ESC European Society of Cardiology

European Heart Journal - Quality of Care and Clinical Outcomes (2023) **9**, 768–777
https://doi.org/10.1093/ehjqcco/qcad002

ORIGINAL ARTICLE

Machine learning for prediction of all-cause mortality after transcatheter aortic valve implantation

Jacek Kwiecinski[1,2], Maciej Dabrowski[1], Luis Nombela-Franco[3], Kajetan Grodecki[4], Konrad Pieszko[2,5], Zbigniew Chmielak[1], Anna Pylko[1], Breda Hennessey[3], Lukasz Kalinczuk[1], Gabriela Tirado-Conte[3], Bartosz Rymuza[4], Janusz Kochman[4], Maksymilian P. Opolski[1], Zenon Huczek[4], Marc R. Dweck[6], Damini Dey[2], Pilar Jimenez-Quevedo[3], Piotr Slomka[2,*,†] and Adam Witkowski[1,†]

[1]Department of Interventional Cardiology and Angiology, Institute of Cardiology, Warsaw, Poland; [2]Departments of Medicine (Division of Artificial Intelligence in Medicine) and Biomedical Sciences, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Metro 203, Los Angeles, CA 90048, USA; [3]Cardiovascular Institute, Hospital Clinico San Carlos, IdISSC, Madrid, Spain; [4]1st Department of Cardiology, Medical University of Warsaw, Warsaw, Poland; [5]Department of Interventional Cardiology and Cardiac Surgery, University of Zielona Gora, Zielona Gora, Poland; and [6]Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK

Received 20 September 2022; revised 18 November 2022; accepted 11 January 2023; online publish-ahead-of-print 13 January 2023

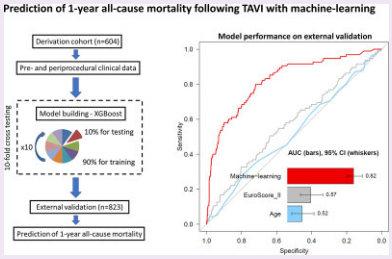
Aims Prediction of adverse events in mid-term follow-up after transcatheter aortic valve implantation (TAVI) is challenging. We sought to develop and validate a machine learning model for prediction of 1-year all-cause mortality in patients who underwent TAVI and were discharged following the index procedure.

Methods and results The model was developed on data of patients who underwent TAVI at a high-volume centre between January 2013 and March 2019. Machine learning by extreme gradient boosting was trained and tested with repeated 10-fold hold-out testing using 34 pre- and 25 peri-procedural clinical variables. External validation was performed on unseen data from two other independent high-volume TAVI centres. Six hundred four patients (43% men, 81 ± 5 years old, EuroSCORE II 4.8 [3.0–6.3]%) in the derivation and 823 patients (46% men, 82 ± 5 years old, EuroSCORE II 4.7 [2.9–6.0]%) in the validation cohort underwent TAVI and were discharged home following the index procedure. Over the 12 months of follow-up, 68 (11%) and 95 (12%) subjects died in the derivation and validation cohorts, respectively. In external validation, the machine learning model had an area under the receiver-operator curve of 0.82 (0.78–0.87) for prediction of 1-year all-cause mortality following hospital discharge after TAVI, which was superior to pre- and peri-procedural clinical variables including age 0.52 (0.46–0.59) and the EuroSCORE II 0.57 (0.51–0.64), $P < 0.001$ for a difference.

Conclusion Machine learning based on readily available clinical data allows accurate prediction of 1-year all-cause mortality following a successful TAVI.

† Equal contribution as senior author.
* Corresponding author. Tel: +310-423-4348, Fax: 310-423-0173, Email: piotr.slomka@cshs.org, Twitter: @Piotr_JSlomka

Graphical Abstract



Keywords Artificial intelligence • Machine learning • Transcatheter aortic valve implantation • Aortic stenosis

Introduction

Transcatheter aortic valve implantation (TAVI) has revolutionized the management of severe, symptomatic aortic valve stenosis.[1–4] While according to recent nationwide registry data, TAVI outcomes are improving over time across a range of important metrics, the 1-year mortality following implantation remains substantial at 10–12% in 2020.[5–7] This relatively high event rate can be largely attributed to the advanced age, frailty, and competing cardiovascular as well as non-cardiovascular risk, which all jointly affect TAVI recipients. The high comorbidity burden of patients undergoing TAVI makes prediction of outcome following a successful bioprosthesis implantation challenging. While several methods for prediction of TAVI outcomes have been proposed, these efforts have largely focused on prediction of in-hospital and/or 30-day mortality, and their performance remained limited.[8–11] Given the extensive diagnostic work-up that precedes TAVI, it is plausible that the wealth of pre- and peri-procedural data could be leveraged for robust risk stratification.

Artificial intelligence with machine learning has emerged as a powerful tool for combining several weak predictors in a single model for enhanced prediction of adverse outcomes.[12,13] Recently, gradient boosting algorithms have been shown to enhance risk stratification across a wide range of diseases and clinical scenarios providing patient-specific prediction beyond conventional risk scores.[14–16] In this study, we leveraged a state-of-the-art gradient boosting algorithm to develop a prediction model for all-cause mortality in the year following a successful TAVI procedure and validated its performance on external datasets from independent sites.

Methods

Study design

The study population included three cohorts of consecutive TAVI recipients from tertiary high-volume centres (each performing >100 procedures annually) who underwent valve implantation between January 2013 and March 2019. All patients underwent a comprehensive baseline clinical assessment with evaluation of their cardiovascular risk factor profile, including calculation of risk scores (European System for Cardiac Operative Risk Evaluation—EuroSCORE II, the France II, and OBSERVANT scores).[8,9] Only subjects who underwent a successful TAVI and were discharged home following the index procedure were included. Data from the Institute of Cardiology, Warsaw served as the derivation cohort for the machine learning model, which was then further tested on unseen external datasets from the Medical University of Warsaw and the Cardiovascular Institute, Hospital Clinico San Carlos, Madrid (validation cohort) Figure (*1*). This paper was written according to recommendations in the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.[17] The study was conducted with the approval of the local research ethics committee at the National Institute of Cardiology, Warsaw, Poland (registration number IK.NPIA.0021.1.1954/22) and in accordance with the Declaration of Helsinki.

Clinical follow-up

The primary endpoint of the study was 1-year all-cause mortality following hospital discharge after a successful TAVI procedure. Outcome information was obtained from local and national healthcare record systems. Categorization of these outcomes was performed blinded to the clinical patient data.

Machine learning

Machine learning was used to derive a joint probabilistic score that could inform the physician on the risk of 1-year all-cause mortality after TAVI and therefore facilitate planning post-discharge care and surveillance. We have therefore excluded patients who died during the index hospitalization

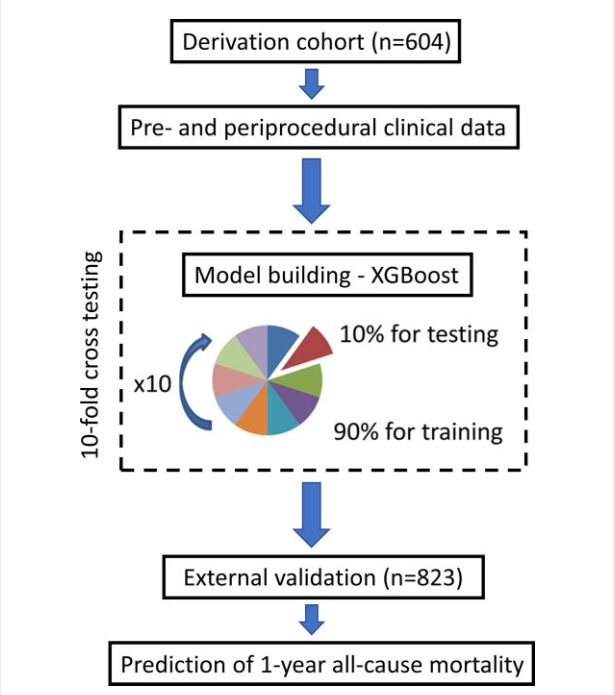


Figure 1 Study design. Our machine learning model was developed on 604 patients from the derivation cohort using 10-fold repeated cross-testing. Subsequently, the model was validated on unseen data from the validation cohort ($n = 823$).

from the analysis. The score was based on 34 pre- and 25 peri-procedural clinical variables: baseline characteristics, cardiovascular risk factors (comorbidities), echocardiography- and blood-derived biomarkers, as well as procedural aspects (access, radiation, and complications), and pre-discharge echocardiography and blood tests (Supplementary material online, *Table 1*).

Model building

XGBoost is a recent implementation of a gradient boosting algorithm, which iteratively trains a set of weak predictors (simple decision trees) using a given set of patient data, to build a combined strong classifier to identify an outcome.[14–16] For every patient, the XGBoost algorithm computes an individualized probability of outcome, considering all input variables. All variables utilized in the machine learning modelling are presented in Supplementary material online, *Table 1*. For optimal model performance, we have performed hypertuning of XGBoost parameters (Supplementary material online, *Methods*). The model configuration providing the best prediction accuracy was selected.

Internal testing

To avoid biased results, limit overfitting, and ascertain generalizability of our model using the derivation cohort, we tested the model using repeated 10-fold cross-testing, which separates training and testing data.[18] The dataset was randomly split into 10 folds with similar all-cause mortality rates in each fold (stratified 10 folds). Ten models were created each from 90% of the data, and each tested in a held-out test sample (10% of the data). These 10 held-out samples containing non-overlapping test results were subsequently concatenated to evaluate the average performance of XGBoost in unseen data.

External testing

To further validate the generalizability of this approach, we have built the model from all the data used for the internal testing. Subsequently, we have conducted external validation of this model on real-world data from two independent high-volume TAVI centres (validation cohort) (Figure *1*).

Feature importance

To elucidate the influence of each of the variables included in the machine learning model, we provided machine learning feature importance scores. Importance is the relative amount that each attribute improves the XGBoost performance measure (similar to information gain). The variable importance was determined directly from the XGBoost model separately in each fold and returned from the XGBoost model for each variable.[19]

Individualized explainability

Further in this study, we provide a description of individualized predictions made by the algorithm.[14,20] This internal XGBoost function allows identification of important patient-specific features and the role of the feature in the predicted score for the specific patient and may facilitate the clinical acceptance of the artificial intelligence approach. The individual explanation can be achieved by analysis of the specific path a subject takes in the model as in each decision stump (or split) of the model, the individual lands in one of two leaves. Each leaf is associated with a weight: one leaf decreases the risk of the event happening, and the other one increases the risk. Ultimately, this information can be graphically presented with waterfall plots.

Statistical analysis

We assessed the distribution of data with the Shapiro–Wilk test. Continuous parametric variables were expressed as mean ± standard deviation, and non-parametric data were presented as median (interquartile interval). Fisher's exact test or χ^2 test was used for the analysis of categorical variables. The performance of machine learning models and single clinical characteristics in predicting all-cause mortality was assessed using receiver-operator characteristic (ROC) analysis, and the area under the curve (AUC) values were compared with the DeLong test.[21] To evaluate the accuracy of predictors, we have also quantified the sensitivity, specificity, positive and negative predictive values of each clinical variable. For continuous variables, the Youden index was employed to define the optimal thresholds. Statistical analysis was performed with SPSS version 24 (IBM SPSS Statistics for Windows, Version 24.0, Armonk, NY: IBM Corp.) and R studio and R software version 4.01 (R Foundation for Statistical Computing, Vienna, Austria). A two-sided $P < 0.05$ was considered statistically significant.

Results

In the derivation cohort, 604 patients (43% men, 81 ± 5.1 years old, EuroSCORE II 4.8 [3.0–6.3]%) underwent TAVI and were discharged home following the index procedure. Over 12 months following the index procedure, 68 (11%) patients died. Baseline demographic, clinical, echocardiographic, and procedural characteristics of the study population are listed in Table*1*. Only a few clinical variables emerged as predictors of 1-year all-cause mortality following hospital discharge after a successful TAVI. These included baseline kidney function, platelet levels, the lowest post-procedural kidney function, left ventricular ejection fraction, length of stay in the hospital after TAVI, and the amount of packed red blood cells transfused (Table *2*). The predictive performance of these variables in isolation was, however, limited with the highest AUC (95% confidence interval) of 0.67 (0.59–0.74) for kidney function following TAVI and 0.64 (0.56–0.72) for the left ventricular ejection fraction on post-procedural echocardiography. Importantly, neither patient age nor the EuroSCORE II were significant predictors of all-cause mortality: AUC 0.51 (0.44–0.57) and 0.56 (0.49–0.63), respectively ($P = 0.83$ and $P = 0.08$). The overall variable importance for the classification of all-cause mortality is depicted in Figure *2*. While the number of packed red blood cell units transfused, the hospital length of stay, and the lowest estimated glomerular filtration rate were the top predictors, baseline blood biomarkers (creatine and platelet levels), and echocardiographic findings (both baseline and post-procedural) were also among the variables that had the greatest contribution to the machine learning model (Figure *2*).

Predictive performance

Our model was validated on a cohort of 823 consecutive patients (46% men, 82 ± 5 years old, EuroSCORE II 4.7 [2.9–6.0]%) who underwent TAVI between January 2014 and March 2019 (Table *1*). Over the 12 months of follow-up, 95 (12%) subjects died in the validation cohort. Our model had an AUC of 0.82 (0.78–0.87) for prediction of 12-month mortality, which was superior to the EuroSCORE II 0.57 (0.51–0.64) and age 0.52 (0.46–0.59), $P < 0.001$ for a difference (Figure *3*). Our model also outperformed the France II and OBSERVANT risk scores AUC of 0.58 (0.53–0.63) and 0.59 (0.54–65), respectively, $P < 0.001$ for a difference (Supplementary material online, *Figure 1*). To generate distinct clinical risk groups, we dichotomized the population according to their machine learning risk score, with the optimal cutoff for event prediction derived using the Youden index. A threshold of 15% achieved a sensitivity, specificity, and negative predictive value of 80 (72–88%), 73 (69–77%), and 96 (95–98%), respectively, for the primary endpoint. The performance of the model is also depicted on the calibration plot (Figure *4*), which allows the evaluation of the agreement between machine learning scores and the actual distribution of the observed events. In the external validation, 77 (91%) events occurred in patients with the machine learning score within the top two deciles. We have also

Table 1 Baseline characteristics of study participants

	Derivation cohort ($n = 604$)	Validation cohort ($n = 823$)		P-value
		Medical University of Warsaw	Hospital Clinico San Carlos	
Age, years	82 [77–86]	81 [76–84]	83 [79–83]	0.37
Females, *n*	345 (57%)	108 (52%)	337 (55%)	0.24
Weight, kg	73.5 ± 16	75 ± 15	71 ± 15	0.27
Body mass index	27.1 ± 5.3	27.2 ± 4.5	27.9 ± 5.2	0.41
Bicuspid aortic valve	53 (9%)	17 (8%)	43 (7%)	0.50
EuroScore II	4.8 [3.0–6.3]	5.1 [3.3–7.4]	4.5 [2.9–7.3]	0.29
Comorbidities and past medical history				
Diabetes	226 (37%)	72 (35%)	208 (34%)	0.53
Atrial fibrillation	206 (34%)	79 (38%)	248 (40%)	0.31
History of ACS	130 (21.5%)	35 (17%)	97 (16%)	0.25
History of PCI	175 (29%)	74 (36%)	149 (24%)	0.19
History of CABG	77 (13%)	22 (11%)	63 (10%)	0.22
History of valve surgery	24 (4%)	10 (5%)	31 (5%)	0.28
History of pacemaker implantation	86 (14%)	36 (17%)	82 (13%)	0.33
History of a cerebrovascular accident	77 (13%)	21 (10%)	57 (9%)	0.23
Baseline biomarkers				
Creatinine, mg/dL	1.1 [0.9–1.4]	1.1 [0.8–1.5]	1.0 [0.8–1.3]	0.51
eGFR, mL/m^2	55 [42–70]	54 [36–79]	59 [44–74]	0.38
Haemoglobin, g/dL	12.4 [11.2–13.3]	11.4 [10.3–12.8]	12.3 [11.0–13.3]	0.43
Platelets, *n*/dL	174 [138–219]	193 [150–231]	187 [156–230]	0.32
Baseline echocardiography				
Left ventricular ejection fraction, %	55 [50–65]	58 [46–64]	55 [50–65]	0.26
Effective orifice area, cm^2	0.6 [0.5–0.8]	0.7 [0.5–0.9]	0.6 [0.5–0.8]	0.36
Peak transvalvular pressure gradient, mmHg	86 [70–104]	77 [61–96]	76 [61–90]	0.29
Mean transvalvular pressure gradient, mmHg	49 [39–62]	43 [34–51]	44 [38–55]	0.27
Procedure—valve implantation				
Contrast media volume, mL	190 [150–200]	200 [150–250]	164 [137–200]	0.38
Fluoroscopy time, min	24 [20–35]	31 [21–45]	25 [18–31]	0.23
Radiation dose, mGy	1051 [608–1737]	1251 [810–2004]	1134 [682–1680]	0.28
Peri- and post-procedural outcomes				
Major vascular complication	30 (5%)	12 (6%)	25 (4%)	0.27
Minor vascular complication	57 (9%)	10 (5%)	69 (11%)	0.034
Life threatening bleeding	24 (4%)	10 (5%)	22 (4%)	0.31
Major bleeding	80 (13%)	21 (10%)	27 (4%)	0.001
Minor bleeding	86 (14%)	19 (9%)	95 (15%)	0.11
Total RBC concentrate transfused, units	0 [0–1]	0 [0–1]	0 [0–1]	0.38
Peri-procedural myocardial infarction	4 (1%)	3 (1%)	8 (1%)	0.12
Peri-procedural stroke	11 (2%)	3 (1%)	9 (1%)	0.38
Coronary occlusion	4 (1%)	2 (1%)	4 (1%)	0.24
Annulus rupture	3 (1%)	0	1 (0%)	0.30
Pacemaker implantation	92 (15%)	38 (18%)	96 (16%)	0.41
Post-procedural biomarkers				
Minimum haemoglobin, g/dL	10.2 [9.3–11.1]	10.7 [9.7–12.0]	10.1 [8.9–11.0]	0.36
Minimum platelets, *n*/dL	105 [81–137]	124 [96–162]	116 [93–147]	0.29
Minimum eGFR, mL/m^2	58 [42–84]	52 [38–81]	54 [37–70]	0.34
Post-procedural echocardiography				
Left ventricular ejection fraction, %	60 [50–65]	58 [48–65]	60 [54–67]	0.42
Peak transvalvular pressure gradient, mmHg	16 [12–23]	14 [9–17]	17 [12–23]	0.26
Mean transvalvular pressure gradient, mmHg	7 [3–11]	8 [6–11]	8 [6–12]	0.61
Aortic insufficiency ≥ moderate	85 (14%)	18 (9%)	82 (13%)	0.16
Hospitalization length (days)	9 [7–15]	8 [6–14]	6 [5–9]	0.07

Statistics presented: median [quartile 1–quartile 3], *n* (%). Abbreviations: ACS, acute coronary syndrome; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting; and eGFR, estimated glomerular filtration rate.

Table 2 Prediction of all-cause mortality

	Receiver-operator curve analysis						
	Area under the curve (95% confidence intervals)	*P*-value	Sensitivity, %	Specificity, %	Positive predictive value, %	Negative predictive value, %	Accuracy, %
Baseline characteristics							
Age	0.51 (0.44–0.57)	0.83	49 (38–60)	63 (59–67)	18 (13–22)	88 (85–92)	61 (60–62)
Weight	0.50 (0.42–0.58)	0.98	30 (19–41)	73 (69–77)	16 (10–23)	82 (77–86)	63 (62–64)
Body mass index	0.51 (0.43–0.57)	0.80	26 (16–36)	84 (80–89)	22 (13–31)	85 (81–90)	74 (73–75)
Bicuspid aortic valve	0.51 (0.44–0.58)	0.72	38 (29–47)	59 (51–66)	25 (16–35)	64 (58–71)	61 (52–70)
EuroScore II	0.56 (0.49–0.63)	0.08	47 (36–58)	36 (32–40)	11 (8–14)	80 (75–86)	37 (36–38)
Comorbidities and past medical history							
Diabetes	0.57 (0.49–0.65)	0.09	51 (40–61)	64 (60–68)	11 (6–15)	84 (81–87)	62 (61–63)
Atrial fibrillation	0.54 (0.47–0.61)	0.24	26 (17–36)	64 (60–68)	11 (6–15)	84 (80–88)	59 (58–60)
History of PCI	0.53 (0.46–0.61)	0.37	29 (20–38)	67 (62–73)	17 (11–23)	76 (70–83)	61 (60–62)
History of valve surgery	0.52 (0.45–0.59)	0.64	25 (18–32)	63 (60–67)	18 (11–26)	82 (77–87)	59 (58–60)
History of pacemaker implantation	0.51 (0.43–0.58)	0.88	17 (9–25)	76 (72–81)	16 (9–24)	77 (73–79)	67 (66–68)
History of a cerebrovascular accident	0.53 (0.46–0.61)	0.42	20 (15–25)	79 (72–86)	24 (19–30)	84 (80–88)	64 (63–65)
Baseline biomarkers							
Creatinine	0.62 (0.56–0.69)	<0.001	72 (62–82)	56 (51–62)	22 (17–27)	92 (89–95)	58 (57–59)
eGFR	0.61 (0.55–0.68)	0.001	57 (47–68)	24 (20–27)	11 (8–14)	77 (71–84)	28 (27–29)
Haemoglobin	0.54 (0.48–0.60)	0.22	38 (27–49)	50 (46–55)	11 (7–14)	83 (79–88)	48 (47–49)
Platelets	0.57 (0.50–0.63)	0.05	63 (53–74)	53 (48–57)	19 (14–23)	89 (86–93)	54 (53–55)
Baseline echocardiography							
Left ventricular ejection fraction	0.56 (0.49–0.62)	0.11	76 (67–85)	11 (8–13)	12 (9–15)	73 (62–83)	20 (19–21)
Effective orifice area	0.51 (0.43–0.59)	0.74	18 (9–26)	85 (82–89)	17 (8–25)	86 (83–89)	75 (74–76)
Peak transvalvular pressure gradient	0.53 (0.45–0.61)	0.40	11 (4–18)	83 (79–86)	10 (4–16)	85 (82–88)	73 (72–74)
Mean transvalvular pressure gradient	0.52 (0.44–0.60)	0.65	10 (3–17)	83 (80–87)	9 (3–15)	85 (82–88)	73 (72–74)
Procedure—valve implantation							
Contrast media volume	0.55 (0.48–0.63)	0.17	64 (54–74)	42 (38–46)	15 (11–19)	88 (84–92)	45 (44–46)
Fluoroscopy time	0.55 (0.47–0.62)	0.22	48 (37–59)	63 (59–67)	17 (12–22)	88 (85–92)	61 (60–62)
Radiation dose	0.56 (0.48–0.64)	0.10	24 (18–30)	86 (83–89)	22 (16–29)	78 (72–85)	72 (71–73)
Peri-procedural and in-hospital outcomes							
Major vascular complication	0.55 (0.48–0.62)	0.13	13 (6–20)	94 (91–98)	37 (19–54)	86 (83–90)	81 (80–82)
Minor vascular complication	0.52 (0.45–0.59)	0.56	23 (17–29)	79 (73–86)	28 (24–33)	75 (70–81)	71 (70–72)
Life threatening bleeding	0.57 (0.49–0.64)	0.06	16 (8–24)	93 (92–95)	54 (34–74)	87 (85–90)	80 (79–81)
Total RBC concentrate transfused	0.62 (0.55–0.70)	<0.001	23 (14–32)	94 (92–97)	52 (36–69)	88 (85–91)	81 (78–85)

Table 2 Continued

	Receiver-operator curve analysis						
	Area under the curve (95% confidence intervals)	*P*-value	Sensitivity, %	Specificity, %	Positive predictive value, %	Negative predictive value, %	Accuracy, %
Post-procedural biomarkers							
Minimum haemoglobin	0.56 (0.50–0.61)	0.16	89 (83–95)	7 (4–9)	13 (10–16)	80 (67–92)	18 (17–19)
Minimum platelets	0.53 (0.44–0.61)	0.53	88 (81–96)	18 (14–21)	15 (11–19)	91 (85–98)	28 (27–29)
Minimum eGFR	0.67 (0.59–0.74)	<0.001	15 (7–24)	53 (49–58)	5 (2–7)	81 (76–86)	49 (48–50)
Post-procedural echocardiography							
Left ventricular ejection fraction	0.64 (0.56–0.72)	0.01	37 (23–52)	36 (31–41)	7 (4–10)	81 (75–88)	36 (35–37)
Peak transvalvular pressure gradient	0.57 (0.50–0.65)	0.16	27 (17–37)	82 (78–86)	19 (11–26)	86 (83–90)	72 (71–73)
Mean transvalvular pressure gradient	0.53 (0.44–0.61)	0.72	28 (18–37)	82 (78–87)	21 (13–28)	86 (83–90)	73 (72–74)
Aortic insufficiency ≥ moderate	0.54 (0.45–0.63)	0.42	32 (29–36)	61 (58–64)	16 (12–20)	63 (59–68)	51 (50–52)
Hospitalization length	0.62 (0.53–0.71)	0.014	73 (63–83)	33 (29–37)	15 (11–18)	88 (84–93)	38 (37–39)

Abbreviations: eGFR, estimated glomerular filtration rate; PCI, percutaneous coronary intervention; and RBC, red blood cell.

developed a machine learning model based only on pre-procedural data; however, its performance was limited—AUC 0.64 (0.59–0.69).

Individualized explainability of the prediction

To further clarify and explain the machine learning predictions, we also provide waterfall plots that highlight the contribution of the predictors for individual patients. As demonstrated in Figure *5* and Supplementary material online, *Figure 1*, the machine learning model can accurately predict events in elderly patients who present with a high EuroSCORE II as well as relatively younger individuals who are at a low risk according to the surgical risk scores.

Discussion

In this multicentre machine learning study, we have demonstrated that by leveraging state-of-the-art artificial intelligence and readily available clinical data, it is feasible to predict 1-year all-cause mortality in patients who have undergone successful TAVI and were discharged from the hospital following the procedure. Our approach outperformed individual clinical metrics, as well as the EuroSCORE II risk score, which despite having been developed to predict short-term adverse events following surgical valve replacement and because of the lack of bespoke risk stratification tools for TAVI, is frequently used in patients, managed percutaneously. The observed efficacy of our model suggests that machine learning could play an important clinical role in evaluating prognostic risk in patients after TAVI. By stratifying TAVI recipients according to the risk of adverse events, it could facilitate offering a more intense follow-up regimen to those at the highest risk of adverse outcomes after the index procedure. Ultimately, it can enable cost-effective allocation of medical resources and might contribute to improving outcomes following TAVI.

In view of the complexity of patients undergoing TAVI, artificial intelligence with machine learning emerges as an ideal tool for combining the information provided by a large set of weak predictors for robust risk stratification. The XGBoost algorithm has been successfully implemented for risk prediction in a wide range of clinical scenarios.[14,15] It enables the incorporation of numerous predictors into the model even when these variables are correlated—a major limitation with conventional regression analyses. Our model was developed using conservative internal 10-fold repeated validation, which limits overfitting and ascertains generalizability. Importantly, the model attained high accuracy in external validation (combined data from two independent tertiary clinical centres) by objectively integrating pre- and peri-procedural data—a task that is challenging to accomplish at the point of care.

While surgical risk scores have been widely used to stratify different groups of patients for comparative clinical trials between SAVR and TAVI, they were developed and validated on series of patients undergoing SAVR and therefore do not necessarily encompass the diverse co-morbidities that have an adverse impact on outcomes in TAVI recipients. Only a few studies have aimed to develop methods for risk stratification beyond 30 days after TAVI and less than a handful included peri-procedural data in the models.[8–12] Moreover, their performance was limited, or the target population was only extreme-risk and high-risk patients.[22–26] Our model addresses this important clinical gap. By providing patient-specific risk estimates, it has the potential to enable tailored post-discharge patient surveillance after TAVI, ultimately providing an opportunity for a more cost-effective allocation of resources.

Prediction of TAVI outcomes is challenging due to the competing cardiovascular and non-cardiovascular risks, as well as the fact that patients are typically fairly homogeneous in terms of conventional metrics that act as robust predictors in different clinical settings

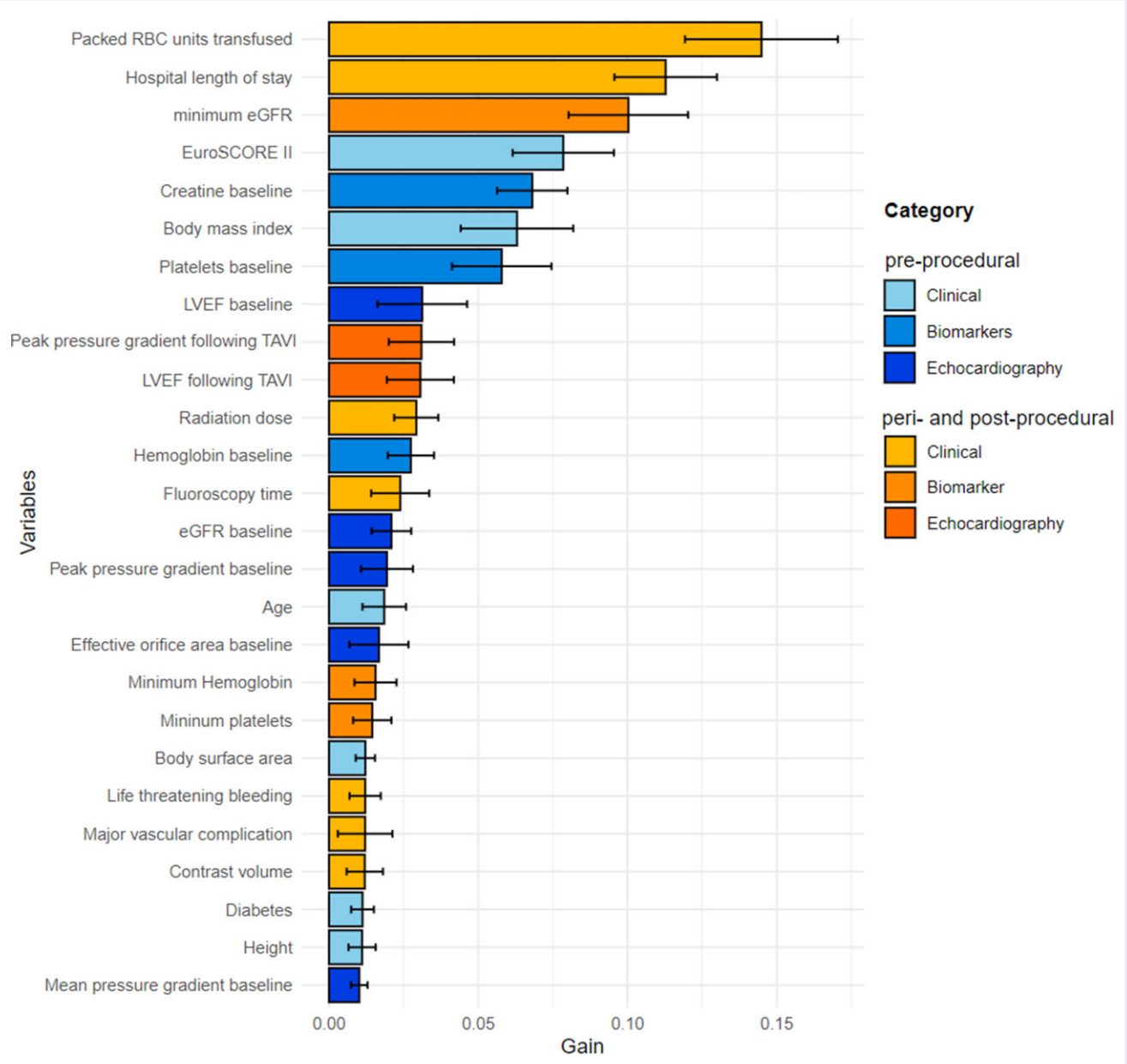


Figure 2 Feature importance for the machine learning model predicting all-cause mortality in TAVI recipients. The solid bars and error bars represent the mean gain and standard deviation derived from the distribution of the importance within 10 folds of the cross-testing for each variable. Abbreviations: eGFR, estimated glomerular filtration rate; LVEF, left ventricular ejection fraction; PPG, peak pressure gradient; and RBC, red blood cell.

(i.e. age).[27] While these are not adequately accounted for by the EuroSCORE II, machine learning has the potential to overcome this issue and risk-stratify the challenging cohort of TAVI patients. In our study, none of the single metrics acted as a strong predictor of all-cause mortality on receiver-operator curve characteristic analysis (AUC $<$ 0.70, Table 2). As shown previously, the lack of association between age and survival in our study population may be caused by a ceiling effect in the overall intermediate-risk cohort.[27]

Importantly, the machine learning model also provides insights into the top predictors by ranking the relative contribution from each variable for a unique patient. This has the potential to improve physicians' confidence in the machine learning results and may potentially help to overcome the perception of artificial intelligence as a 'black box'.[28] Our model represents a substantial improvement in risk stratification. While multiple variables have been shown to act as independent predictors of adverse outcomes following TAVI, prior studies did not separate data for deriving independent predictors and testing their clinical utility, and therefore the predictive value of such parameters may be overestimated and not applicable to TAVI recipients globally. In contrast, in our study we employed rigorous external validation, establishing the generalizability of our model on unseen data from two independent centres.

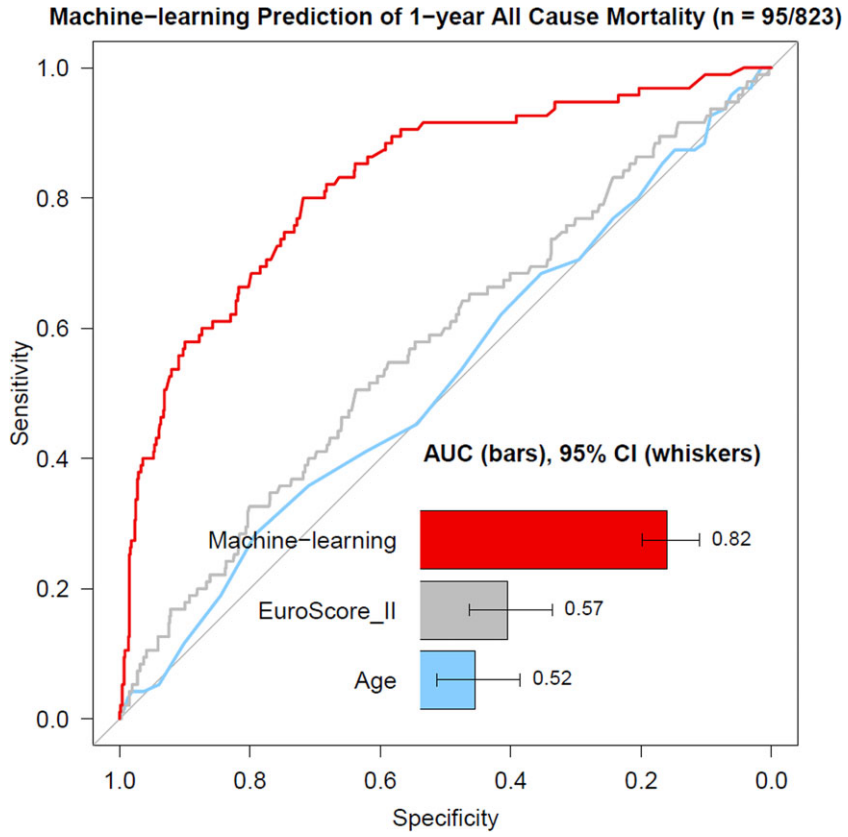


Figure 3 Prediction of 1-year all-cause mortality on external testing. Receiver-operating characteristic curves for prediction of 1-year all-cause mortality following hospital discharge after successful transcatheter aortic valve implantation. The machine learning XGBoost model had a significantly higher area under the curve for all-cause mortality prediction than an established risk score (EuroSCORE II) and conventional predictors of outcomes (including age), $P < 0.001$.

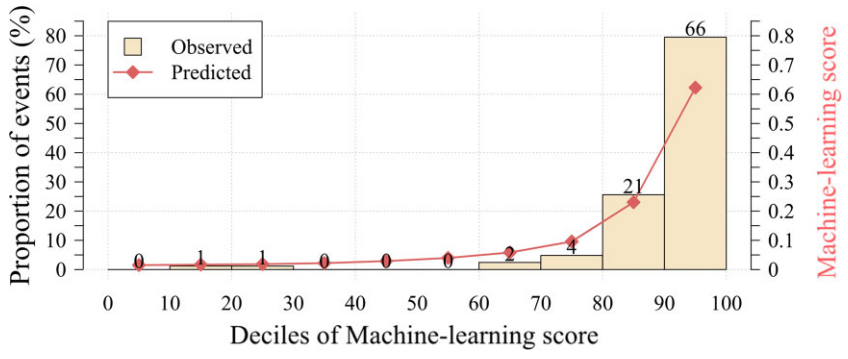


Figure 4 Calibration plot for machine learning XGBoost model. The calibration plot shows the relationship between the observed and predicted proportion of events, grouped by decile of risk. The XGBoost model showed good calibration with the observed 1-year risk of all-cause mortality.

Limitations

Our study has notable strengths and weaknesses. Our model was trained and tested on data from tertiary TAVI centres that reflect everyday clinical practice (a real-world setting). It is based on readily available data, and therefore the proposed approach could easily benefit patients without the need for additional testing or tedious data crunching. Indeed, our machine learning approach could be easily incorporated into clinical practice. At the time of discharge from the hospital after a successful TAVI, it could inform the physician on the

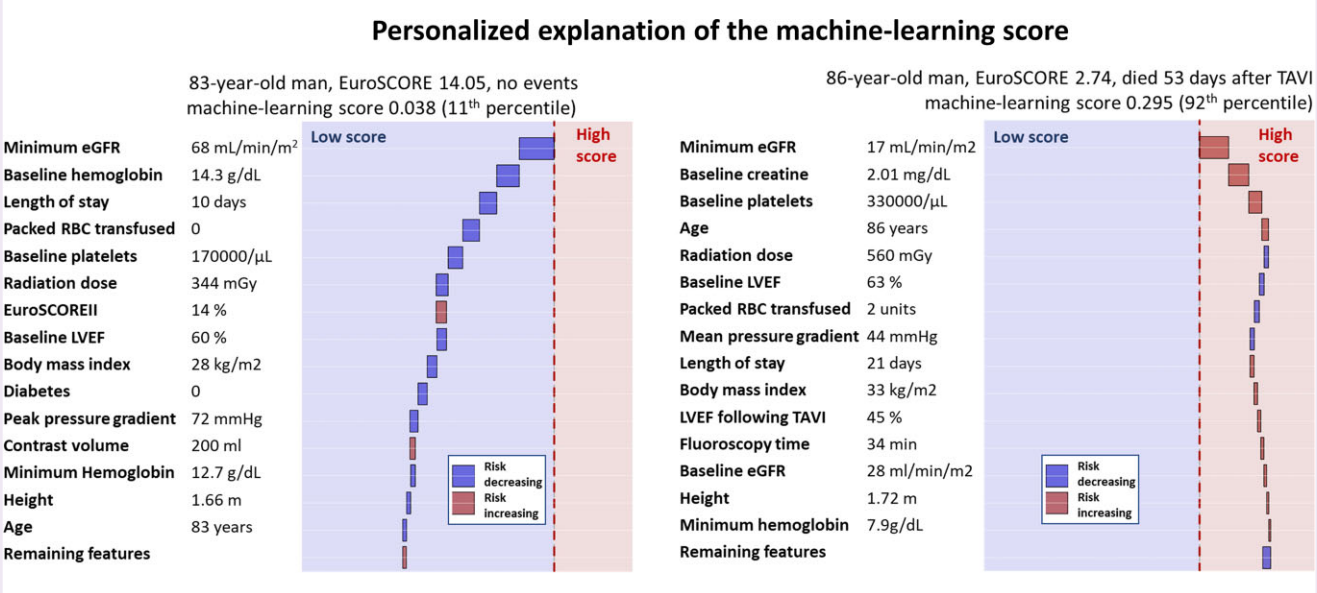


Figure 5 Individual prediction of all-cause mortality with explainable artificial intelligence. Explanations of individual prediction with subject-specific feature importance for an 83-year-old man who survived over a year following TAVI and an 86-year-old man who died 53 days after TAVI. The *x*-axis corresponds to the machine learning risk score. The arrows represent the influence of each covariate on the overall prediction; blue and red arrows indicate whether the associated parameters decrease (blue) or increase (red) the risk of future events. The combination of all covariates' influence provides the final machine learning risk score. The red and blue colours provide the separation between low and high machine learning risk scores. Abbreviations: eGFR, estimated glomerular filtration rate; LVEF, left ventricular ejection fraction; and RBC, red blood cell.

risk of all-cause mortality during the first year following the procedure, enabling patient-specific post-discharge care planning. While our databases lack STS-PROM scores, we have provided the EuroSCORE II, which was shown to have comparable discrimination and calibration to STS-PROM in patients receiving aortic valve replacement.[29,30] In the derivation cohort, we included patients who underwent TAVI in 2013–19, ascertaining a large population for establishing our model but inevitably also including a relatively large proportion of high-risk patients who currently represent a minority of TAVI recipients. Our model did not include scores that characterize patients' frailty, concomitant coronary artery disease, or the perception of their health status. Additionally, due to the observational nature of our study, we acknowledge the inherent risk for selection bias and residual confounding, which is, however, limited given the multicentre character of the current study. Finally, while prediction of outcomes following hospital discharge is valuable, our study does not address the need for a pre-procedural tool for prediction of adverse outcomes that could facilitate selection of patients for TAVI. Given the limited performance of models based exclusively on pre-procedural clinical data, it appears that inclusion of advanced cardiac imaging data might be necessary.[31,32] Ultimately, robust models based on clinical and imaging pre-procedural data could have a significant impact on clinical practice.

Conclusions

In conclusion, machine learning based on readily available clinical data allows accurate prediction of 1-year all-cause mortality following TAVI. The machine learning model could potentially be used to guide the intensity of patients' follow-up and post-discharge care after TAVI.

Supplementary material

Supplementary material is available at *European Heart Journal—Quality of Care and Clinical Outcomes* online.

Funding

This research was funded by the National Science Centre, Poland (grant 2021/41/B/NZ5/02630). K.G. is supported by the Foundation for Polish Science. M.R.D. is a recipient of the Sir Jules Thorn Award for Biomedical Research Award (2015) and is supported by the British Heart Foundation (FS/14/78/31020). This research was supported in part by grants R01HL135557, R01HL148787, and R01HL151266 from the National Heart, Lung, and Blood Institute (NHLBI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: The authors declare that they have no relevant conflict of interest or material/financial interests that relate to the research described in this paper.

Data availability

The data underlying this article were provided by the Institute of Cardiology, Warsaw, the Medical University of Warsaw, and the Cardiovascular Institute, Hospital Clinico San Carlos, Madrid. Data can be shared on request to the corresponding author with permission from the Institute of Cardiology, the Medical University of Warsaw, the Cardiovascular Institute, Hospital Clinico San Carlos, and the Polish/Spanish Data Protection Agency.

References

1. Cribier A, Eltchaninoff H, Bash A, Borenstein N, Tron C, Bauer F *et al.* Percutaneous transcatheter implantation of an aortic valve prosthesis for calcific aortic stenosis: first human case description. *Circulation* 2002; **106**:3006–3008.
2. Mack MJ, Leon MB, Thourani VH, Makkar R, Kodali SK, Russo M *et al.* Transcatheter aortic valve replacement with a balloon-expandable valve in low-risk patients. *N Engl J Med* 2019; **380**: 1695–1705.
3. Popma JJ, Deeb GM, Yakubov SJ, Mumtaz M, Gada H, O'Hair D *et al.* Transcatheter aortic-valve replacement with a self-expanding valve in low-risk patients. *N Engl J Med* 2019; **380**:1706–1715.
4. Søndergaard L, Ihlemann N, Capodanno D, Jørgensen TH, Nissen H, Kjeldsen BJ *et al.* Durability of transcatheter and surgical bioprosthetic aortic valves in patients at lower surgical risk. *J Am Coll Cardiol* 2019 **73**;:546–553.
5. Carroll JD, Mack MJ, Vemulapalli S, Herrmann HC, Gleason TG, Hanzel G *et al.* STS-ACC TVT registry of transcatheter aortic valve replacement. *J Am Coll Cardiol* 2020;**76**:2492–2516.
6. Werner N, Zahn R, Beckmann A, Bauer T, Bleiziffer S, Hamm CW *et al.* Patients at intermediate surgical risk undergoing isolated interventional or surgical aortic valve implantation for severe symptomatic aortic valve stenosis. *Circulation* 2018;**138**:2611–2623.
7. Bekeredjian R, Szabo G, Balaban Ü, Bleiziffer S, Bauer T, Ensminger S *et al.* Patients at low surgical risk as defined by the Society of Thoracic Surgeons Score undergoing isolated interventional or surgical aortic valve implantation: in-hospital data and 1-year results from the German Aortic Valve Registry (GARY). *Eur Heart J* 2019;**40**:1323–1330.
8. Iung B, Laouénan C, Himbert D, Eltchaninoff H, Chevreul K, Donzeau-Gouge P *et al.* Predictive factors of early mortality after transcatheter aortic valve implantation: individual risk assessment using a simple score. *Heart* 2014;**100**:1016–1023.
9. Capodanno D, Barbanti M, Tamburino C, D'Errigo P, Ranucci M, Santoro G *et al.* A simple risk tool (the OBSERVANT score) for prediction of 30-day mortality after TAVI. *Am J Cardiol* 2014;**113**:1851–1858.
10. Edwards FH, Cohen DJ, O'Brien SM, Peterson ED, Mack MJ, Shahian DM *et al.* Development and validation of a risk prediction model for in-hospital mortality after transcatheter aortic valve replacement. *JAMA Cardiol* 2016;**1**:46–52.
11. Martin GP, Sperrin M, Ludman PF, de Belder MA, Redwood SR, Townend JN *et al.* Novel United Kingdom prognostic model for 30-day mortality following transcatheter aortic valve implantation. *Heart* 2018;**104**:1109–1116.
12. Hernandez-Suarez DF, Kim Y, Villablanca P, Gupta T, Wiley J, Nieves-Rodriguez BG *et al.* A machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. *JACC: Cardiovasc Interv* 2019;**12**:1328–1338.
13. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP *et al.* Artificial intelligence in cardiovascular imaging *J Am Coll Cardiol* 2019;**73**: 1317–1335.
14. Commandeur F, Slomka PJ, Goeller M, Chen X, Cadet S, Razipour A *et al.* Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study. *Cardiovasc Res* 2020;**116**:2216–2225.
15. Kwiecinski J, Tzolos E, Meah MN, Cadet S, Adamson PD, Grodecki K *et al.* Machine learning with 18F-sodium fluoride PET and quantitative plaque analysis on CT angiography for the future risk of myocardial infarction. *J Nucl Med* 2022;**63**: 158–165.
16. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. New York, NY:Association for Computing Machinery, pp. 785–794.https://doi.org/10.1145/2939672.2939785.
17. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;**61**:344–349.
18. Kim JH. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. *Comput Stat Data Anal* 2009;**53**:3735–3745.
19. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning Data Mining, Inference and Prediction*. New York: Springer. 2001; 367.
20. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, August 13-17, 2016, San Francisco, CA. pp. 785–794.
21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–845.
22. Agasthi P, Ashraf H, Pujari SH, Girardo ME, Tseng A, Mookadam F *et al.* Artificial intelligence trumps TAVI2-SCORE and CoreValve score in predicting 1-year mortality post-transcatheter aortic valve replacement. *Cardiovasc Revasc Med* 2021;**24**:33–41.
23. Lantelme P, Eltchaninoff H, Rabilloud M, Souteyrand G, Dupré M, Spaziano M *et al.* Development of a risk score based on aortic calcification to predict 1-year mortality after transcatheter aortic valve replacement. *JACC: Cardiovasc Imaging* 2019;**12**:123–132.
24. Hermiller JB, Jr, Yakubov SJ, Reardon MJ, Deeb GM, Adams DH, Afilalo J *et al.* Predicting early and late mortality after transcatheter aortic valve replacement. *J Am Coll Cardiol* 2016;**68**:343–352.
25. Lopes RR, van Mourik MS, Schaft EV, Ramos LA, Baan J, Jr, Vendrik J *et al.* Value of machine learning in predicting TAVI outcomes. *Neth Heart J* 2019;**27**:443–450.
26. Penso M, Pepi M, Fusini L, Muratori M, Cefalù C, Mantegazza V *et al.* Predicting long-term mortality in TAVI patients using machine learning techniques. *J Cardiovasc Dev Dis* 2021;**8**:44. doi: 10.3390/jcdd8040044.
27. Martinsson A, Nielsen SJ, Milojevic M, Redfors B, Omerovic E, Tønnessen T *et al.* Life expectancy after surgical aortic valve replacement. *J Am Coll Cardiol* 2021;**78**:2147–2157.
28. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019;**1**: 206–215.
29. Wang TK, Choi DH, Stewart R, Gamble G, Haydock D, Ruygrok P *et al.* Comparison of four contemporary risk models at predicting mortality after aortic valve replacement. *J Thorac Cardiovasc Surg* 2015;**149**:443–448.
30. Wendt D, Thielmann M, Kahlert P, Kastner S, Price V, Al-Rashid F *et al.* Comparison between different risk scoring algorithms on isolated conventional or transcatheter aortic valve replacement. *Ann Thorac Surg* 2014;**97**:796–802.
31. Kwiecinski J, Chin CWL, Everett RJ, White AC, Semple S, Yeung E *et al.* Adverse prognosis associated with asymmetric myocardial thickening in aortic stenosis. *Eur Heart J Cardiovasc Imaging* 2018;**19**:347–356.
32. Kwiecinski J, Tzolos E, Cartlidge TRG, Fletcher A, Doris MK, Bing R *et al.* Native aortic valve disease progression and bioprosthetic valve degeneration in patients with transcatheter aortic valve implantation. *Circulation* 2021;**144**:1396–1408.