Check for updates

Harnessing the power of artificial intelligence in predicting all-cause mortality in transcatheter aortic valve replacement: a systematic review and meta-analysis

Faizus Sazzad[1]*, Ashlynn Ai Li Ler[1], Mohammad Shaheryar Furqan[2], Linus Kai Zhe Tan[1], Hwa Liang Leo[3], Ivandito Kuntjoro[4], Edgar Tay[4,5] and Theo Kofidis[1]

[1]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [2]Department of Biomedical Informatics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [3]Department of Biomedical Engineering, College of Design and Engineering, National University of Singapore, Singapore, Singapore, [4]Department of Cardiology, National University Heart Centre, Singapore, National University Hospital, Singapore, Singapore, [5]Asian Heart & Vascular Centre (AHVC), Mount Elizabeth Medical Centre, Singapore, Singapore

Objectives: In recent years, the use of artificial intelligence (AI) models to generate individualised risk assessments and predict patient outcomes post-Transcatheter Aortic Valve Implantation (TAVI) has been a topic of increasing relevance in literature. This study aims to evaluate the predictive accuracy of AI algorithms in forecasting post-TAVI mortality as compared to traditional risk scores.

Methods: Following the Preferred Reporting Items for Systematic Reviews and Meta-analyses for Systematic Reviews (PRISMA) standard, a systematic review was carried out. We searched four databases in total—PubMed, Medline, Embase, and Cochrane—from 19 June 2023–24 June, 2023.

Results: From 2,239 identified records, 1,504 duplicates were removed, 735 manuscripts were screened, and 10 studies were included in our review. Our pooled analysis of 5 studies and 9,398 patients revealed a significantly higher mean area under curve (AUC) associated with AI mortality predictions than traditional score predictions (MD: −0.16, CI: −0.22 to −0.10, $p < 0.00001$). Subgroup analyses of 30-day mortality (MD: −0.08, CI: −0.13 to −0.03, $p = 0.001$) and 1-year mortality (MD: −0.18, CI: −0.27 to −0.10, $p < 0.0001$) also showed significantly higher mean AUC with AI predictions than traditional score predictions. Pooled mean AUC of all 10 studies and 22,933 patients was 0.79 [0.73, 0.85].

Conclusion: AI models have a higher predictive accuracy as compared to traditional risk scores in predicting post-TAVI mortality. Overall, this review demonstrates the potential of AI in achieving personalised risk assessment in TAVI patients.

Registration and protocol: This systematic review and meta-analysis was registered under the International Prospective Register of Systematic Reviews (PROSPERO), under the registration name "All-Cause Mortality in Transcatheter Aortic Valve Replacement Assessed by Artificial Intelligence" and registration

---

Abbreviations
AI, artificial intelligence; ANN, artificial neural network; CI, confidence interval; GB, Gradient boosting; LR, logistic regression; MD, mean difference; MLP, multilayer perceptron; TAVI, transcatheter aortic valve implantation.

number CRD42023437705. A review protocol was not prepared. There were no amendments to the information provided at registration.

Introduction

Transcatheter aortic valve implantation (TAVI) is a crucial procedure in treating severe aortic stenosis, which is characterised by the narrowing of the aortic valve (1). TAVI provides a minimally invasive alternative to open-heart surgery for older patients and those with numerous comorbidities (2) with faster recovery times, shorter hospital stays, and less procedural risks (3–6). In patients with severe aortic stenosis, TAVI has been found to reduce symptoms, as well as increase quality of life and overall survival rates (7, 8). The emphasis of this paper is to predict the risk for TAVI, given that it has already been proven to have multiple benefits.

The mortality rate after TAVI might vary depending on a number of factors, including patient characteristics, comorbidities, and intra-procedural concerns. For intermediate-risk patients undergoing TAVI, longer-term follow-up in the PARTNER 2 study revealed a death rate of 26.2% at 5 years (9, 10). The SURTAVI trial showed that all-cause mortality was 31.3% at 5 years using the CoreValve self-expanding prosthesis in intermediate-risk patients (11). In a similar vein, the CoreValve U.S. Pivotal High-Risk Study found that high-risk patients had a 5-year mortality rate of 47.9% (12), and a separate study found the mortality rate to be 58.8% (13). The NOTION trial demonstrated that in low-risk patients, the mortality rate of TAVI with the first-generation CoreValve self-expanding prosthesis was 27.6% after 5 years of follow up (14). The PARTNER 3 trial also showed that the 5-year all-cause mortality for TAVI was 10.0% (15). One of the more recent trials, Evolut trial, found that the 4-year mortality for TAVI using self-expanding CoreValve was 10.7% (16). A separate meta-analysis has also found the all-cause mortality to be higher in low-risk patients undergoing TAVI over surgical aortic valve replacement (27.5% vs. 17.3%) (17). Thus, given the high probability of fatality and differences in mortality depending on the risk levels of patients, prediction prior to the start of the procedure is increasingly necessary.

Improving risk assessment and patient outcomes may be possible with the application of artificial intelligence (AI) in predicting death after TAVI. In-depth patient data analysis by AI algorithms has the ability to uncover pertinent trends and risk factors that can help forecast post-TAVI death (18). These models can help identify high-risk individuals who might need more measures or more intensive postoperative surveillance. On the other hand, AI models could also identify patients who may not benefit from TAVI, given that the risk of mortality outweighs the benefits of undergoing the procedure.

Thus, our goal in analysing the accuracy of AI-generated death forecasts in TAVI procedures was to determine how well AI algorithms can predict mortality outcomes for patients undergoing TAVI. This evaluation aims to assess the effectiveness of AI models in predicting post-TAVI death rates and evaluate the accuracy of these forecasts.

Methods

We conducted a systematic review following guidelines from the Preferred Reporting Items for Systematic Reviews and Meta-analyses for systematic review (PRISMA) standard (19). In total, we performed our search on 4 databases, including PubMed, Medline, Embase and Cochrane, from the date of inception to 24 October 2023. Across all databases, combinations of different search terms, including Medical Subject Headings (MeSH) terms were generated. The following search strings were used: "Transcatheter aortic valve replacement AND artificial intelligence", "Transcatheter aortic valve replacement AND machine learning", "Transcatheter aortic valve replacement AND deep learning", "Transcatheter aortic valve implantation AND artificial intelligence", "Transcatheter aortic valve implantation AND machine learning", "Transcatheter aortic valve implantation AND deep learning", "Aortic stenosis AND artificial intelligence", "Aortic Stenosis AND machine learning" and "Aortic stenosis AND deep learning". The search was perfomed using MeSH terms only.

Inclusion and exclusion criteria

Any retrospective studies that reported the use of AI to predict post-operative mortality in TAVI patients were included in our analysis. There were no randomised controlled trials identified in our search. We excluded articles that used AI to predict aortic stenosis, intra-cardiac parameters such as aortic valve annulus, studies on heart murmurs, hypertrophic cardiomyopathy, paediatric studies, narrative articles and conference abstracts. We also excluded studies on complicated TAVI patients (e.g., TAVI with infective endocarditis or cancer) and articles that used AI to predict other post-TAVI parameters such as cerebrovascular complications, pacemaker implantation, heart failure, readmission, length of hospital stay and bleeding.

Study selection

To determine suitability for inclusion of each study, we first assessed the studies by their titles and abstracts and then retrieved the full-text records should the study either fulfil the inclusion criteria or if the reviewer was uncertain of the article's suitability. In order to ensure reproducibility of our study selection, the studies were independently screened and evaluated by three reviewers. All disagreements were solved by consensus amoung the reviewers with no modification of the search and inclusion criteria.

Quality of evidence and risk of bias

GradePro quality of evidence assessment software was used to evaluate the included studies as illustrated in the Cochrane handbook of reviews (20). The risk of bias in all observational cohort studies was also assessed according to guidelines from the Cochrane handbook. Risk of bias was evaluated using the Risk of Bias in Non-randomised Studies of Interventions (ROBINS-I) tool (21).

Outcomes of interest

Data from each article was collected by two authors (AL, LT). The following variables were abstracted for analysis: Authors, year of publication, study type, patient sample size, post-TAVI mortality, age, female, AI algorithms used. Our primary outcome measure was the area under curve (AUC) value [mean and 95% confidence interval (CI)] of AI models predicting post-TAVI all-cause mortality. We included data on intra-hospital mortality, 30-day mortality, 1-year mortality and 5-year mortality in our analyses. For studies that reported AUC values of both an internal and external validation cohort, data from the latter was abstracted for analysis.

Statistical analysis

Studies that reported data comparing AI and traditional scores were analysed using Review Manager 5.3 (RevMan 5.3) software (22). Due to the limitations of RevMan 5.3 in pooling AUC values, reported means were converted to negative values and inputted into the software. We calculated the mean difference (MD) as the outcome effect measure used in our double-arm meta-analysis. For studies that reported 95% CI, the in-built RevMan calculator was used to estimate the missing standard deviations (SD). For single-arm meta-analysis, the STATA 17 software was used to pool study data and generate forest plots (23). In order to adjust for statistical heterogeneity across all study populations, all data was analysed using random effects models. All data is reported in mean and 95% CI [lower limit, upper limit]. In our subgroup analysis, we divided the studies by intra-hospital, 30-day, 1-year and 5-year mortality. For the 1-year mortality subgroup, gradient boosting and extreme gradient

boosting models were grouped together as the models are similar and we had insufficient data to compare each individually. Similarly, non-gradient-boosting algorithms (random forest, decision tree, artificial neural network, multilayer perceptron) were also grouped together for subgroup analysis.

AI algorithms

The AI algorithms of interest to this study included random forest (RF), gradient boosting (GB), artificial neural network (ANN), multilayer perceptron (MLP) and logistic regression (LR). This section includes short descriptions of each algorithm. Further elaboration on the pros and cons of each algorithm can be found in the discussion section.
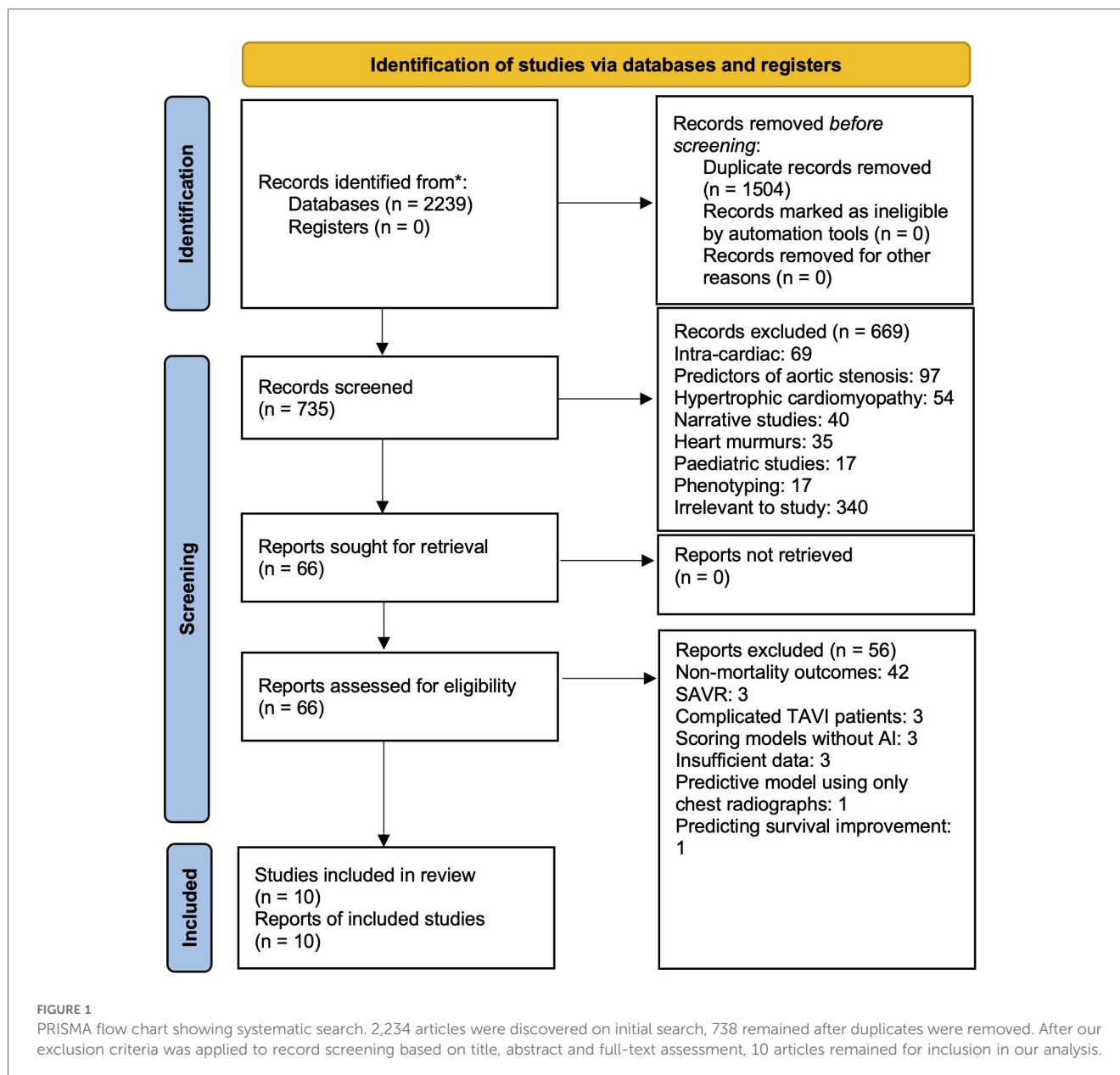
The RF model constructs a series of decision trees and utilises bagging and randomisation of predictors in order to accurately predict outcomes (24). The GB model involves an ensemble of weak learner decision trees. Each weak learner progresses through a stepwise progression, where previous iterations of weak learners are combined into a successive strong learner, correcting for the errors of the preceding learner each time (25, 26). The ANN algorithm is a mathematical model developed based on the transmission of signals amoung neural networks in biological nervous systems. An input signal is channelled through a collection of nodes, which are artificial neurons (27). Each node analyses the input signal and then transmits an output to each of its connected neurons, mimicking the transmission of action potentials across neurons and synapses in the human brain. The nodes are also organised into layers. The input signal travels from the first input layer to the last output layer, undergoing different transformations at each layer. The MLP algorithm is a class of feedforward ANN that is comprised of an input layer, one or more hidden layers and an output layer. The LR algorithm fits a logistic function onto a dataset and predicts the probability of an independent variable, such as mortality, from a dependent variable. The maximimum likelihood estimation method is most commonly used to maximise the likelihood function and find the optimal fit of the model (28).

Results

In total, our systematic search yielded 2,239 records. After 1,504 duplicate records were removed, 735 remained for title and abstract review. The high number of duplicates was likely due to our search strings throughout the various databases producing similar articles. Based on our exclusion criteria, we eliminated 669 studies and retrieved full-text articles for 66 articles. Finally, a further 56 articles were excluded based on full-text assessment, leaving 10 studies (29–38) for data abstraction (Figure 1).

Risk of bias assessment

We conducted a risk of bias assessment on the included studies. Overall, all 10 studies were retrospective cohort studies

FIGURE 1
PRISMA flow chart showing systematic search. 2,234 articles were discovered on initial search, 738 remained after duplicates were removed. After our exclusion criteria was applied to record screening based on title, abstract and full-text assessment, 10 articles remained for inclusion in our analysis.

and were thus prone to the inherent bias associated with observational study designs. No serious risk of bias was detected in any of the included studies (Table 1).

Summary of included studies

Of the 10 included studies, 6 (30, 31, 33–36) reported data on the RF algorithm, 4 (29, 32, 35, 37) on GB, 4 (30, 31, 35, 38) on ANN and MLP, and 3 (31, 36, 38) on LR. 5 studies (30, 31, 35, 36, 38) reported using more than 1 AI model for mortality prediction. The traditional scores used in the studies included the STS score (30, 33), TAVI$_2$-SCORE (29), EuroSCORE II (32, 38) and CoreValve score (34). A summary of the predictive variables used in each study is provided in Supplementary Table S1.

Use of AI algorithms

A summary of the different AI algorithms, as well as their pros and cons, is provided in Figure 2. The RF model, which divides the dataset into subgroups using bootstrap sampling and produces numerous decision trees from the same subset, is one of the AI techniques utilized in the included studies. The final decision tree is then produced by combining these trees. GB, another used approach, sequentially merges weak learning models in a step-wise process. By raising the weight of incorrect predictions in each iteration to enhance succeeding models, the goal is to produce an ensemble of models with a minimum amount of prediction errors. An artificial neural network, specifically an MLP has at least three layers of nodes: an input layer, a hidden layer, and an output layer. Backpropagation is used to change weights and reduce prediction errors throughout the learning

TABLE 1 Summary of studies and gradePro quality of evidence.

Study	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Patients (n)	Mortality (%)	Risk score (%)	Age (years)	Female (%)	AI algorithms used
Agasthi et al. (29)	R/O	Not serious	Not serious	Not serious	Not serious	1,055	14	STS score 7.9±5.1 (Alive) 9.7±5.3 (Deceased)	80.9±8.0	42	GB
Gomes et al. (30)	R/O	Not serious	Not serious	Not serious	Not serious	451	5.5	STS score 5.1±3.4	82.7±5.7	51.7	ANN, RF, support vector machine
Hernandez-Suarez et al. (31)	R/O	Not serious	Not serious	Not serious	Not serious	10,883	3.6	-	81.0±8.5	47.7	ANN, RF, LR, Naïve Bayes
Kwiecinski et al. (32)	R/O	Not serious	Not serious	Not serious	Not serious	823	12	EuroScoreII 4.8 [3.0–6.3] (Derivation cohort) 5.1 [3.3–7.4] (Validation cohort: Medical University of Warsaw) 4.5 [2.9–7.3] (Validation cohort: Hospital Clinico San Carlos)	82.0±5.0	54	GB
Leha et al. (33)	R/O	Not serious	Not serious	Not serious	Not serious	22,283 (Cross-validation) 5,864 (Generalisation)	3.3	STS score 5.9±4.6	81.0±6.1	53.2	RF
Lertsanguansinchai et al. (34)	R/O	Not serious	Not serious	Not serious	Not serious	178	4.1	STS score 7.79±5.51	81.6±8.3	56.2	Decision tree
Lopes et al. (35)	R/O	Not serious	Not serious	Not serious	Not serious	1,160 (Centre A) 631 (Centre B)	10	-	81.0±7.0 (Centre A Survived, n=1,039)	44 (Centre A survived, n=1,039)	ANN (narrow and wide), RF, GB Catboost
Mamprin, Lopes et al. (36)	R/O	Not serious	Not serious	Not serious	Not serious	1,300 (AMC centre) 631 (CZE-TU centre)	9.9	-	81.0±7.0 (AMC Survived, N=1,171)	56 (AMC Survived, N=1,171)	RF, LR, Catboost
Mamprin, Zelis et al. (37)	R/O	Not serious	Not serious	Not serious	Not serious	270	11.1	-	80.7±6.2	48	GB
Penso et al. (38)	R/O	Not serious	Not serious	Not serious	Not serious	471	45	EuroScoreII 16 (10–21)	81.0±6.0	63.7	ANN (MLP), RF, GB, LR

R/O, retrospective/observational; AMC, Amsterdam University Medical Centre; ANN, artificial neural network; CZE-TU, collaboration between Catharina Ziekenhuis Eindhoven and Technische Universiteit Eindhoven; GB, gradient boosting; LR, logistic regression; MLP, multilayer perceptron; RF, Random forest.
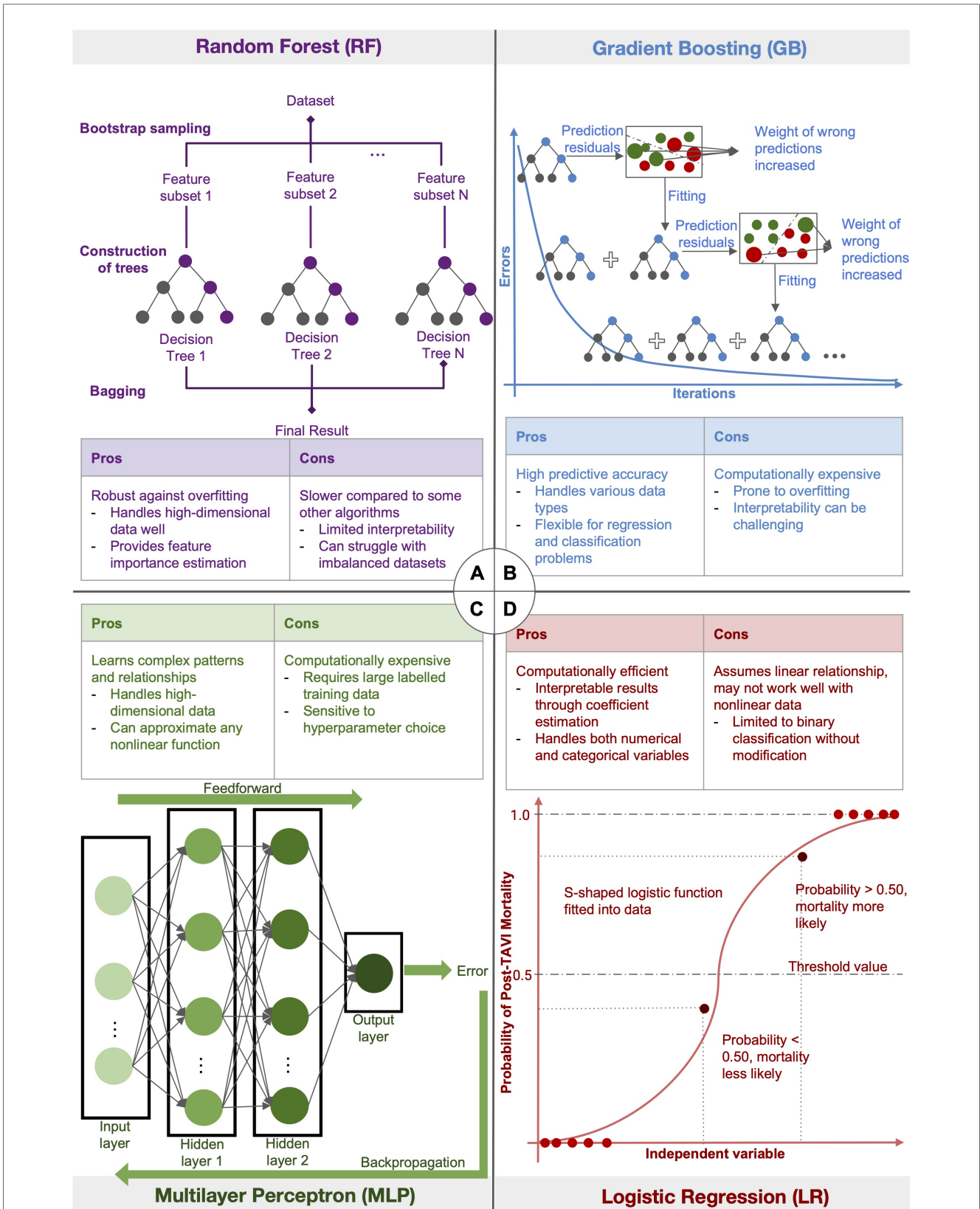
FIGURE 2
Diagram illustrating the different AI algorithms: figure summarising four of the different AI algorithms used in the included studies. (A) The random forest model first divides the dataset into subsets using bootstrap sampling, then generates different decision trees from the same subset. The trees are then averaged to generate the final decision tree. (B) Gradient Boosting uses a stage-wise progression to combine weak learning models sequentially to produce an ensemble of different models with minimal prediction errors. In each iteration, the weight of wrong predictions is increased in order to improve the learning model in the successive iteration. (C) Multilayer Perceptron, a type of artificial neural network, inputs data into at least 3 layers of nodes: an input layer, a hidden layer and an output layer. Backpropogation of data is used for learning to minimise prediction errors. (D) Logistic Regression fits a logistic function onto a dataset. Maximum likelihood estimation is used to produce a curve with maximum likelihood.
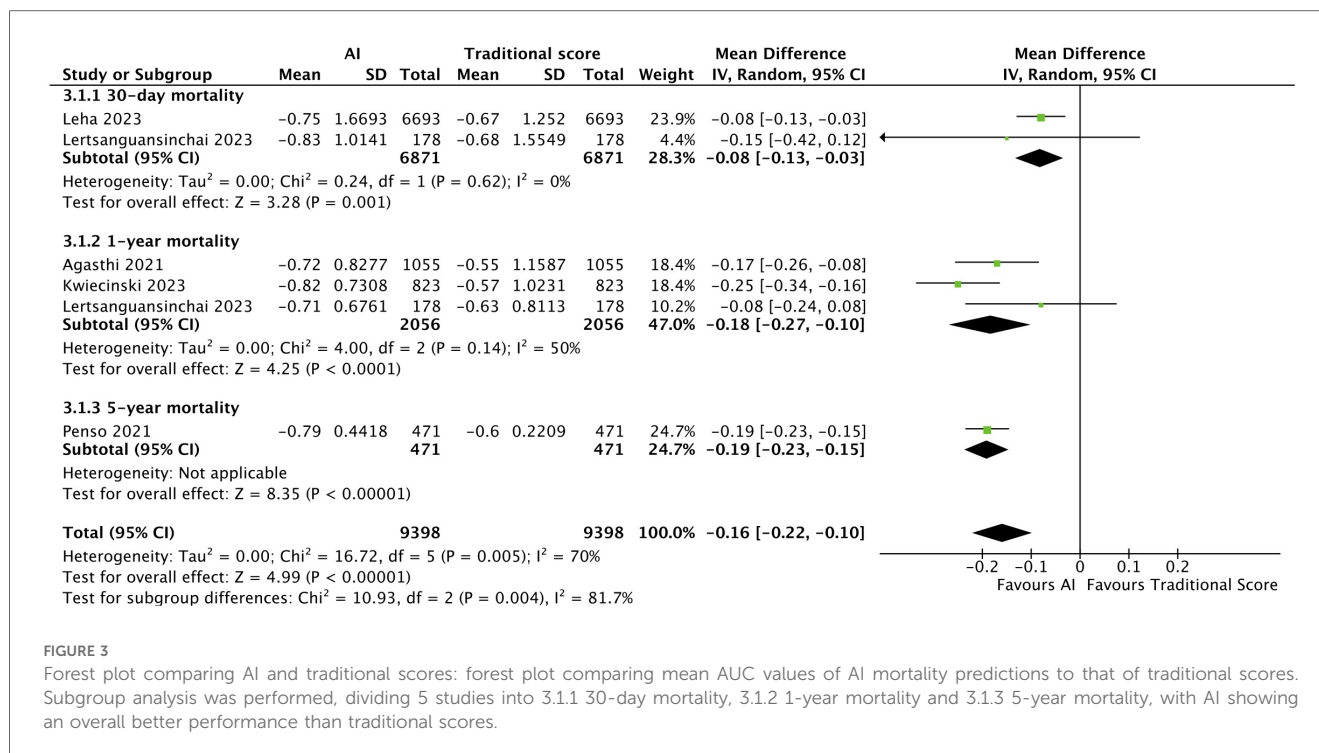
FIGURE 3

Forest plot comparing AI and traditional scores: forest plot comparing mean AUC values of AI mortality predictions to that of traditional scores. Subgroup analysis was performed, dividing 5 studies into 3.1.1 30-day mortality, 3.1.2 1-year mortality and 3.1.3 5-year mortality, with AI showing an overall better performance than traditional scores.

phase after data is fed into the network. Lastly, LR adapts a logistic function to the dataset. The likelihood of the observed data is maximized by using maximum likelihood estimate to draw the curve. For binary classification tasks, this approach is frequently employed. Further elaboration on the different AI algorithms can be found in the Supplementary materials and individual AI methods are also categorized in the Supplementary Table S2.

Meta-analysis of mortality

All 10 studies were subjected to meta-analysis. 5 studies (29, 32–34, 38) that reported data comparing the predictive ability of AI with traditional clinical risk scores were included in a two-arm meta-analysis. The results from all 10 studies were pooled in a single-arm meta-analysis.

Meta-analysis comparing predictive ability of AI vs. traditional scores

From our pooled analysis of 5 studies and a total of 9,398 patients, we observed a significantly higher mean AUC in cohorts where post-TAVI mortality was predicted with AI than when traditional scores were used on the same population (MD: −0.16, CI: −0.22 to −0.10, $p < 0.00001$, I2: 70%). 30-day mortality subgroup analysis of 2 studies (6,871 patients) (33, 34) also showed a significantly higher mean AUC with AI predictions than with traditional score predictions (MD: −0.08, CI: −0.13 to −0.03, $p = 0.001$, I2: 0%). Similar findings were observed in the 1-year mortality subgroup, consisting of 2 studies (2,056 patients) (29, 32), with AI showing an overall better performance than traditional scores (MD: −0.18, CI: −0.27 to −0.10, $p < 0.0001$, I2: 50%). (Figure 3.).

Meta-analysis of pooled AUC means of 10 studies

The single-arm meta-analysis of all 10 included studies, and a combined cohort of 22,933 patients, showed a pooled mean AUC of 0.79 [0.73, 0.85]. Due to the high overall heterogeneity (I2 = 99.06%), a subgroup analysis was conducted, separating the studies into intra-hospital mortality, 30-day mortality, 1-year mortality with gradient boosting, 1-year mortality with non-gradient-boosting and 5-year mortality. The pooled mean AUC values of 2 studies reporting intra-hospital mortality was 0.95 [0.90, 1.00], I2: 88.29%. 2 studies with 30-day mortality outcomes had a mean AUC value of 0.75 [0.72, 0.79], I2: 0%. For the 1-year mortality subgroup, 3 studies featuring the use of gradient boosting algorithms had a pooled AUC of 0.79 [0.72, 0.86], I2: 91.97%, while that of the 2 studies that used non-gradient-boosting models was 0.68 [0.67, 0.69], I2:0.19%. For 5-year mortality, 1 study reported an AUC of 0.79 [0.75, 0.83]. Even after subgroup analysis the heterogeneities of the intra-hospital and 1-year mortality with gradient boosting subgroups were still high (Figure 4). This was likely due to the differing parameters used to train the AI models, discordant datasets, and/or the fact that AI model performance varies with the training dataset. A further elaboration on the heterogeneity can be found in the discussion section. Finally, the results of our single-arm meta-analysis of traditional risk scores reported in 5 studies demonstrated a pooled mean AUC value of 0.61 [0.56, 0.65]. The traditional risk score AUC for 30-day and 1-year mortality AUC was 0.67 [0.64, 0.70] and 0.57 [0.53, 0.61], respectively. For 5-year mortality, the traditional risk score AUC reported in 1 study was 0.60 [0.57, 0.64] (Figure 5).
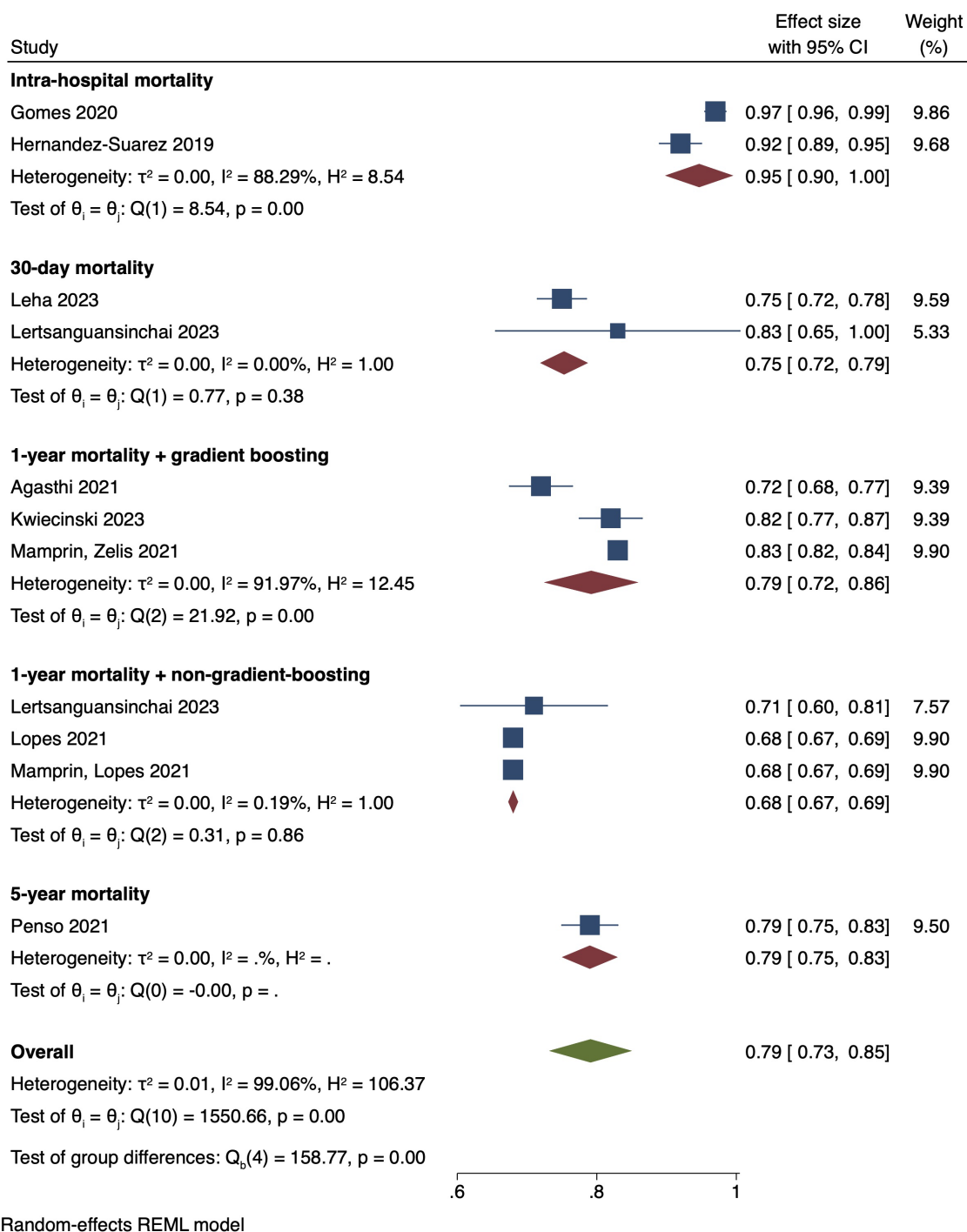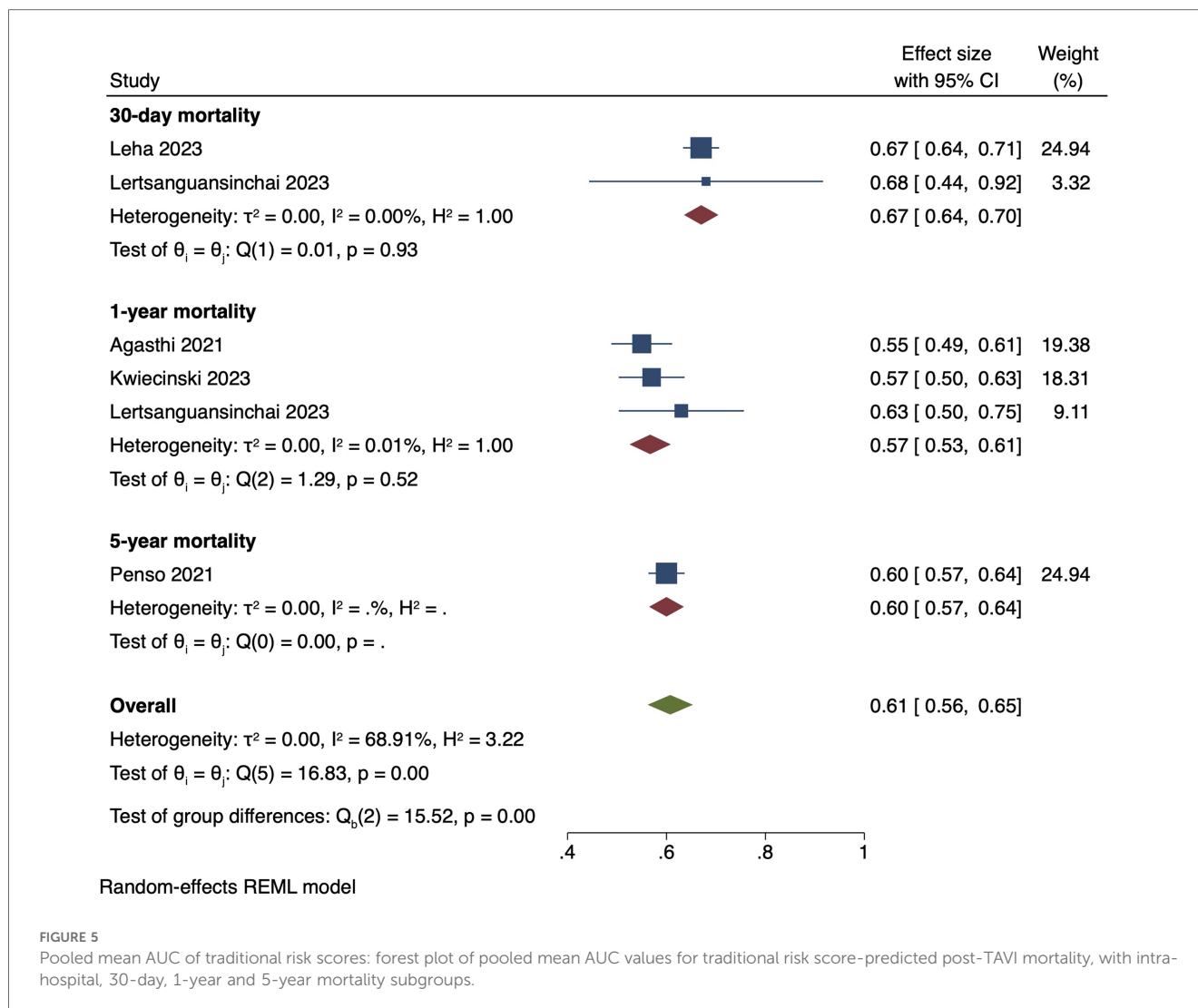
FIGURE 4
Pooled mean AUC of included studies: forest plot of pooled mean AUC values for AI-predicted post-TAVI mortality, with intra-hospital, 30-day, 1-year and 5-year mortality subgroups.

Discussion

To the best of our knowledge, this is the first meta-analysis in literature comparing AI models against traditional scores in predicting post-TAVI mortality. Overall, the results of our double-arm meta-analysis demonstrate that applying AI to predict death in TAVI cases has a higher predictive accuracy than conventional clinical scoring techniques. In particular, the 30-day (AUC: 0.75) and 1-year mortality (AUC: 0.79) with non-gradient-boosting algorithm subgroups both demonstrated high mean AUC values with low heterogeneity. In our combined analysis of traditional risk scores, the pooled mean AUC value for intra-hospital mortality, 30-day mortality, 1-year mortality with gradient boosting, 1-year mortality without gradient

FIGURE 5
Pooled mean AUC of traditional risk scores: forest plot of pooled mean AUC values for traditional risk score-predicted post-TAVI mortality, with intra-hospital, 30-day, 1-year and 5-year mortality subgroups.

boosting, and 5-year mortality was 0.95 [0.90, 1.00] 0.75 [0.72, 0.79], 0.79 [0.72, 0.86], 0.68 [0.67, 0.69] and 0.79 [0.75, 0.83], respectively. Hence from gross comparison of short- and long-term AUC alone, AI-based models may fare better than current scoring methods at correctly predicting death outcomes in TAVI patients. This suggests that AI models have the potential to aid clinicians in predicting post-TAVI mortality in patients and act as an adjunct or alternative to traditional clinical scores. AI models are able to process large amounts of diverse patient data and can be modelled to continuously analyse new data in order to make accurate predictions in real time. This is a significant advantage over traditional scores, which can only utilise a limited number of variables to predict outcomes.

In the study by Kwiecinski et al. (29), the number of packed red blood cell units transfused, length of hospital stay and minimum estimated glomerular filtration rate provided the greatest contribution to the AI model, with specificities of 94 (92–97)%, 33 (29–37)% and 53 (49–58)%, respectively (29). No other studies reported the specificity values for the individual clinical variables used in their predictive models. As an illustration of the multiplicity of variables that AI models can evaluate, the main

30-day mortality predictive variables reported in the study by Lertsanguansinchai et al. (31) were height, chronic lung disease, STS score, preoperative left ventricular ejection fraction (LVEF), age, and preoperative left ventricular outflow tract velocity time integral (VOT VTI), while the main 1-year mortality variables were preoperative LVEF, STS score, hear rate, systolic blood pressure, home oxygen use, serum creatinine level, and preoperative LVOT Vmax. A more detailed summary of the variables used across the included studies can be found in Supplementary Table S1.

Researchers often compare the AI-generated death predictions with the actual mortality results seen in real-world TAVI patients to assess model accuracy. This evaluation aids in determining the degree to which the AI models are capable of making trustworthy and precise forecasts of post-operative mortality. Here, we provide a brief summary of the advantages and disadvantages associated with each AI algorithm featured in our study.

Firstly, the RF model is capable of analysing high-dimensional data with a huge number of diverse predictors, which can even exceed the number of observations (39).While a single decision tree is prone to noise and overfitting when grown on its training

set, the RF model improves accuracy by averaging multiple decision trees. However, due to their complexity, RF models tend to have low intrinsic interpretability (40). Additionally, the presence of dependent observations in data may contribute to increased bias and inaccurate predictive variable selection (41). While the GB algorithm minimises errors and maximises predictive accuracy in the final model, GB algorithms, like the RF models, may also suffer from low interpretability. The GB model may also be prone to over-fitting if the additive process of gradient boosting is not regularised. ANN and MLP algorithms are suitable for analysing non-linear relationships between dependent and independent variables. However, these models are difficult to apply to real-time predictions and are also prone to overfitting (42). Finally, LR algorithms can only be used to predict discrete functions and cannot predict continuous outcomes. It also assumes a linear relationship between dependent and independent variables.

Overall, AI-based prediction models are able to take into account a myriad of patient data to generate more accurate predictions on post-TAVI mortality than traditional scores. However, such models require a large number of patient variables to generate predictions, some of which may not be readily available to clinicians in the immediate clinical setting. Hence, at present, AI-based prediction models may be less user-friendly than simple traditional risk scores. In future, further research into simpler AI-based models that are able to use easily-available clinical parameters in predictions is needed to increase the clinical utility of these models.

Firstly, the RF model constructs a series of decision trees and utilises bagging and randomisation of predictors in order to accurately predict outcomes (24). The original dataset is first divided into smaller subsets through random feature selection. Individual decision trees are then grown on the subsets, allowing for the construction of many decision trees with low correlation to one another. Finally, multiple decision trees are averaged, thereby minimising variance and improving the accuracy of the final predictive model (24). The RF model is thus capable of analysing high-dimensional data with a huge number of diverse predictors, which can even exceed the number of observations (39).

While a single decision tree is prone to noise and overfitting when grown on its training set, the RF model improves accuracy by averaging multiple decision trees. However, due to their complexity, RF models tend to have low intrinsic interpretability (40). Additionally, the presence of dependent observations in data may contribute to increased bias and inaccurate predictive variable selection (41).

The GB model involves an ensemble of weak learner decision trees. Each weak learner progresses through a stepwise progression, where previous iterations of weak learners are combined into a successive strong learner, correcting for the errors of the preceding learner each time (25, 26). Eventually, this process is repeated until errors are minimised and predictive accuracy maximised in the final model. However, similar to the RF model, GB algorithms may also suffer from low interpretability. The GB model may also be prone to over-fitting if the additive process of gradient boosting is not regularised.

The ANN algorithm is a mathematical model developed based on the transmission of signals among neural networks in biological nervous systems. An input signal is channelled through a collection of nodes, which are artificial neurons (27). Each node analyses the input signal and then transmits an output to each of its connected neurons, mimicking the transmission of action potentials across neurons and synapses in the human brain. The nodes are also organised into layers. The input signal travels from the first input layer to the last output layer, undergoing different transformations at each layer.

The MLP algorithm is a class of feedforward ANN that is comprised of an input layer, one or more hidden layers and an output layer. The input signal first passes sequentially through the layers in a feedforward process until an output is generated. Subsequently, the data is then backpropagated from the output layer back into the hidden layers. The error signal of each node that contributed to the overall error is determined and the weights of the network updated until the gradient of the mean squared error converges. While ANN and MLP are able to analyse non-linear relationships between dependent and independent variables, these models are difficult to apply to real-time predictions and are also prone to overfitting (42).

The LR algorithm fits a logistic function onto a dataset and predicts the probability of an independent variable, such as mortality, from a dependent variable. The maximimum likelihood estimation method is most commonly used to maximise the likelihood function and find the optimal fit of the model (28). LR algorithms can only be used to predict discrete functions and cannot predict continuous outcomes. It also assumes a linear relationship between dependent and independent variables.

Overall, AI-based prediction models are able to take into account a myriad of patient data to generate more accurate predictions on post-TAVI mortality than traditional scores. However, such models require a large number of patient variables to generate predictions, some of which may not be readily available to clinicians in the immediate clinical setting. Hence, at present, AI-based prediction models may be less user-friendly than simple traditional risk scores. In future, further research into simpler AI-based models that are able to use easily-available clinical parameters in predictions is needed to increase the clinical utility of these models.

Limitations

The main limitations of this study were firstly, the high heterogeneity of the data, specifically in comparing the intra-hospital mortality subgroup, 1-year mortality with gradient boosting subgroup and the overall pooled AUC of all 10 studies. We postulated that this was due to significant differences in the training datasets used for each AI model. Fundamentally, there was large variability in the quantity of data, as well as in the type and number of parameters used to predict mortality in each model. In addition, due to the lack of studies comparing the same AI algorithms and the same control group, it was difficult to perform a more homogenous subgroup

analysis on each AI algorithm. Hence, our results may not be generalisable to all AI algorithms or all datasets predicting mortality in patients post-TAVI.

Secondly, all the studies included in our analysis were of a retrospective nature. There was also a lack of data comparing AI model predictive performance to traditional scores. Therefore, further research, including randomized controlled trials, comparing the use of AI algorithms with traditional scores in predicting post-TAVI mortality will be needed in the future in order to determine the validity of our observations.

Finally, our study did not analyse other endpoints that are known post-TAVI complications, such as stroke, pacemaker need and heart failure, and focused solely on mortality. Hence, this may limit the applicability of our findings in clinical practice.

Conclusion

Personalized risk assessments are the ultimate goal of research into AI systems for predicting patient outcomes in TAVI procedures. The potential of AI-generated mortality forecasts to improve the precision and value of risk assessment in TAVI is highlighted in this systematic review. AI can help healthcare providers predict and monitor patient outcomes, which will hopefully result in better decision-making and more desirable post-TAVI outcomes in future.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. Derived data supporting the findings of this study are available from the corresponding author on request.

Author contributions

FS: Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. AL: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. MF: Conceptualization, Formal Analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. LT: Data curation, Formal Analysis, Writing – review & editing. HL: Conceptualization, Validation, Writing – original draft, Writing – review & editing. IK: Conceptualization, Resources, Validation, Writing – original draft, Writing – review & editing. ET: Conceptualization, Validation, Writing – original draft, Writing – review & editing. TK: Conceptualization, Funding acquisition, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2024.1343210/full#supplementary-material

References

1. Investigators TUTT. Effect of transcatheter aortic valve implantation vs surgical aortic valve replacement on all-cause mortality in patients with aortic stenosis: a randomized clinical trial. *JAMA*. (2022) 327:1875–87. doi: 10.1001/jama.2022.5776

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(18):1695–705. doi: 10.1056/NEJMoa1814052

3. Conte JV, Hermiller J Jr, Resar JR, Deeb GM, Gleason TG, Adams DH, et al. Complications after self-expanding transcatheter or surgical aortic valve replacement. *Semin Thorac Cardiovasc Surg*. (2017) 29(3):321–30. doi: 10.1053/j.semtcvs.2017.06.001

4. Rodés-Cabau J, Abbas AE, Serra V, Vilalta V, Nombela-Franco L, Regueiro A, et al. Balloon- vs self-expanding valve systems for failed small surgical aortic valve bioprostheses. *J Am Coll Cardiol*. (2022) 80(7):681–93. doi: 10.1016/j.jacc.2022.05.005. Erratum in: J Am Coll Cardiol. (2022) 80(14):1419. PMID: 35597385

5. Färber G, Bleiziffer S, Doenst T, Bon D, Böning A, Weiler H, et al. Transcatheter or surgical aortic valve implantation in chronic dialysis patients: a German aortic valve registry analysis. *Clin Res Cardiol*. (2021) 110(3):357–67. doi: 10.1007/s00392-020-01717-7

6. Van Belle E, Vincent F, Labreuche J, Auffret V, Debry N, Lefèvre T, et al. Balloon-expandable versus self-expanding transcatheter aortic valve replacement: a propensity-

matched comparison from the FRANCE-TAVI registry. *Circulation*. (2020) 141 (4):243–59. doi: 10.1161/CIRCULATIONAHA.119.043785

7. Hoogma DF, Venmans E, Al Tmimi L, Tournoy J, Verbrugghe P, Jacobs S, et al. Postoperative delirium and quality of life after transcatheter and surgical aortic valve replacement: a prospective observational study. *J Thorac Cardiovasc Surg*. (2023) 166 (1):156–166.e6. doi: 10.1016/j.jtcvs.2021.11.023

8. Magro PL, Sousa-Uva M. In low-risk patients aged >70–75 with severe aortic stenosis, is transcatheter superior to surgical aortic valve replacement in terms of reported cardiovascular composite outcomes and survival? *Interact Cardiovasc Thorac Surg*. (2021) 34:40–4. doi: 10.1093/icvts/ivab218

9. Pibarot P, Ternacle J, Jaber WA, Salaun E, Dahou A, Asch FM, et al. Structural deterioration of transcatheter versus surgical aortic valve bioprostheses in the PARTNER-2 trial. *J Am Coll Cardiol*. (2020) 76(16):1830–43. doi: 10.1016/j.jacc.2020.08.049

10. Webb JG, Mack MJ, White JM, Dvir D, Blanke P, Herrmann HC, et al. Transcatheter aortic valve implantation within degenerated aortic surgical bioprostheses: pARTNER 2 valve-in-valve registry. *J Am Coll Cardiol*. (2017) 69 (18):2253–62. doi: 10.1016/j.jacc.2017.02.057

11. Reardon MJ, Van Mieghem NM, Popma JJ, Kleiman NS, Søndergaard L, Mumtaz M, et al. Surgical or transcatheter aortic-valve replacement in intermediate-risk patients. *N Engl J Med*. (2017) 376(14):1321–31. doi: 10.1056/NEJMoa1700456

12. Gleason TG, Reardon MJ, Popma JJ, Deeb GM, Yakubov SJ, Lee JS, et al. 5-year outcomes of self-expanding transcatheter versus surgical aortic valve replacement in high-risk patients. *J Am Coll Cardiol*. (2018) 72(22):2687–96. doi: 10.1016/j.jacc.2018.08.2146

13. Ichibori Y, Mizote I, Tsuda M, Mukai T, Maeda K, Onishi T, et al. Long-term outcomes of high-risk or inoperable patients who underwent transcatheter aortic valve implantation. *Am J Cardiol*. (2019) 124(4):573–9. doi: 10.1016/j.amjcard.2019.05.025

14. Jørgensen TH, Thyregod HGH, Ihlemann N, Nissen H, Petursson P, Kjeldsen BJ, et al. Eight-year outcomes for patients with aortic valve stenosis at low surgical risk randomized to transcatheter vs. surgical aortic valve replacement. *Eur Heart J*. (2021) 42(30):2912–9. doi: 10.1093/eurheartj/ehab375

15. Pibarot P, Salaun E, Dahou A, Avenatti E, Guzzetti E, Annabi MS, et al. Echocardiographic results of transcatheter versus surgical aortic valve replacement in low-risk patients: the PARTNER 3 trial. *Circulation*. (2020) 141(19):1527–37. doi: 10.1161/CIRCULATIONAHA.119.044574

16. Forrest JK, Deeb GM, Yakubov SJ, Gada H, Mumtaz MA, Ramlawi B, et al. 4-Year Outcomes of patients with aortic stenosis in the evolut low risk trial. *J Am Coll Cardiol*. (2023) 82(22):2163–5. doi: 10.1016/j.jacc.2023.09.813

17. Çelik M, Milojevic MM, Durko AP, Oei FBS, Bogers A, Mahtab EAF. Mortality in low-risk patients with aortic stenosis undergoing transcatheter or surgical aortic valve replacement: a reconstructed individual patient data meta-analysis. *Interact Cardiovasc Thorac Surg*. (2020) 31:587–94. doi: 10.1093/icvts/ivaa179

18. Thoenes M, Agarwal A, Grundmann D, Ferrero C, McDonald A, Bramlage P, et al. Narrative review of the role of artificial intelligence to improve aortic valve disease management. *J Thorac Dis*. (2021) 13(1):396–404. doi: 10.21037/jtd-20-1837

19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br Med J*. (2021) 372:n71. doi: 10.1136/bmj.n71

20. Schünemann HJ, Higgins JP, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Completing "summary of findings" tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, et al., editors. *Cochrane Handbook for Systematic Reviews*. London, UK: John Wiley & Sons, Inc (2019). p. 375–402.

21. Sterne JA, Hernán MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Br Med J*. (2016) 355:i4919. doi: 10.1136/bmj.i4919

22. Collaboration C. *Review Manager (RevMan) Version 5.3*. Copenhagen: The Nordic Cochrane Centre (2014).

23. StataCorp L. *Stata Statistical Software: Release 17*. College Station, Texas, United States of America: A Stata Press Publication, StataCorp LLC (2021).

24. Rigatti SJ. Random forest. *J Insur Med*. (2017) 47:31–9. doi: 10.17849/insm-47-01-31-39.1

25. 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(14):2216–25. doi: 10.1093/cvr/cvz321

26. Kwiecinski J, Tzolos E, Meah MN, Cadet S, Adamson PD, Grodecki K, et al. Machine learning with ^{18}F-sodium fluoride PET and quantitative plaque analysis on CT angiography for the future risk of myocardial infarction. *J Nucl Med*. (2022) 63 (1):158–65. doi: 10.2967/jnumed.121.262283

27. Kriegeskorte N, Golan T. Neural network models and deep learning. *Curr Biol*. (2019) 29:R231–6. doi: 10.1016/j.cub.2019.02.034

28. Nistal-Nuño B. Artificial intelligence forecasting mortality at an intensive care unit and comparison to a logistic regression system. *Einstein (Sao Paulo)*. (2021) 19: eAO6283. doi: 10.31744/einstein_journal/2021AO6283

29. 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. doi: 10.1016/j.carrev.2020.08.010

30. Gomes B, Pilz M, Reich C, Leuschner F, Konstandin M, Katus HA, et al. Machine learning-based risk prediction of intrahospital clinical outcomes in patients undergoing TAVI. *Clin Res Cardiol*. (2021) 110(3):343–56. doi: 10.1007/s00392-020-01691-0

31. Hernandez-Suarez DF, Kim Y, Villablanca P, Gupta T, Wiley J, Nieves-Rodriguez BG, et al. Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. *JACC: Cardiovasc Interv*. (2019) 12 (14):1328–38. doi: 10.1016/j.jcin.2019.06.013

32. Kwiecinski J, Dabrowski M, Nombela-Franco L, Grodecki K, Pieszko K, Chmielak Z, et al. Machine learning for prediction of all-cause mortality after transcatheter aortic valve implantation. *Eur Heart J Quality Care Clin Outcomes*. (2023) 9(8):768–77. doi: 10.1093/ehjqcco/qcad002

33. Leha A, Huber C, Friede T, Bauer T, Beckmann A, Bekeredjian R, et al. Development and validation of explainable machine learning models for risk of mortality in transcatheter aortic valve implantation: tAVI risk machine scores. *Eur Heart J Dig Health*. (2023) 4(3):225–35. doi: 10.1093/ehjdh/ztad021

34. Lertsanguansinchai P, Chokesuwattanaskul R, Petchlorian A, Suttirut P, Buddhari W. Machine learning-based predictive risk models for 30-day and 1-year mortality in severe aortic stenosis patients undergoing transcatheter aortic valve implantation. *Int J Cardiol*. (2023) 374:20–6. doi: 10.1016/j.ijcard.2022.12.023

35. Lopes RR, Mamprin M, Zelis JM, Tonino PAL, van Mourik MS, Vis MM, et al. Local and distributed machine learning for inter-hospital data utilization: an application for TAVI outcome prediction. *Front Cardiovasc Med*. (2021) 8:787246. doi: 10.3389/fcvm.2021.787246

36. Mamprin M, Lopes RR, Zelis JM, Tonino PAL, van Mourik MS, Vis MM, et al. Machine learning for predicting mortality in transcatheter aortic valve implantation: an inter-center cross validation study. *J Cardiovasc Dev Dis*. (2021) 8(6):65. doi: 10.3390/jcdd8060065

37. Mamprin M, Zelis JM, Tonino PAL, Zinger S, de With PHN. Decision trees for predicting mortality in transcatheter aortic valve implantation. *Bioengineering*. (2021) 8:22. doi: 10.3390/bioengineering8020022

38. 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(4):44. doi: 10.3390/jcdd8040044

39. Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics*. (2012) 99:323–9. doi: 10.1016/j.ygeno.2012.04.003

40. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinformatics*. (2023) 24(2):bbad002. doi: 10.1093/bib/bbad002

41. Fokkema M, Smits N, Zeileis A, Hothorn T, Kelderman H. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behav Res Methods*. (2018) 50:2016–34. doi: 10.3758/s13428-017-0971-x

42. Salgado CM, Dam RSF, Puertas EJA, Salgado WL. Calculation of volume fractions regardless scale deposition in the oil industry pipelines using feed-forward multilayer perceptron artificial neural network and MCNP6 code. *Appl Radiat Isot*. (2022) 185:110215. doi: 10.1016/j.apradiso.2022.110215