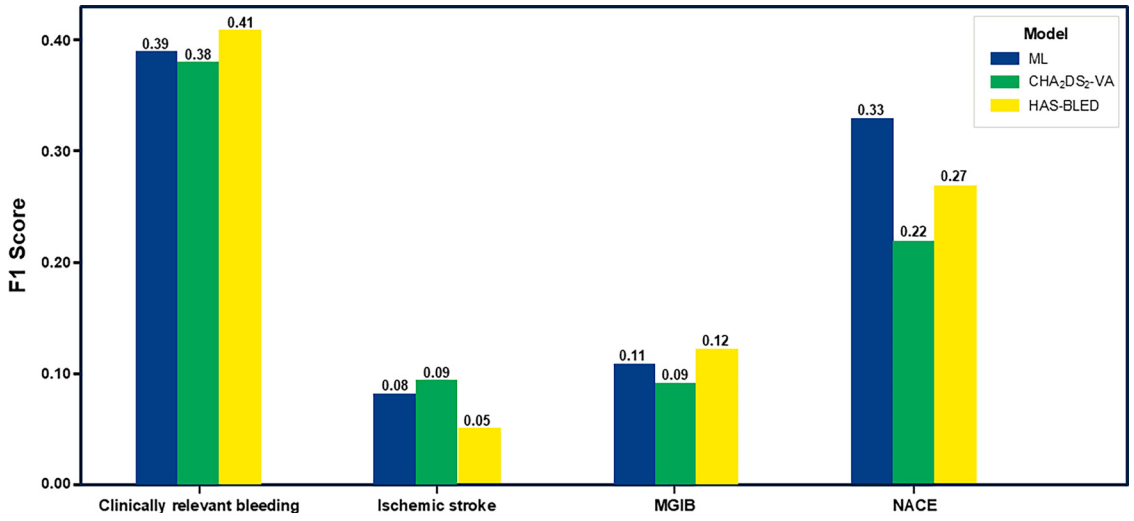


Figure 1. Performance of ML models versus CHA_2DS_2-VA and HAS-BLED in predicting clinically relevant bleeding, ischemic stroke, MGIB, and NACE in patients with AF after TAVR. AF = atrial fibrillation; CHA_2DS_2-VA = Congestive heart failure, Hypertension, age ≥75 (doubled), Diabetes, Stroke (doubled), Vascular disease, Age 65 to 74 years; HAS-BLED = Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or disposition, Labile international normalized ratio, Elderly, Drugs/alcohol concomitantly; MGIB = major gastrointestinal bleeding; ML = machine learning; NACE = net adverse clinical events; TAVR = transcatheter aortic valve replacement.

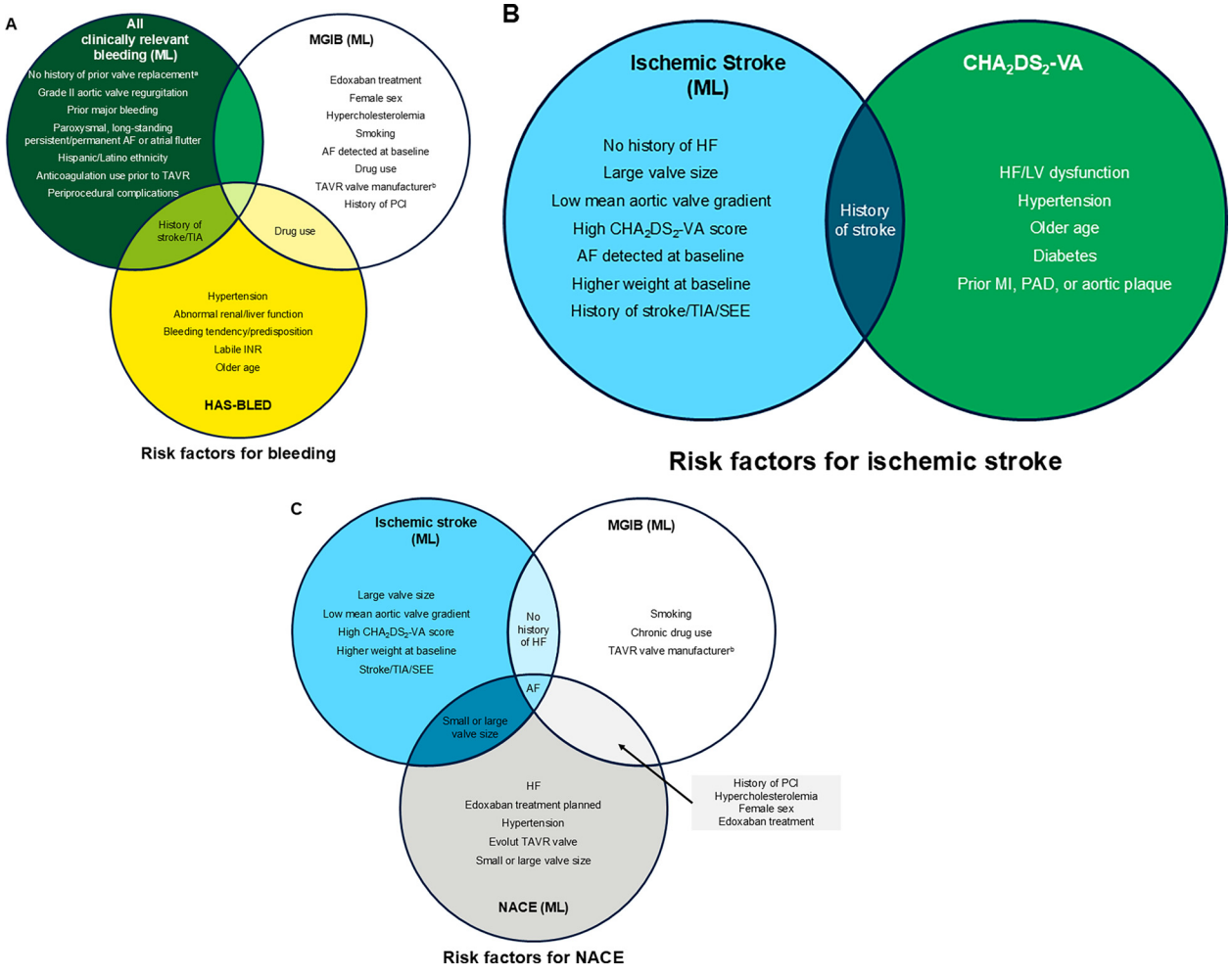


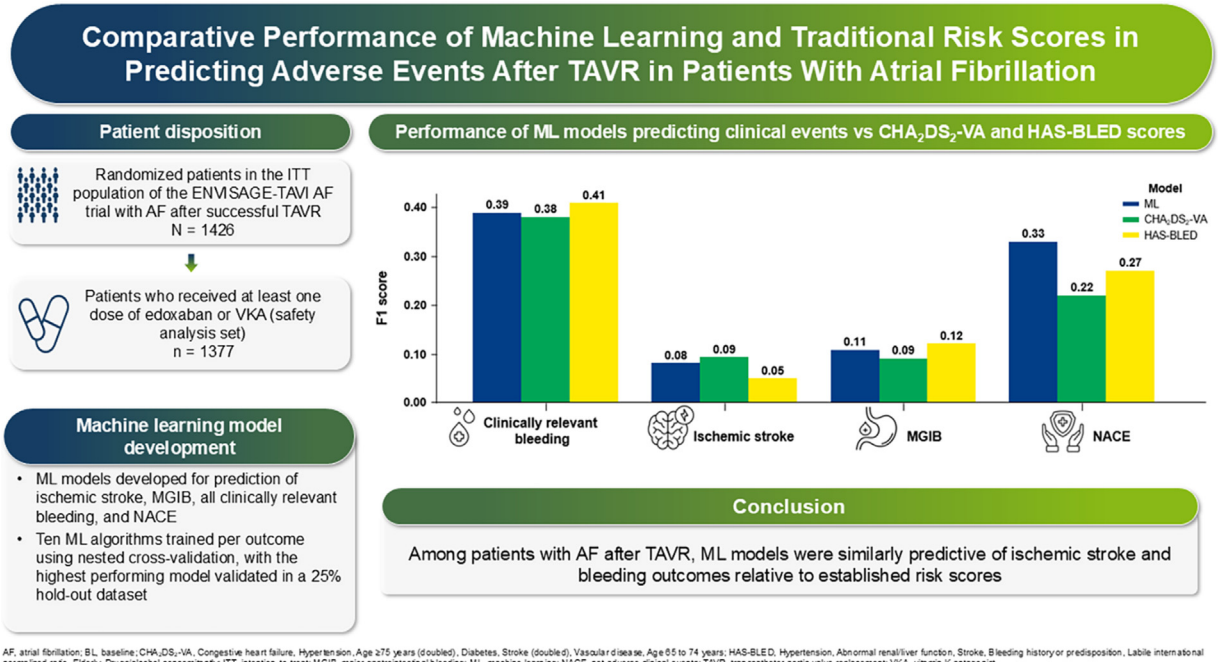
Figure 2. Overlap of risk factors identified by ML models and traditional scores (HAS-BLED, CHA_2DS_2-VA) for post-TAVR bleeding (A) and ischemic stroke (B); overlap and distinction of ML –identified risk factors for NACE, MGIB, and ischemic stroke in patients with AF following TAVR (C)Factors with the highest average SHAP values (top 10) for each outcome are reported. The HAS-BLED score as well as the constituents of the CHA_2DS_2-VA score were used as input features for the ML models. When comparing SHAP values from different ML models, some of the models were linear and some were non-linear. [a]No history of previous valve replacement before index procedure. [b]Device companies are Boston/Symetis, Edwards, Medtronic, Meril Lifesciences, NVT, and St. Jude. AC = anticoagulant; AF = atrial fibrillation; BL = baseline; CHA_2DS_2-VA = Congestive heart failure, Hypertension, Age ≥75 years (doubled), Diabetes, Stroke (doubled), Vascular disease, Age 65 to 74 years; HAS-BLED = Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly, Drugs/alcohol concomitantly; HF = heart failure; INR = international normalized ratio; LV = left ventricle; MB = major bleeding; MGIB = major gastrointestinal bleeding; MI = myocardial infarction; ML = machine learning; NACE = net adverse clinical events; PAD = peripheral artery disease; PCI = percutaneous coronary intervention; SEE = systemic embolic event; SHAP = SHapley Additive exPlanations; TAVR = transcatheter aortic valve replacement; TIA = transient ischemic attack.

traditional bleeding risk scores such as HAS-BLED in TAVR cohorts, which has led to the development of more specific models that incorporate variables relevant to TAVR patients.[18] PREDICT-TAVR is a practical 6-item bleeding risk stratification tool developed using ML and multivariate regression, which assessed over 100 clinical variables in 5,185 TAVR patients from a prospective multicenter registry.[18] The model was then validated in another prospective multicenter registry of 5,043 TAVR patients; external validation produced a 30-day AUC of 0.78.[18] While PREDICT-TAVR was not developed specifically for patients with AF after TAVR, use of oral anticoagulation is included as 1 of the 6 items conferring increased risk of bleeding at 30 days post-TAVR. It was not possible to compare the predictive ability of our ML model for NACE to a traditional clinical risk score evaluating a similar outcome, as currently no such score combines both thrombotic and bleeding risks into a single predictive metric. Our ML model for NACE effectively integrated these distinct risk factors and helped to identify high-risk patients who might otherwise be overlooked when thrombotic and bleeding risks are assessed separately.

Although ML models did not consistently outperform traditional risk scores, they provided additional insights by identifying a broad range of clinically and procedurally relevant variables associated with IS and bleeding after TAVR. A potential concern in translating ML-based risk tools into practice is whether the variables used for model development are routinely available in real-world settings. Importantly, all 180 candidate variables in our analysis were standard demographic, clinical, laboratory, and procedural parameters commonly collected at TAVR centers. Thus, our models did not depend on novel or nonroutinely measured factors but rather leveraged existing data that could be feasibly integrated into electronic health record –based risk-prediction systems. Indeed, the ML models identified variables beyond those captured by traditional scores, particularly anatomical and procedurerelated characteristics (e.g., valve type, aortic valve gradient, and valve size). In contrast, traditional risk scores predominantly reflect well-established risk factors that are important but not specific to patients undergoing TAVR. Therefore, ML models could refine risk prediction and potentially guide tailored management decisions after TAVR beyond what conventional scores provide.

Putting it all into perspective, our findings indicate that the present ML models offer rather marginal improvement in predictive performance compared to conventional risk scores for TAVR patients with AF, especially for stroke risk where event rates are low. All patients in the current study were anticoagulated, which likely mitigated stroke occurrence and contributed to the limited predictive power of both ML and traditional scores. Nevertheless, further refining of this type of methodology may pave the way for future discoveries. The ML approach provided additional insights by incorporating TAVR-specific anatomical and procedural factors beyond those captured in standard risk scores, highlighting its potential to refine risk stratification. These models could complement existing risk scores in clinical practice to better identify high-risk patients for tailored post-TAVR management; a cautious interpretation is warranted given their modest incremental benefit and the need for further validation. Future research should focus on enhancing and externally validating these ML models in larger, diverse cohorts, as well as developing integrated risk-prediction tools (e.g., combining thrombotic and bleeding risk) to guide therapy. By pragmatically integrating ML algorithms with clinical judgment and electronic health records, we can aim to improve risk assessment and outcomes for TAVR patients with AF in the coming years.

This study has several important limitations. First, it is a retrospective analysis of a randomized trial. Despite the high quality of the data used for this analysis, the findings may not be generalizable to all patient populations due to the strict inclusion and exclusion criteria. Second, although we used nested cross-validation and a hold-out test set for internal validation, we did not have an independent external validation cohort. Since all model fitting and internal validation were performed within a single dataset, discrimination and calibration indices may be biased due to study-specific variable definitions. External validation in independent cohorts is necessary. Third, our ML models were built using the variables collected in the ENVISAGE-TAVI AF trial; other important and clinically relevant procedural variables might not be captured. Fourth, the overall event rates for certain outcomes, particularly IS, were relatively low, which limited the potential of any predictive model and may explain why IS



AF, atrial fibrillation; BL, baseline; CHA_2DS_2-VA, Congestive heart failure, Hypertension, Age ≥75 years (doubled), Diabetes, Stroke (doubled), Vascular disease, Age 65 to 74 years; HAS-BLED, Hypertension, Abnormal renal/liver function, Stroke, Bleeding history or predisposition, Labile international normalized ratio, Elderly, Drugs/alcohol concomitantly; ITT, intention-to-treat; MGIB, major gastrointestinal bleeding; ML, machine learning; NACE, net adverse clinical events; TAVR, transcatheter aortic valve replacement; VKA, vitamin K antagonist.

Central Illustration

predictions were modest. Fifth, we compared our ML models to the CHA_2DS_2-VA and HAS-BLED scores as these are well-established and commonly used tools in clinical practice. However, other risk scores are available in the literature and may also be considered in future analyses. Lastly, it is important to note that although the SHAP method identifies potential predictors for a given model and a dataset, it does not inherently establish causal relationships and may be prone to spurious correlations as observed in our data. Consequently, any newly identified features should be subjected to rigorous medical scrutiny and thorough investigation before any assumptions of causality are made.

Conclusions

Among TAVR patients with AF enrolled in the ENVISAGE-TAVI AF trial, ML models demonstrated similar risk prediction for both thrombotic and bleeding outcomes compared with conventional risk scores. A ML model also effectively predicted NACE, highlighting the capacity to capture overall risk balance in this high-risk patient population. Additional studies are needed for model training and external validation on larger datasets.

Clinical Perspectives

This analysis of the ENVISAGE-TAVI AF trial demonstrates that ML algorithms provide event-risk estimates after TAVR that are comparable to the CHA_2DS_2-VA and HAS-BLED scores. However, the predictive performance for ischemic stroke, major gastrointestinal bleeding, clinically relevant bleeding, and net adverse clinical events remains modest across all methods. Using 155 clinical and procedural variables, the ML models suggest procedure specific factors, such as valve type and aortic-valve gradient, that are not otherwise captured by conventional scores, which consider far fewer variables. Consequently, this enables a more nuanced approach to pre-TAVR planning and post-TAVR management. Given the low event rates and single-trial dataset, broader development and external validation are crucial before ML-based tools can be incorporated into routine practice.

Uncited References

[19–23].

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: George Dangas reports financial support was provided by Daiichi Sankyo. Christian Hengstenberg reports financial support was provided by Daiichi Sankyo. Nicolas Van Mieghem reports financial support was provided by Daiichi Sankyo. George Dangas reports a relationship with Daiichi Sankyo that includes: consulting or advisory. Christian Hengstenberg reports a relationship with Boston Scientific that includes: speaking and lecture fees. Christian Hengstenberg reports a relationship with Edwards Lifesciences that includes: speaking and lecture fees. Christian Hengstenberg reports a relationship with Daiichi Sankyo that includes: consulting or advisory, funding grants, and speaking and lecture fees. Amanda Borrow reports a relationship with Daiichi Sankyo that includes: employment and equity or stocks. Cathy Chen reports a relationship with Daiichi Sankyo that includes: employment and equity or stocks. Martin Unverdorben reports a relationship with Daiichi Sankyo that includes: employment and equity or stocks. Rudiger Smolnik reports a relationship with Daiichi Sankyo Europe GmbH that includes: employment. Felix Just reports a relationship with Daiichi Sankyo Europe GmbH that includes: employment. Eva-Maria Fronk reports a relationship with Daiichi Sankyo Europe GmbH that includes: employment. Krishna Padmanabhan reports a relationship with Cytel that includes: employment. Christian Hengstenberg reports a relationship with Abbott Vascular that includes: funding grants. Johny Nicolas reports a relationship with National Heart Lung and Blood Institute that includes: funding grants. Christian Hengstenberg reports a relationship with Amgen that includes: funding grants. Christian Hengstenberg reports a relationship with Biosensors that includes: funding grants. Christian Hengstenberg reports a relationship with Biotronik that includes: funding grants. Christian Hengstenberg reports a relationship with Boehringer Ingelheim that includes: funding grants. Christian Hengstenberg reports a relationship with Boston Scientific that includes: funding grants. Christian Hengstenberg reports a relationship with Edwards Lifesciences that includes: funding grants. Christian Hengstenberg reports a relationship with Ferrer that includes: funding grants. Christian Hengstenberg reports a relationship with Medtronic that includes: funding grants. Christian Hengstenberg reports a relationship with Novartis that includes: funding grants. Christian Hengstenberg reports a relationship with Philips that includes: funding grants. Christian Hengstenberg reports a relationship with Siemens that includes: funding grants. Christian Hengstenberg reports a relationship with Terumo that includes: funding grants. Nicolas Van Mieghem reports a relationship with Abbott that includes: funding grants. Nicolas Van Mieghem reports a relationship with Abiomed that includes: funding grants. Nicolas Van Mieghem reports a relationship with Boston Scientific that includes: funding grants. Nicolas Van Mieghem reports a relationship with Daiichi Sankyo that includes: funding grants. Nicolas Van Mieghem reports a relationship with Edwards Lifesciences that includes: funding grants. Nicolas Van Mieghem reports a relationship with Medtronic that includes: funding grants. Nicolas Van Mieghem reports a relationship with PulseCath BV that includes: funding grants. Nicolas Van Mieghem reports a relationship with Siemens that includes: funding grants. Co-author, Christian Hengstenberg, is a clinical proctor for Boston Scientific and Edwards Lifesciences. Christian Hengstenberg reports support for attending meetings from Daiichi Sankyo. Corresponding author, George Dangas, reports support for attending meetings from Bayer and Daiichi Sankyo. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Johny Nicolas: Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **George Dangas:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Amanda Borrow:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Rüdiger Smolnik:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization. **Felix Just:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Cathy Chen:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Krishna Padmanabhan:** Writing – review & editing, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Eva-Maria Fronk:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Christian Hengstenberg:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Nicolas Van Mieghem:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization. **Martin Unverdorben:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization.

Acknowledgments

Editorial and manuscript support were provided by Stephanie Justice-Bitner, PhD, and Kimberly Dent-Ferguson, MBS, MPH, of Red Nucleus, and funded by Daiichi Sankyo.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.amjcard.2025.12.002.

References

1 Mangieri A, Montalto C, Poletti E, Sticchi A, Crimi G, Giannini F, Latib A, Capodanno D, Colombo A. Thrombotic versus bleeding risk after transcatheter aortic valve replacement: JACC review topic of the week. J Am Coll Cardiol 2019;74:2088–101.
2 Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, Cronin EM, Deswal A, Eckhardt LL, Goldberger ZD, Gopinathannair R, Gorenek B, Hess PL, Hlatky M, Hogan G, Ibeh C, Indik JH, Kido K, Kusumoto F, Link MS, Linta KT, Marcus GM, McCarthy PM, Patel N, Patton KK, Perez MV, Piccini JP, Russo AM, Sanders P, Streur MM, Thomas KL, Times S, Tisdale JE, Valente AM, Van Wagoner DR. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2024;149:e1–e156.
3 Van Gelder IC, Rienstra M, Bunting KV, Casado-Arroyo R, Caso V, Crijns H, De Potter TJR, Dwight J, Guasti L, Hanke T, Jaarsma T, Lettino M, Lochen ML, Lumbers RT, Maesen B, Molgaard I, Rosano GMC, Sanders P, Schnabel RB, Suwalski P, Svennberg E, Tamargo J, Tica O, Traykov V, Tzeis S, Kotecha D, Group ESCSD. 2024 ESC guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J 2024;45:3314–414.
4 Zhu W, He W, Guo L, Wang X, Hong K. The HAS-BLED score for predicting major bleeding risk in anticoagulated patients with atrial fibrillation: a systematic review and meta-analysis. Clin Cardiol 2015;38:555–61.
5 Teppo K, Lip GYH, Airaksinen KEJ, Halminen O, Haukka J, Putaala J, Mustonen P, Linna M, Hartikainen J, Lehto M. Comparing CHA(2)DS(2)-VA and CHA(2)DS(2)-VASc scores for stroke risk stratification in patients with atrial fibrillation: a temporal trends analysis from the retrospective Finnish AntiCoagulation in Atrial Fibrillation (FinACAF) cohort. Lancet Reg Health Eur 2024;43:100967.
6 Van Mieghem NM, Unverdorben M, Hengstenberg C, Möllmann H, Mehran R, López-Otero D, Nombela-Franco L, Moreno R, Nordbeck P, Thiele H, Lang I, Zamorano JL, Shawl F, Yamamoto M, Watanabe Y, Hayashida K, Hambrecht R, Meincke F, Vranckx P, Jin J, Boersma E, Rodés-Cabau J, Ohlmann P, Capranzano P, Kim HS, Pilgrim T, Anderson R, Baber U, Duggal A, Laeis P, Lanz H, Chen C, Valgimigli M, Veltkamp R, Saito S, Dangas GD. Edoxaban versus vitamin K antagonist for atrial fibrillation after TAVR. N Engl J Med 2021;385:2150–60.
7 Van Mieghem NM, Unverdorben M, Valgimigli M, Mehran R, Boersma E, Baber U, Hengstenberg C, Shi M, Chen C, Saito S, Veltkamp R, Vranckx P, Dangas GD. Edoxaban versus standard of care and their effects on clinical outcomes in patients having undergone transcatheter aortic valve implantation in atrial fibrillation-rationale and design of the ENVISAGE-TAVI AF trial. Am Heart J 2018;205:63–9.
8 Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost 2005;3:692–4.
9 Kaatz S, Ahmad D, Spyropoulos AC, Schulman S. Subcommittee on Control of A. Definition of clinically relevant non-major bleeding in studies of anticoagulants in atrial fibrillation and venous thromboembolic disease in non-surgical patients: communication from the SSC of the ISTH. J Thromb Haemost 2015;13:2119–26.
10 Kappetein AP, Head SJ, Généreux P, Piazza N, van Mieghem NM, Blackstone EH, Brott TG, Cohen DJ, Cutlip DE, van Es GA, Hahn RT, Kirtane AJ, Krucoff MW, Kodali S, Mack MJ, Mehran R, Rodés-Cabau J, Vranckx P, Webb JG, Windecker S, Serruys PW, Leon MB. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document. J Am Coll Cardiol 2012;60:1438–54.
11 Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, Cronin EM, Deswal A, Eckhardt LL, Goldberger ZD, Gopinathannair R, Gorenek B, Hess PL, Hlatky M, Hogan G, Ibeh C, Indik JH, Kido K, Kusumoto F, Link MS, Linta KT, Marcus GM, McCarthy PM, Patel N, Patton KK, Perez MV, Piccini JP, Russo AM, Sanders P, Streur MM, Thomas KL, Times S, Tisdale JE, Valente AM, Van Wagoner DR. Peer Review Committee M. 2023 ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2024;149:e1–e156.
12 Lundberg SMaL S-I. A unified approach to interpreting model predictions. In: In: Proceddings of the 31st International Conference on Neural Information Processing Systems; 2017. p. 4765–74.
13 Just F., Padmanabhan, K., Jinger, P., Borrow, A., Smolnik, R. Methodology for AI-driven outcome prediction for patients with atrial fibrillation after transcatheter aortic valve implantation (TAVI), 2025.
14 Abdel-Qadir H, Singh SM, Pang A, Austin PC, Jackevicius CA, Tu K, Dorian P, Ko DT. Evaluation of the risk of stroke without anticoagulation therapy in men and women with atrial fibrillation aged 66 to 74 years without other CHA2DS2-VASc factors. JAMA Cardiol 2021;6:918–25.
15 Bernardini A, Bindini L, Antonucci E, Berteotti M, Giusti B, Testa S, Palareti G, Poli D, Frasconi P, Marcucci R. Machine learning approach for prediction of outcomes in anticoagulated patients with atrial fibrillation. Int J Cardiol 2024;407:132088.
16 Conrotto F, D'Ascenzo F, D'Onofrio A, Agrifoglio M, Chieffo A, Cioni M, Regesta T, Tarantini G, Gabbieri D, Saia F, Tamburino C, Ribichini F, Cugola D, Aiello M, Sanna F, Iadanza A, Pompei E, Stolcova M, Cappai A, Minati A, Cassese M, Martinelli GL, Agostinelli A, Gerosa G, Gaita F, Rinaldi M, Salizzoni S. Predictive ability of the CHADS2 and CHA2DS2-VASc scores for stroke after transcatheter aortic balloon-expandable valve implantation: an italian Transcatheter Balloon-expandable Valve Implantation Registry (ITER) sub-analysis. Eur J Cardiothorac Surg 2016;50:867–73.
17 Honda Y, Yamawaki M, Araki M, Tada N, Naganuma T, Yamanaka F, Watanabe Y, Yamamoto M, Shirai S, Hayashida K. Impact of HAS-BLED score to predict trans femoral transcatheter aortic valve replacement outcomes. Catheter Cardiovasc Interv 2018;92:1387–96.
18 Navarese EP, Zhang Z, Kubica J, Andreotti F, Farinaccio A, Bartorelli AL, Bedogni F, Rupji M, Tomai F, Giordano A, Reimers B, Spaccarotella C, Wilczek K, Stepinska J, Witkowski A, Grygier M, Kukulski T, Wanha W, Wojakowski W, Lesiak M, Dudek D, Zembala MO, Berti S. Development and validation of a practical model to identify patients at risk of bleeding after TAVR. JACC Cardiovasc Interv 2021;14:1196–206.
19 Lee PT, Huang MH, Huang TC, Sung PS, Hsu CH, Huang YL, Lin SH, Liu PY. Improvement of oral anticoagulant prescription and long-term clinical outcomes of patients with atrial fibrillation after implementation of a clinical decision support system in outpatient practice. J Am Heart Assoc 2025:e038640.
20 Wang SV, Rogers JR, Jin Y, Bates DW, Fischer MA. Use of electronic healthcare records to identify complex patients with atrial fibrillation for targeted intervention. J Am Med Inform Assoc 2017;24:339–44.
21 Chen JY, Zhang AD, Lu HY, Guo J, Wang FF, Li ZC. CHADS2 versus CHA2DS2-VASc score in assessing the stroke and thromboembolism risk stratification in patients with atrial fibrillation: a systematic review and meta-analysis. J Geriatr Cardiol 2013;10:258–66.
22 Gao X, Cai X, Yang Y, Zhou Y, Zhu W. Diagnostic accuracy of the HAS-BLED bleeding score in VKA- or DOAC-treated patients with atrial fibrillation: a systematic review and meta-analysis. Front Cardiovasc Med 2021;8:757087.
23 Ellis RJ, Sander RM, Limon A. Twelve key challenges in medical machine learning and solutions. Intelligence-Based Med 2022;6:100068.