BMC Cardiovascular Disorders

RESEARCH

Open Access

Evaluation of machine learning methods for prediction of heart failure mortality and readmission: meta-analysis

Hamed Hajishah[1], Danial Kazemi[2], Ehsan Safaee[3], Mohammad Javad Amini[4], Maral Peisepar[5,6], Mohammad Mahdi Tanhapour[1] and Arian Tavasol[7,8*]

Abstract

Background Heart failure (HF) impacts nearly 6 million individuals in the U.S., with a projected 46% increase by 2030, is creating significant healthcare burdens. Predictive models, particularly machine learning (ML)-based models, offer promising solutions to identify patients at greater risk of adverse outcomes, such as mortality and hospital readmission. This review aims to assess the effectiveness of ML models in predicting HF-related outcomes, with a focus on their potential to improve patient care and clinical decision-making. We aim to assess how effectively machine learning models predict mortality and readmission in heart failure patients to improve clinical outcomes.

Method The study followed PRISMA 2020 guidelines and was registered in the PROSPERO database (CRD42023481167). We conducted a systematic search in PubMed, Scopus, and Web of Science databases using specific keywords related to heart failure, machine learning, mortality and readmission. Extracted data focused on study characteristics, machine learning details, and outcomes, with AUC or c-index used as the primary outcomes for pooling analysis. The PROBAST tool was used to assess bias risk, evaluating models based on participants, predictors, outcomes, and statistical analysis. The meta-analysis pooled AUCs for different machine learning models predicting mortality and readmission. Prediction accuracy data was categorized by timeframes, with high heterogeneity determined by an I^2 value above 50%, leading to a random-effects model when applicable. Publication bias was assessed using Egger's and Begg's tests, with a p-value below 0.05 considered significant

Result A total of 4,505 studies were identified, and after screening, 64 were included in the final analysis, covering 943,941 patients. Of these, 40 studies focused on mortality, 17 on readmission, and 7 on both outcomes. In total, 346 machine learning models were evaluated, with the most common algorithms being random forest, logistic regression, and gradient boosting. The neural network model achieved the highest overall AUC for mortality prediction (0.808), while the support vector machine performed best for readmission prediction (AUC 0.733). The analysis revealed a significant risk of bias, primarily due to reliance on retrospective data and inadequate sample size justification.

*Correspondence:
Arian Tavasol
ariantavasol1378@gmail.com

Full list of author information is available at the end of the article

Conclusion In conclusion, this review emphasizes the strong potential of ML models in predicting HF readmission and mortality. ML algorithms show promise in improving prognostic accuracy and enabling personalized patient care. However, challenges like model interpretability, generalizability, and clinical integration persist. Overcoming these requires refined ML techniques and a robust regulatory framework to enhance HF outcomes.

Keywords Machine learning, Heart failure, Mortality, Readmission, Meta-analysis

Introduction

Heart failure (HF) is a complex condition characterized by the heart's inability to adequately pump blood and oxygen to sustain other organs, as defined by the US Centers for Disease Control and Prevention which is highly prevalent in the United States [1], affecting nearly 6 million Americans aged elder than 20 years [2]. By the year 2030, it is anticipated that the prevalence of this condition will surge by 46%, surpassing the 8 million marks [3]. This rising prevalence places a substantial burden on healthcare systems and underscores the need for improved management strategies.

One of the most critical challenges in HF care is the high rate of readmissions and mortality.

Elevated rates of readmission and mortality often signal inadequate care during the initial hospitalization and discharge planning process leading to detrimental effects on patient health and well-being [4]. Thus, forecasting the mortality of heart failure patients holds significant importance for various reasons, such as enhancing the quality of care, mitigating healthcare expenses, optimizing resource distribution, and ultimately improving patient outcomes [5].

Predictive models have the potential to transform HF care by helping clinicians identify patients at the highest risk of adverse outcomes, such as death or readmission. Despite their potential benefits, predictive models for readmission and mortality after heart failure (HF) hospitalization often exhibit unsatisfactory performance compared to models predicting mortality. This performance gap underscores the need for further research and refinement to improve the accuracy and reliability of readmission prediction models in HF care settings [6].

In recent years, machine learning (ML) and artificial intelligence (AI) have gained prominence in healthcare due to their ability to analyze complex data and detect patterns that may not be apparent with traditional methods.

ML models, in particular, have demonstrated superior performance in predicting HF-related outcomes, with 76% of studies reporting better results compared to conventional statistical models. Leveraging ML to develop more accurate predictive models could significantly enhance the care of HF patients by providing clinicians with powerful tools to anticipate risks and make more informed treatment decisions [7, 8].

In this systematic review and meta-analysis, we aim to evaluate the effectiveness of ML-based predictive models for forecasting mortality and readmission in HF patients, providing insights into their potential to improve clinical outcomes.

Method

This study was conducted according to the criteria of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) [9]. To ensure transparency and adherence to planned methods, the study protocol was prospectively registered with PROSPERO, an international database for systematic reviews (CRD42023481167).

Search strategy

PubMed, Scopus, and Web of Science were searched systemically up to November 11, 2023. To identify all pertinent studies, we employed a combination of standardized medical terminology (MeSH headings) and normal search terms. The tree main keywords with their synonyms and similarities was as follows:1. "Heart Failure" OR "congestive heart failure" OR "cardiac failure" OR "heart decompensation" OR "myocardial failure" 2. "Machine Learning" OR "artificial intelligence" OR "artificial neural network" OR "deep learning" OR "prediction model" 3. "Mortality" OR "death" OR "fatality" OR "survival" OR "admission" OR "readmission" OR "Rehospitalization". The detailed search strategy for each database is available in additional file.

The inclusion criteria for this study were as follows: studies investigating the apply of machine learning models on heart failure patients, studies investigating on mortality or readmission or both, and studies written in English.

We excluded studies conducted in a non-English language or in the case of full-text unavailability. Also, all the other types of study except original studies with a detailed machine learning evaluation were excluded.

Data extraction

Four reviewers separated into two groups, independently extracted data from the included studies into a pre-defined form, and any dissimilarity was resolved by the first author. The Extracted data can be divided into three main categories: (1) Study characteristics (Author, year, country, age, sex, population, heart failure type) (2)

Machine learning details (model type, predictive variables, test-train split, evaluation metric) (3) Outcomes. Because the most common type of reported outcome was AUC or c-index, we decided to consider this variable as our main outcomes for the pooling analysis. However, sensitivity, specificity, and accuracy were also analyzed. When a study explores multiple models, or even in different time-framing, we extracted the data pertaining to each model and timeframe individually. In case of reporting different number of predictive variables, we just considered the one included in the final operated model.

Quality assessment

To evaluate the risk of bias of the included models, PROBAST assessment tool was utilized [10]. The defined questionnaire contains 20 questions categorized into four domains: participants, predictors, outcomes, and statistical analysis. On the basis of pre-defined answers ("yes", "probably yes", "no", "may or may not", and "no information), the models were evaluated as low risk, high risk, or unclear risk of bias. The same four reviewers which did the data extraction were also responsible, and any conflicts was concluded by a third party.

Data analysis

The meta-analysis was conducted using STATA 17 to pool the AUCs of various machine learning models designed to predict the mortality and readmission of HF patients. It has to be mentioned that the data of different deep learning models was pooled due to the low number of studies employing each specific deep learning model.

Prediction accuracy data for each model were reported over different time periods in some studies. To ensure comparability, we selected and pooled data with similar prediction periods. For mortality prediction accuracies, the data were categorized as follows: "under 1-year", "1-year", "more than 1-year", and "overall". In the "overall" category, one prediction data point per model from each study was selected and pooled, preferably data around the 1-year mark. The categories for readmission prediction data were similar to those for mortality, with the distinction that the "under 1-year" and "1-year" categories were merged.

Finally, the mean and standard error (SE) of the AUCs of the models from the included studies were extracted and pooled. Also, other evaluation metrics including sensitivity, specificity, and accuracy were pooled. Heterogeneity in the meta-analyses was assessed using the I^2 statistic. An I^2 value greater than 50% indicated high heterogeneity, in which case a random-effects model was employed for the analysis. Publication bias was evaluated using Egger's regression test and Begg's test. A p-value less than 0.05 was considered statistically significant.

Result

Screening results

In total, 4505 articles were gathered through a systematic search in three databases. By removing 1226 duplicate papers, a total of 3279 papers were included for the first screen. Following that, by screening the titles and abstracts, 180 papers were chosen to be reviewed through full texts. Finally, 64 papers were qualified to be included in the study. The screening process is illustrated in Fig. 1.

Characteristics of studies

Out of the 64 studies reviewed, mortality was the primary outcome investigated in 40 papers [11–50], while readmission was the sole focus of 17 papers [51–67]. Seven studies [68–74] examined both mortality and readmission. This study examined a total of 943,941 patients. Among them, the mortality group accounted for 741,050 cases (78.51%), while the readmission group comprised 202,891 patients (21.49%). Except for six studies [14, 30, 57, 58, 60, 61], all the other papers were published after 2020. In terms of the origins of publications, the USA had 22 papers, and China had 18 papers. Entirely, 346 machine learning models were developed, including 225 and 121 models for mortality and readmission, respectively. The algorithms by order of prevalence were as follows: random forest ($n=74$), logistic regression ($n=69$), gradient boosting ($n=62$), support vector machine ($n=43$), neural network ($n=32$), decision tree ($n=16$), lasso regression ($n=15$), KNN ($n=10$), Bayesian network ($n=7$), and others ($n=18$). Samples of studies were acquired from a total of four origins: public databases ($n=22$), national data registries ($n=18$), clinical trials ($n=13$), and electronic health records($n=11$). The table presenting key characteristics of studies is available in additional files.

Quality assessments

PRPBAST was used to evaluate the risk of bias in all models. Many machine learning models in this analysis were susceptible to bias. This was primarily caused by two factors: (1) a heavy reliance on studies with retrospective data, where participants were enrolled after the event of interest had occurred, and (2) a lack of sample size justification. In Fig. 2, we illustrated risk assessments divided into four domains.

Meta-analysis of prediction models

AUC for the prediction of mortality – overall

The results of our meta-analyses revealed that the "Neural network" model, based on 13 included studies, had the best AUC for the prediction of mortality in the HF patients with the pooled AUC of 0.808 [95% CI: 0.754–0.863]. Also, the "Gradient boosting" model, based on

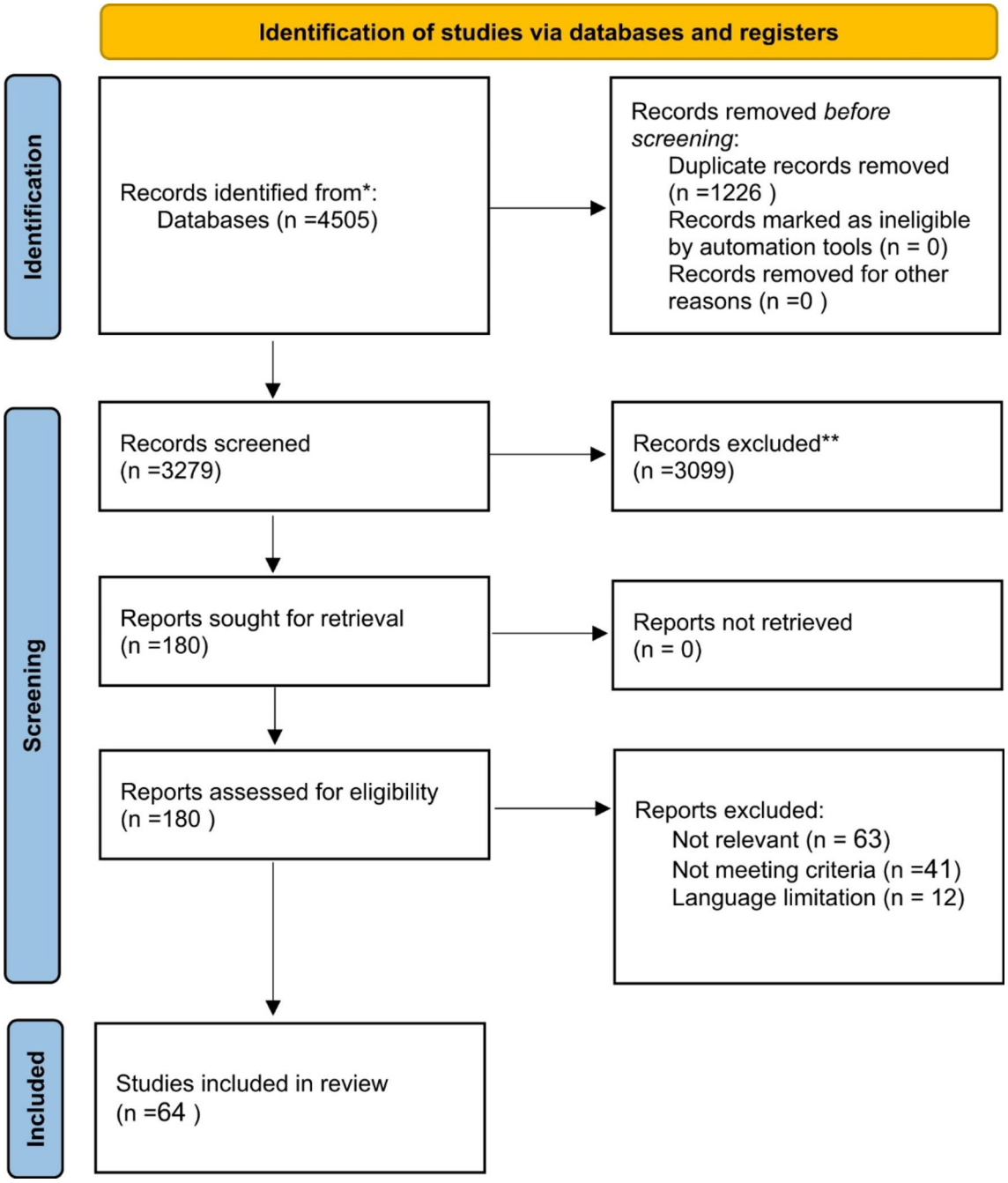

Fig. 1 Prisma flowchart illustrating the process of screening

30 included studies, had a pooled AUC of 0.796 [95% CI: 0.746–0.845], which was approximately similar to the neural network model. The lowest pooled AUC belonged to the "K-nearest neighbors" model, based on 5 included studies, with a pooled AUC of 0.571 [95% CI: 0.532–0.611]. The results of the pooled AUCs of all of the models are reported in Table 1; Figs. 3 and 4.

AUC for the prediction of mortality – under 1-year

The results of our meta-analyses showed that the "Neural network" model, based on 4 included studies, had the best AUC for the prediction of mortality in the HF patients with the pooled AUC of 0.829 [95% CI: 0.788–0.871]. Also, the "Gradient boosting" model, based on 10 included studies, had a pooled AUC of 0.817 [95% CI: 0.792–0.843], which was approximately similar to the "Neural network" model. The lowest pooled AUC belonged to the "Decision tree" model, based on 5

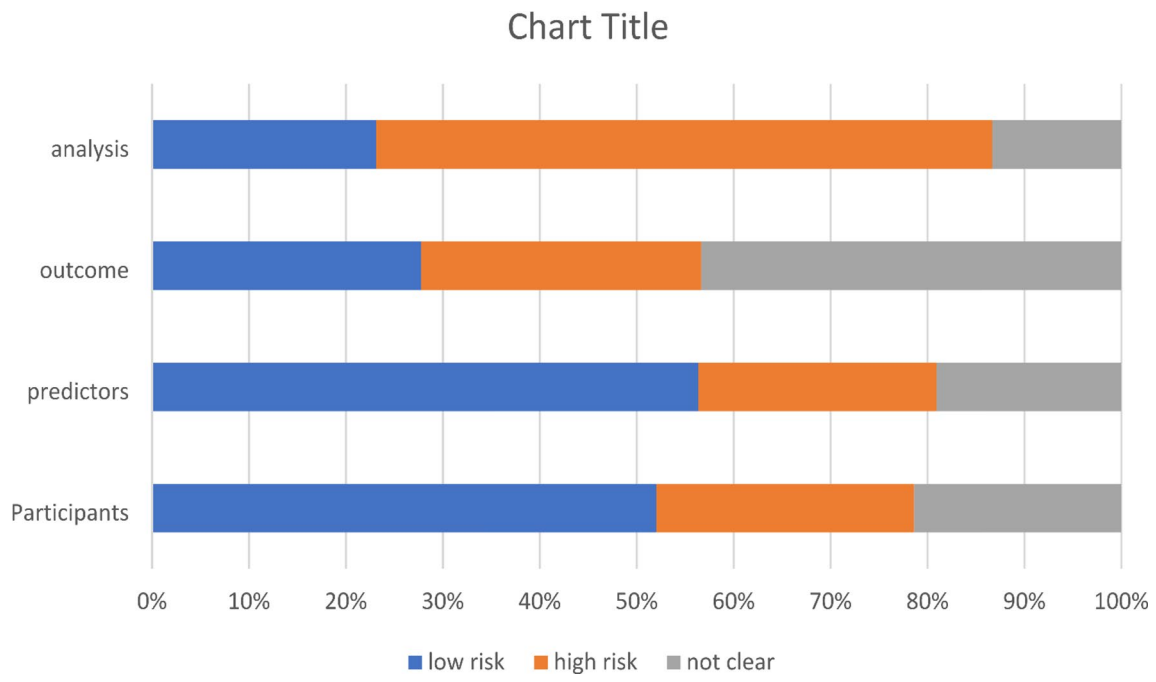

Fig. 2 Percentage staked chart regarding risk of bias assessment

included studies, with a pooled AUC of 0.681 [95% CI: 0.583–0.778]. The results of the pooled AUCs of all of the models are reported in Table 1.

AUC for the prediction of mortality – 1-year

The results of our meta-analyses showed that the "Gradient boosting" model, based on 5 included studies, had the best AUC for the prediction of mortality in the HF patients with the pooled AUC of 0.820 [95% CI: 0.765–0.875]. The lowest pooled AUC belonged to the "Support vector machine" model, based on 5 included studies, with a pooled AUC of 0.689 [95% CI: 0.583–0.795]. The results of the pooled AUCs of all of the models are reported in Table 1.

AUC for the prediction of mortality – more than 1-year

The results of our meta-analyses showed that the "Neural network" model, based on 2 included studies, had the best AUC for the prediction of mortality in the HF patients with the pooled AUC of 0.798 [95% CI: 0.767–0.828]. Also, Among the models with more than 3 included studies, the "Logistic regression" and "Random Forest" models, based on 5 included studies, had a pooled AUC of 0.687 [95% CI: 0.623–0.751] and 0.682 [95% CI: 0.630–0.735], respectively, which were lower than the "Neural network" model. The lowest pooled AUC belonged to the "Support vector machine" model, based on 5 included studies, with a pooled AUC of 0.642 [95% CI: 0.583–0.778]. The results of the pooled AUCs of all of the models are reported in Table 1.

AUC for the prediction of mortality – not specific

We pooled the data of studies that did not report specific periods for mortality. The results of our meta-analyses showed that the "Neural Network" model, based on 6 included studies, had the best AUC for the prediction of mortality in the HF patients with the pooled AUC of 0.851 [95% CI: 0.796–0.905]. The lowest pooled AUC belonged to the "Support vector machine" model, based on 9 included studies, with a pooled AUC of 0.718 [95% CI: 0.578–0.857]. The results of the pooled AUCs of all of the models are reported in Table 1.

Other metrics for the prediction of mortality – overall

The performance of various machine learning models to predict mortality was also evaluated based on accuracy, sensitivity, and specificity (Table 2).

For accuracy, the Decision Tree ($N=2$) achieved the highest performance with an AUC of 0.810 (95% CI: 0.504–1.117), followed by Support Vector Machine (SVM) ($N=8$) with an AUC of 0.743 (95% CI: 0.694–0.791) and Gradient Boosting ($N=12$) with an AUC of 0.706 (95% CI: 0.467–0.945). Random Forest ($N=15$) and Logistic Regression ($N=10$) showed moderate accuracy, with AUC values of 0.600 (95% CI: 0.415–0.786) and 0.576 (95% CI: 0.366–0.786), respectively. K-Nearest Neighbors (KNN) ($N=2$) had an AUC of 0.601 (95% CI: 0.576–0.627), while Neural Network ($N=4$) exhibited the lowest accuracy with an AUC of 0.441 (95% CI: 0.061–0.820). Similarly, Lasso Regression ($N=4$) performed poorly with an AUC of 0.413 (95% CI: 0.079–0.747). Bayesian Network ($N=1$) did not provide sufficient data

Table 1 Pooled AUCs of mortality prediction algorithms

Model	Under 1-year mortality			1-year mortality			More than 1-year mortality			Not specified timing			Overall		
	N	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval
Random forest	9	0.786	0.738–0.834	10	0.740	0.671–0.808	5	0.682	0.630–0.735	19	0.794	0.689–0.899	32	0.763	0.691.0.834
Logistic regression	9	0.780	0.720–0.840	9	0.746	0.693-0.800	5	0.687	0.623–0.751	17	0.755	0.680–0.829	31	0.740	0.706–0.775
Support vector machine	6	0.738	0.700-0.777	5	0.689	0.583–0.795	5	0.642	0.582–0.702	9	0.727	0.582–0.871	19	0.70	0.589–0.801
Gradient boosting	10	0.821	0.789–0.853	5	0.820	0.765–0.875	1	NA	NA	19	0.818	0.724–0.911	30	0.796	0.746–0.845
Neural Network	4	0.829	0.788–0.871	6	0.774	0.744–0.803	2	0.798	0.767–0.828	6	0.851	0.789–0.912	13	0.808	0.754–0.863
Decision tree	5	0.681	0.583–0.778	1	NA	NA	1	NA	NA	5	0.844	0.768–0.921	9	0.708	0.623–0.794
K-nearest neighbors	2	0.717	0.456–0.978	1	NA	NA	1	NA	NA	2	NA	NA	5	0.571	0.532–0.611
Lasso regression	2	0.746	0.734–0.758	2	0.766	0.750–0.801	1	NA	NA	4	0.750	0.678–0.821	8	0.756	0.722–0.791
Bayesian network	2	0.810	0.653–0.967	1	NA	NA	1	NA	NA	0	NA	NA	4	0.654	0.605–0.703

for evaluation. The forest plots of some models are presented in Figs. 5 and 6.

For sensitivity, Lasso Regression ($N=2$) had the highest AUC of 0.776 (95% CI: 0.564–0.988), followed by Gradient Boosting ($N=12$) with an AUC of 0.678 (95% CI: 0.527–0.829). Random Forest ($N=12$) and Neural Network ($N=3$) demonstrated moderate sensitivity, with AUC values of 0.619 (95% CI: 0.547–0.691) and 0.623 (95% CI: 0.503–0.743), respectively. Logistic Regression ($N=9$) and Support Vector Machine (SVM) ($N=7$) showed comparable sensitivity, with AUC values of 0.592 (95% CI: 0.508–0.676) and 0.590 (95% CI: 0.492–0.687), respectively. Decision Tree ($N=3$) had an AUC of 0.609 (95% CI: 0.224–0.994), whereas K-Nearest Neighbors (KNN) ($N=2$) had the lowest sensitivity with an AUC of 0.339 (95% CI: 0.314–0.364). Bayesian Network ($N=0$) did not provide data for sensitivity evaluation. The forest plots of some models are presented in Fig. 7.

For specificity, Neural Network ($N=3$) achieved the highest AUC of 0.774 (95% CI: 0.657–0.891), followed by Support Vector Machine (SVM) ($N=5$) with an AUC of 0.761 (95% CI: 0.572–0.950) and Gradient Boosting ($N=8$) with an AUC of 0.757 (95% CI: 0.664–0.850). Random Forest ($N=8$) and Logistic Regression ($N=7$) performed similarly, with AUC values of 0.755 (95% CI: 0.704–0.806) and 0.745 (95% CI: 0.685–0.805), respectively. Lasso Regression ($N=2$) demonstrated moderate specificity with an AUC of 0.701 (95% CI: 0.610–0.793), while K-Nearest Neighbors (KNN) ($N=2$) had a lower specificity of 0.685 (95% CI: 0.592–0.778). Decision Tree ($N=1$) did not provide sufficient specificity data. Bayesian Network ($N=0$) also lacked specificity data for evaluation. The forest plots of some models are presented in Fig. 8.

AUC for the prediction of the readmission – overall

The results of our meta-analyses revealed that the "Support vector machine" model, based on 10 included studies, had the best AUC for the prediction of readmission in the HF patients with the pooled AUC of 0.726 [95% CI: 0.639–0.813]. Also, the "Gradient boosting" model, based on 18 included studies, had a pooled AUC of 0.703 [95% CI: 0.649–0.758], which was less accurate than the "Support vector machine" model but included more studies. The lowest pooled AUC belonged to the "Decision tree" model, based on 3 included studies, with a pooled AUC of 0.618 [95% CI: 0.558–0.679]. The results of the pooled AUCs of all of the models are reported in Table 3; Figs. 9 and 10.

AUC for the prediction of the readmission – under 1-year

The results of our meta-analyses revealed that the "Support vector machine" model, based on 3 included studies, had the best AUC for the prediction of readmission

Fig. 3 Forest plots reporting the results of the meta-analysis of the accuracy of machine learning models for predicting mortality in heart failure (HF) patients. (1) It represents the pooled area under the curve (AUC) values for the Random Forest model across multiple studies, (2) It shows the pooled AUC for the Logistic Regression model, and (3) It presents the pooled AUC for the Support Vector Machine model. (4) It shows the pooled accuracy values for the Gradient Boosting model. Each plot includes individual study estimates with corresponding confidence intervals and study weights. The diamond at the bottom of each plot represents the overall pooled estimate with its confidence interval

in the HF patients with the pooled AUC of 0.764 [95% CI: 0.617–0.910]. The lowest pooled AUC belonged to the "Neural Network" model, based on 3 included studies, with a pooled AUC of 0.580 [95% CI: 0.558–0.602]. The results of the pooled AUCs of all of the models are reported in Table 3.

AUC for the prediction of the readmission – 1 year and more

The results of our meta-analyses revealed that the "Random Forest" model, based on 4 included studies, had the best AUC for the prediction of readmission in the HF patients with the pooled AUC of 0.721 [95% CI: 0.614–0.828]. The lowest pooled AUC belonged to the "Logistic regression" model, based on 3 included studies, with a pooled AUC of 0.649 [95% CI: 0.535–0.763]. The results

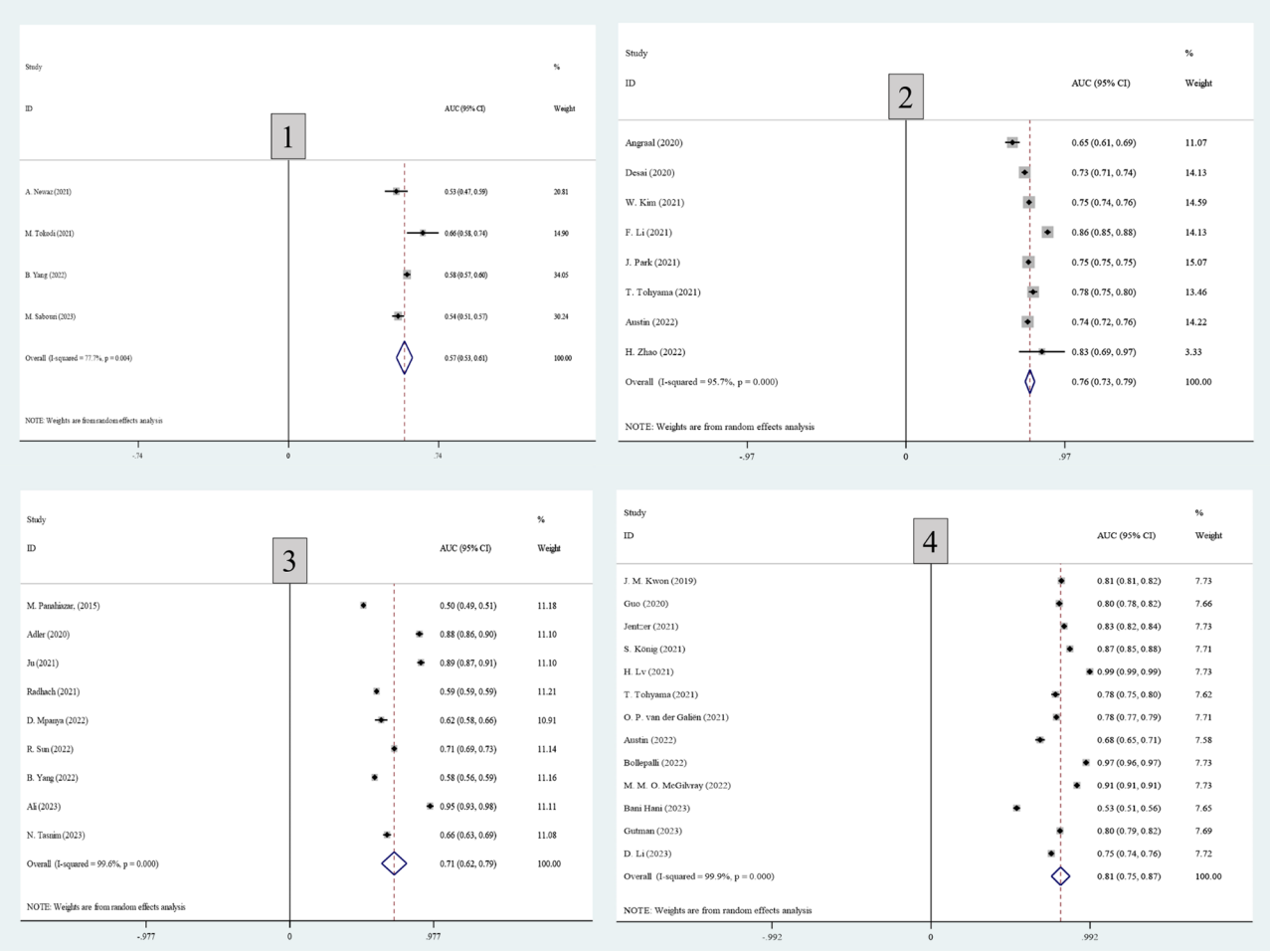

Fig. 4 Forest plots summarizing the meta-analysis of the accuracy of machine learning models for predicting mortality in heart failure (HF) patients. (1) It represents the pooled area under the curve (AUC) values for the K-Nearest Neighbors model across multiple studies, (2) It shows the pooled AUC for the Lasso Regression model, (3) It presents the pooled AUC for the Decision Tree model, and (4) It shows the pooled AUC for the Neural Network model. Each plot includes individual study estimates with corresponding confidence intervals and study weights. The diamond at the bottom of each plot represents the overall pooled estimate with its confidence interval

Table 2 Pooled analysis of other evaluation metrics (accuracy, sensitivity, specificity) for mortality prediction

Model	Accuracy			Sensitivity			Specificity		
	N	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval
Random forest	15	0.600	0.415–0.786	12	0.619	0.547–0.691	8	0.755	0.704–0.806
Logistic regression	10	0.576	0.366–0.786	9	0.592	0.508–0.676	7	0.745	0.685–0.805
Support vector machine	8	0.743	0.694–0.791	7	0.590	0.492–0.687	5	0.761	0.572–0.950
Gradient boosting	12	0.706	0.467–0.945	12	0.678	0.527–0.829	8	0.757	0.664–0.850
Neural Network	4	0.441	0.061–0.820	3	0.623	0.503–0.743	3	0.774	0.657–0.891
Decision tree	2	0.810	0.504–1.117	3	0.609	0.224–0.994	1	NA	NA
K-nearest neighbors	2	0.601	0.576–0.627	2	0.339	0.314–0.364	2	0.685	0.592–0.778
Lasso regression	4	0.413	0.079–0.747	2	0.776	0.564–0.988	2	0.701	0.610–0.793
Bayesian network	1	NA	NA	0	NA	NA	0	NA	NA

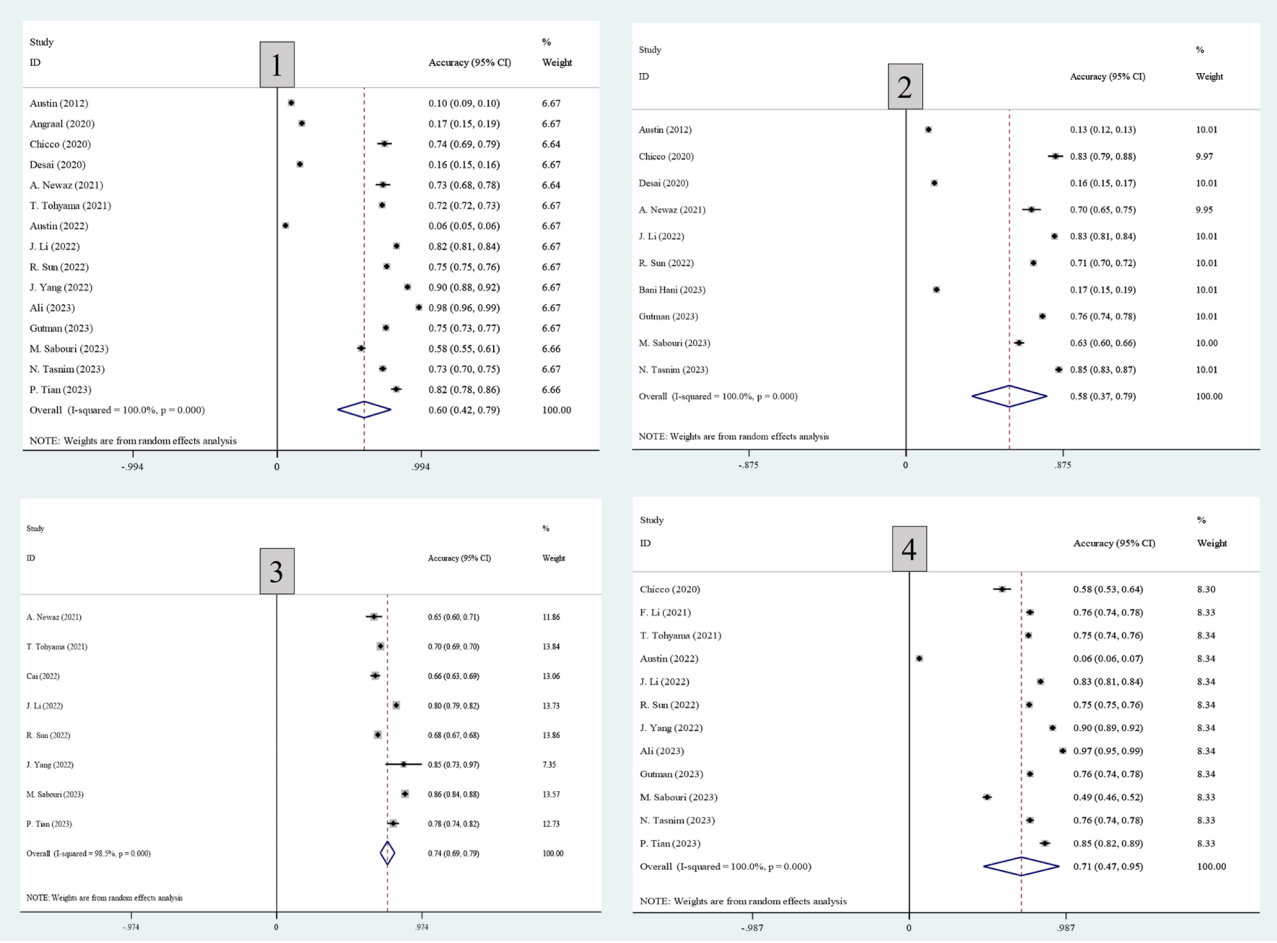

Fig. 5 Forest plots reporting the results of the meta-analysis of the accuracy of machine learning models for predicting mortality in heart failure (HF) patients. (1) It represents the pooled accuracy for the Random Forest model across multiple studies, (2) It shows the pooled accuracy for the Logistic Regression model, and (3) It presents the pooled accuracy for the Support Vector Machine model. (4) It shows the pooled accuracy values for the Gradient Boosting model. Each plot includes individual study estimates with corresponding confidence intervals and study weights. The diamond at the bottom of each plot represents the overall pooled estimate with its confidence interval

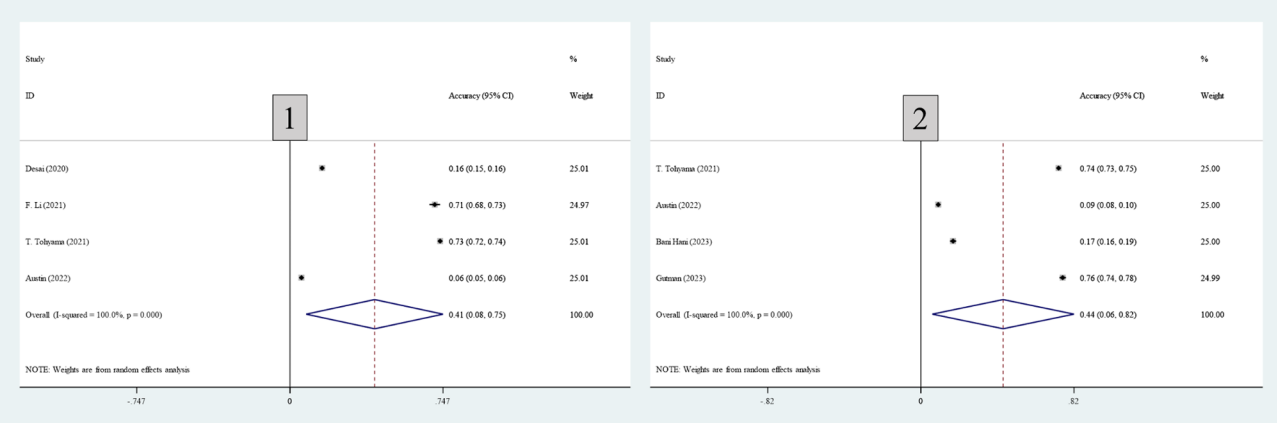

Fig. 6 Forest plots summarizing the meta-analysis of the accuracy of machine learning models for predicting mortality in heart failure (HF) patients. (2) It shows the pooled accuracy for the Lasso Regression model and (4) It shows the pooled accuracy for the Neural Network model. Each plot includes individual study estimates with corresponding confidence intervals and study weights. The diamond at the bottom of each plot represents the overall pooled estimate with its confidence interval

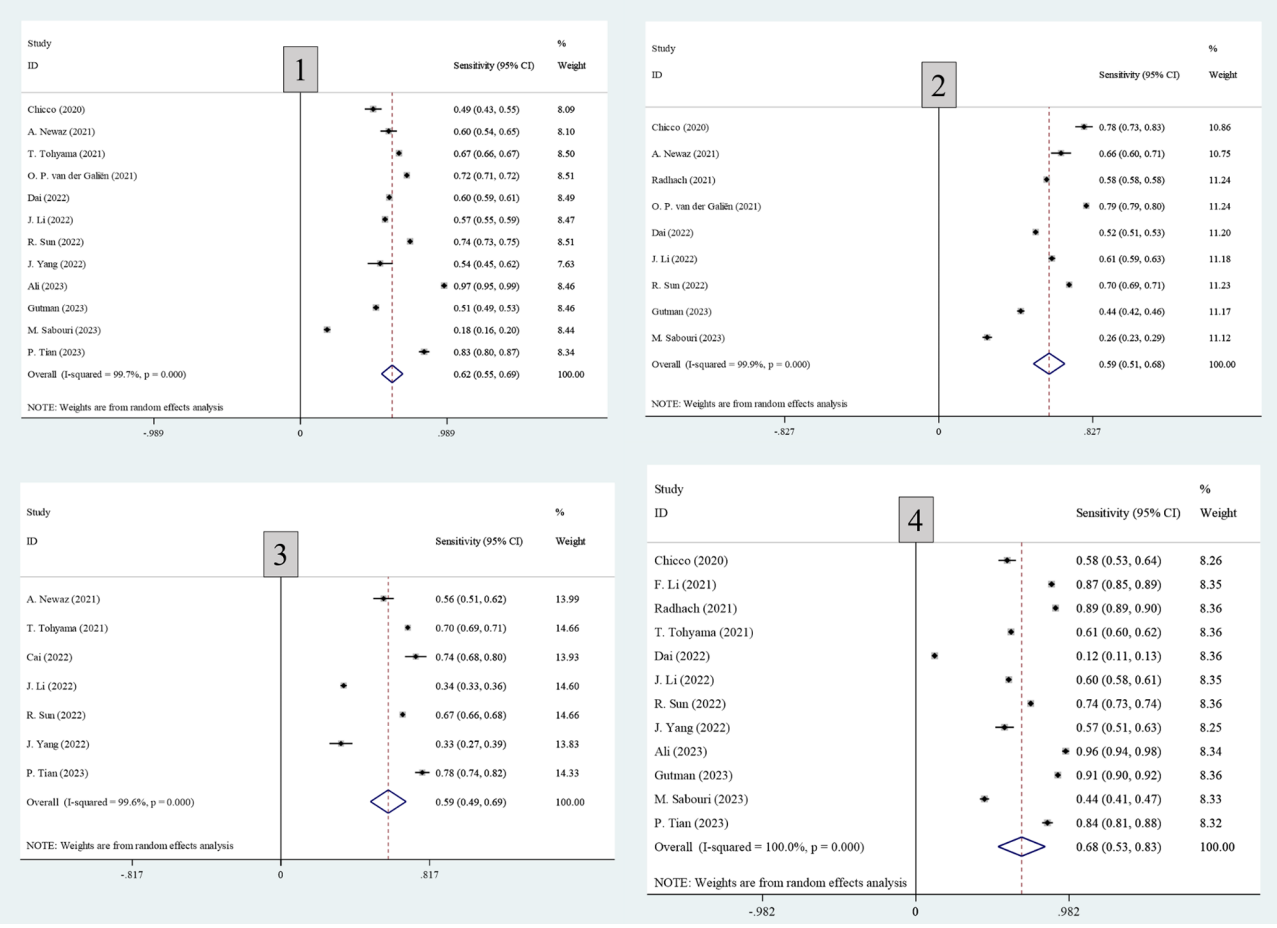

Fig. 7 Forest plots reporting the results of the meta-analysis of the accuracy of machine learning models for predicting mortality in heart failure (HF) patients. (1) It represents the pooled sensitivity for the Random Forest model across multiple studies, (2) It shows the pooled sensitivity for the Logistic Regression model, and (3) It presents the pooled sensitivity for the Support Vector Machine model. (4) It shows the pooled sensitivity for the Gradient Boosting model. Each plot includes individual study estimates with corresponding confidence intervals and study weights. The diamond at the bottom of each plot represents the overall pooled estimate with its confidence interval

of the pooled AUCs of all of the models are reported in Table 3.

AUC for the prediction of the readmission – not specific

We pooled the data from studies that did not report specific periods for readmission to the hospital. The results of our meta-analyses revealed that the "Gradient boosting" model, based on 9 included studies, had the best AUC for the prediction of readmission in the HF patients with the pooled AUC of 0.767 [95% CI: 0.697–0.836]. The lowest pooled AUC belonged to the "Neural Network" model, based on 4 included studies, with a pooled AUC of 0.671 [95% CI: 0.512–0.830]. The results of the pooled AUCs of all of the models are reported in Table 3.

Other metrics for the prediction of readmission – overall

The pooled analysis of various machine learning models for readmission prediction was also evaluated using accuracy, sensitivity, and specificity metrics (Table 4).

For accuracy, the Support Vector Machine (SVM) model, based on 3 included studies, achieved the highest AUC of 0.833 [95% CI: 0.696–0.970], followed by the Gradient Boosting model, based on 6 included studies, with an AUC of 0.782 [95% CI: 0.683–0.882]. The Random Forest model, based on 5 included studies, showed moderate accuracy with an AUC of 0.736 [95% CI: 0.573–0.898], while K-Nearest Neighbors (KNN), based on 2 included studies, had an AUC of 0.668 [95% CI: 0.379–0.957]. The lowest accuracy was observed for Logistic Regression, based on 4 included studies, with an AUC of 0.590 [95% CI: 0.532–0.649].

For sensitivity, the Support Vector Machine (SVM) model, based on 2 included studies, had the highest AUC of 0.835 [95% CI: 0.767–0.903], followed closely by the Random Forest model, based on 4 included studies, with an AUC of 0.801 [95% CI: 0.689–0.912]. The Gradient Boosting model, based on 4 included studies, showed an AUC of 0.756 [95% CI: 0.669–0.842], while K-Nearest

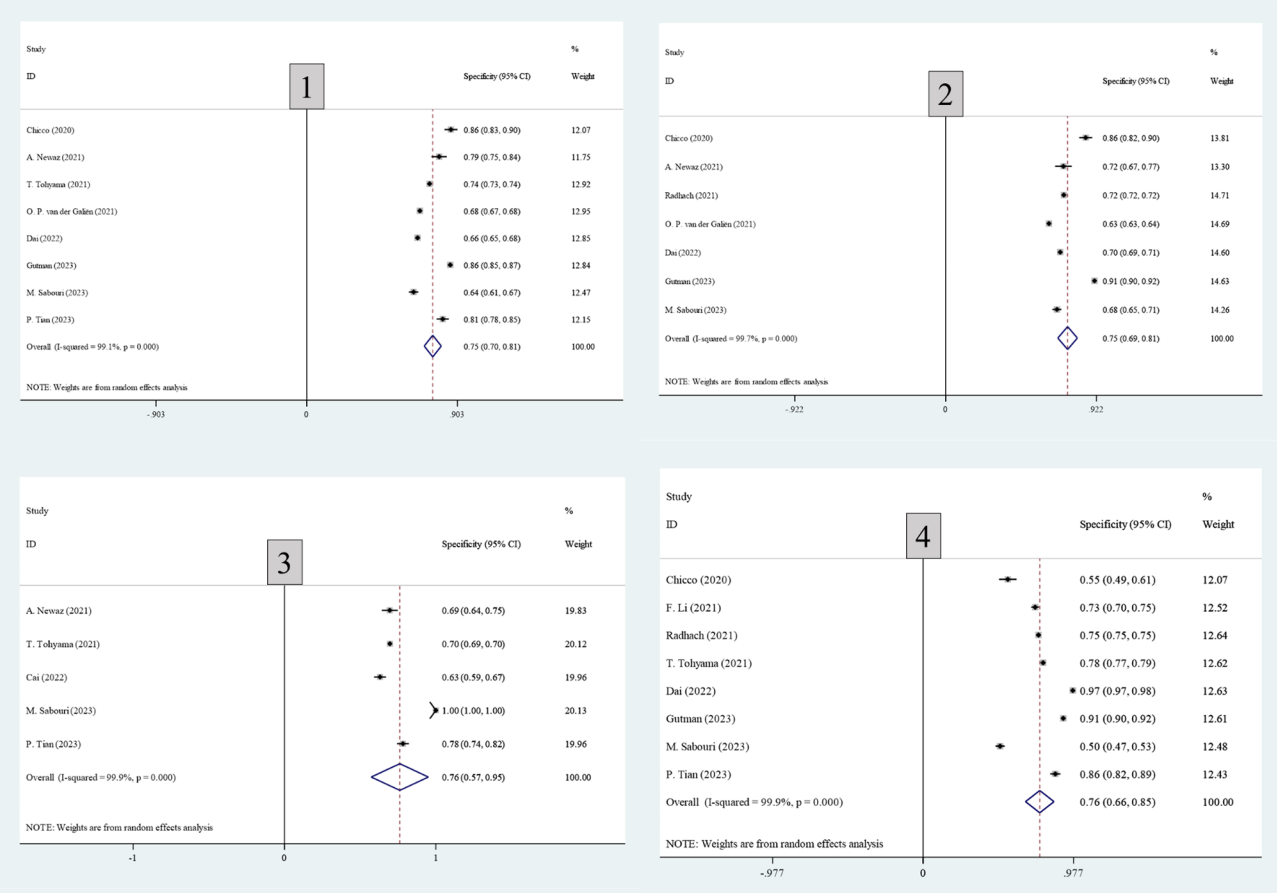

Fig. 8 Forest plots reporting the results of the meta-analysis of the accuracy of machine learning models for predicting mortality in heart failure (HF) patients. (1) It represents the pooled specificity for the Random Forest model across multiple studies, (2) It shows the pooled specificity for the Logistic Regression model, and (3) It presents the pooled specificity for the Support Vector Machine model. (4) It shows the pooled specificity for the Gradient Boosting model. Each plot includes individual study estimates with corresponding confidence intervals and study weights. The diamond at the bottom of each plot represents the overall pooled estimate with its confidence interval

Neighbors (KNN) had a lower sensitivity of 0.656 [95% CI: 0.370–0.942]. The lowest sensitivity was observed for Logistic Regression, based on 4 included studies, with an AUC of 0.560 [95% CI: 0.406–0.713].

For specificity, the Support Vector Machine (SVM) model again achieved the highest AUC of 0.950 [95% CI: 0.852–1.049], followed by the Gradient Boosting model, based on 3 included studies, with an AUC of 0.885 [95% CI: 0.755–1.015]. The Random Forest model, based on 4 included studies, had an AUC of 0.797 [95% CI: 0.527–1.066], while Logistic Regression, based on 4 included studies, had an AUC of 0.615 [95% CI: 0.541–0.689]. The specificity of K-Nearest Neighbors (KNN) was not available due to insufficient data.

Discussion

In this systematic review, we discovered that machine learning models are able to strongly predict readmission and mortality outcomes in HF. Our analysis included 64 studies, encompassing a total of 943,941 patients, with 346 machine learning models developed for mortality and readmission predictions. Notably, neural network models demonstrated the highest predictive accuracy for overall mortality with a pooled AUC of 0.808, followed closely by gradient boosting models with a pooled AUC of 0.796. In contrast, logistic regression and decision tree models showed lower predictive performance. To predict the overall readmission, the support vector machine model showed the highest pooled AUC of 0.726. On the other hand, the decision tree model with a pooled AUC of 0.618 had the lowest predictive ability. These findings underscore the potential of machine learning algorithms to enhance prognostic accuracy in HF, thereby facilitating more personalized and effective patient management strategies.

Heart failure symptoms as predictors of risk: clinical implications

HF is a complex clinical syndrome characterized by the heart's inability to maintain adequate blood circulation,

Table 3 Pooled AUCs of readmission prediction algorithms

Model	Under 1-year readmission			1-year and more readmission			Not specified timing			Overall		
	N	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval
Random forest	8	0.682	0.591–0.774	4	0.721	0.614–0.828	10	0.698	0.612–0.784	20	0.688	0.630–0.746
Logistic regression	7	0.622	0.557–0.686	3	0.649	0.535–0.763	9	0.682	0.642–0.722	19	0.652	0.611–0.692
Support vector machine	3	0.764	0.617–0.910	1	NA	NA	6	0.740	0.610–0.870	10	0.733	0.647–0.756
Gradient boosting	7	0.645	0.583–0.706	2	0.673	0.518–0.829	9	0.767	0.697–0.836	18	0.702	0.649–0.758
Neural Network	3	0.580	0.558–0.602	2	0.660	0.520–0.799	4	0.671	0.512–0.830	7	0.647	0.567–0.727
K-nearest neighbors	3	0.716	0.544–0.887	1	NA	NA	0	NA	NA	3	0.706	0.527–0.885
Lasso regression	0	NA	NA	0	NA	NA	4	0.715	0.626–0.804	5	0.716	0.633–0.798
Decision tree	3	0.618	0.558–0.679	0	NA	NA	1	NA	NA	3	0.618	0.558–0.679
naïve bayes	2	0.656	0.591–0.720	0	NA	NA	1	NA	NA	3	0.618	0.558–0.679

leading to symptoms such as dyspnea, fatigue, and fluid retention [75]. These symptoms are not only indicative of the disease's presence but also serve as critical predictors of patient risk and prognosis [76]. For instance, the severity of dyspnea correlates strongly with mortality and hospitalization rates [77]. Additionally, biomarkers like natriuretic peptides, which reflect cardiac stress and fluid overload, are valuable in predicting adverse outcomes [78]. Understanding and quantifying these symptoms and biomarkers enable clinicians to stratify risk more accurately, guiding therapeutic decisions and improving patient management in HF [79]. Identification of these biomarkers may additionally lead to the generation of more accurate ML models.

Existing prognostic tools for heart failure: the role of GRACE and beyond

Current predictive tools for HF, such as the Global Registry of Acute Coronary Events (GRACE) score and the TIMI score, have been instrumental in advancing the prognosis of HF by incorporating a wide range of clinical parameters [80, 81]. The GRACE score, which includes variables like age, heart rate, and creatinine levels, has been validated across diverse populations and has shown robust performance in predicting mortality and adverse events in HF patients [82]. However, despite its strengths, the GRACE model and similar traditional tools often face limitations in handling the complexity and heterogeneity of HF data [83]. ML models, on the other hand, offer significant improvements by integrating diverse data sources and identifying complex and non-linear patterns that are not apparent through conventional methods [84]. This capability allows ML models to provide more accurate and personalized predictions, thereby enhancing clinical decision-making and patient outcomes in HF management.

Superiority of machine learning models over traditional prognostic methods

ML models have demonstrated significant superiority over traditional methods in predicting HF prognosis due to their ability to handle large and complex datasets, identify intricate patterns, and continuously improve with more data. Unlike conventional methods, which often rely on static and limited clinical variables, ML models can integrate diverse data sources, including genetic information, biomarkers, and patient-reported outcomes, leading to more accurate and personalized predictions [85]. Additionally, ML models exhibit enhanced discrimination and risk stratification capabilities, which are crucial for effective clinical decision-making [86]. This improved performance underscores the potential of ML models to revolutionize HF management by providing more reliable and actionable insights.

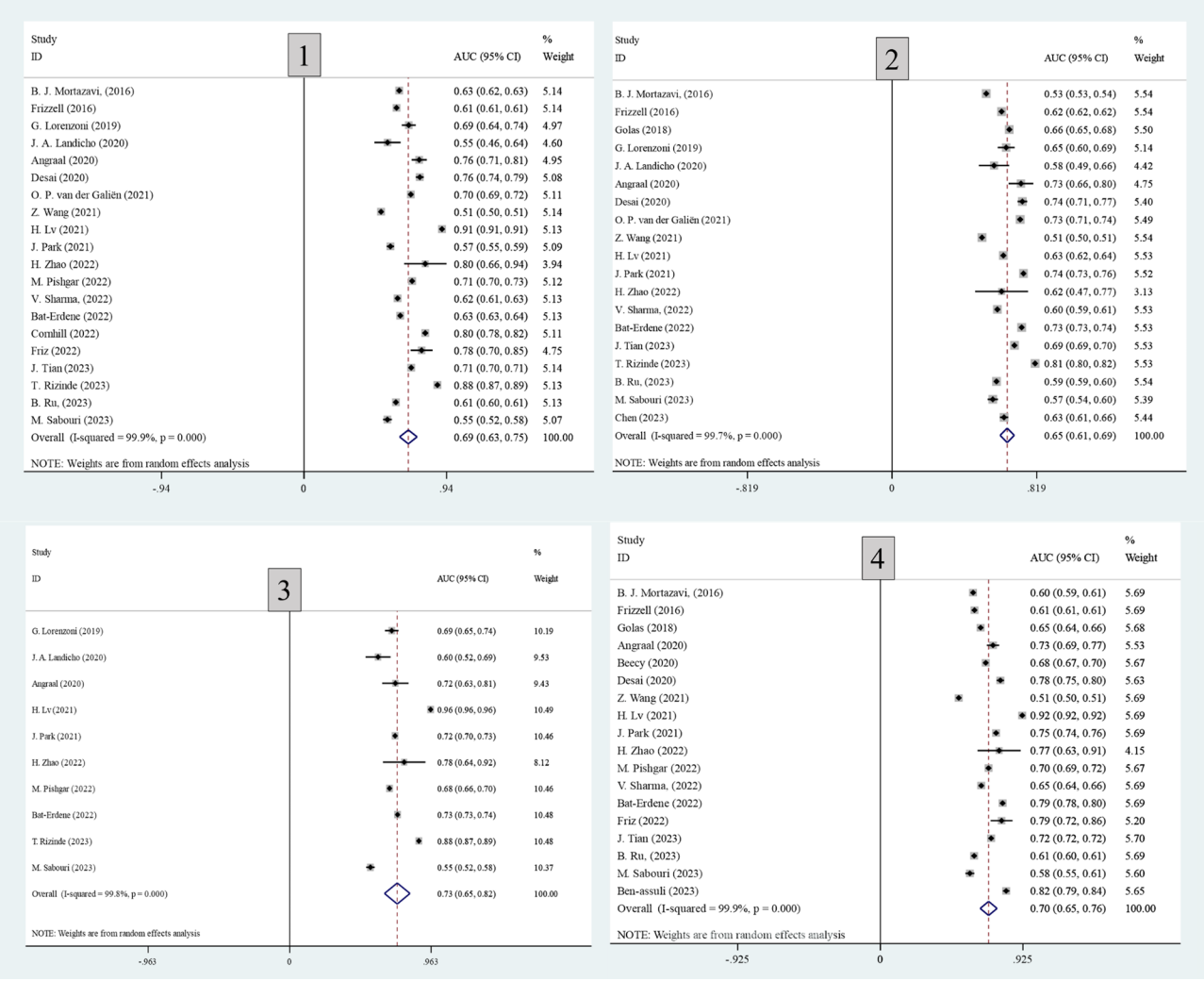

Fig. 9 Forest plots reporting the results of the meta-analysis of the accuracy of machine learning models for predicting hospital readmission in heart failure (HF) patients. (1) It represents the pooled area under the curve (AUC) values for the Random Forest model across multiple studies, (2) It shows the pooled AUC for the Logistic Regression model, and (3) It presents the pooled AUC for the Support Vector Machine model. (4) It shows the pooled accuracy values for the Gradient Boosting model. Each plot includes individual study estimates with corresponding confidence intervals and study weights. The diamond at the bottom of each plot represents the overall pooled estimate with its confidence interval

About neural network models

Studies have shown that ML algorithms, such as neural networks and deep learning models, outperform traditional models in predicting HF outcomes [87, 88]. These advanced models can automatically learn and extract complex patterns from large, high-dimensional datasets, which is often challenging for traditional models. For instance, deep learning models excel at handling unstructured data such as medical images, clinical notes, and time-series data from wearable devices, enabling a more comprehensive analysis [89]. This ability to integrate and analyze diverse data sources leads to more accurate and personalized predictions. Furthermore, neural networks can capture non-linear relationships and complex interactions between variables, which are frequently overlooked by traditional models [88]. This results in enhanced predictive performance and better identification of high-risk patients, ultimately facilitating more effective and timely interventions. The application of these models in real-world data scenarios underscores their potential to revolutionize the field of medical prognosis and improve patient outcomes significantly.

Limitations of current predictive models

Current machine-learning predictive models for HF prognosis in clinical settings face several significant limitations. The interpretability of some of these models such as neural networks is a concern, as they often do not provide clear insights into the underlying factors driving the predictions, making it difficult for clinicians to apply the results effectively in practice [90]. To address these challenges, explainable AI (XAI) techniques have emerged as

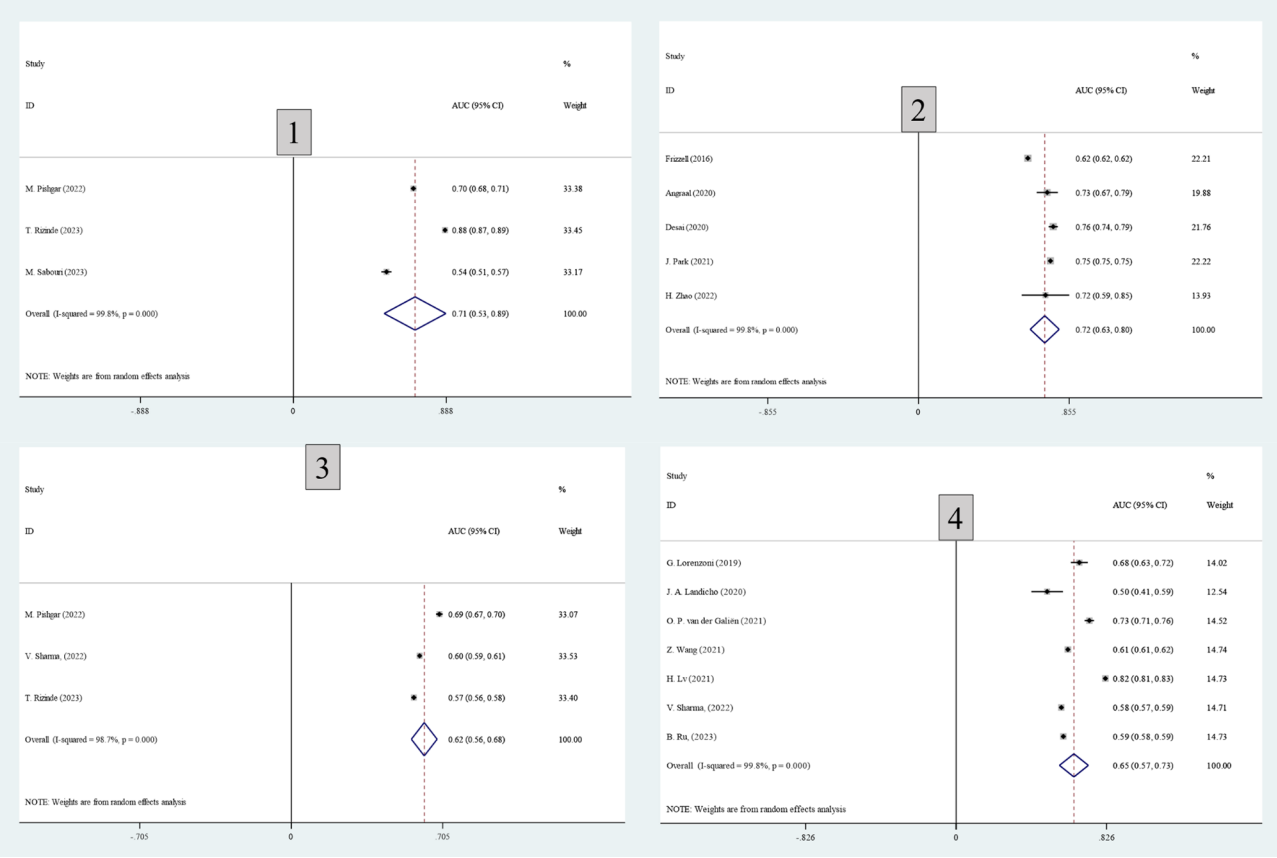

Fig. 10 Forest plots summarizing the meta-analysis of the accuracy of machine learning models for predicting hospital readmission in heart failure (HF) patients. (1) It represents the pooled area under the curve (AUC) values for the K-Nearest Neighbors model across multiple studies, (2) It shows the pooled AUC for the Lasso Regression model, (3) It presents the pooled AUC for the Decision Tree model, and (4) It shows the pooled AUC for the Neural Network model. Each plot includes individual study estimates with corresponding confidence intervals and study weights. The diamond at the bottom of each plot represents the overall pooled estimate with its confidence interval

Table 4 Pooled analysis of other evaluation metrics (accuracy, sensitivity, specificity) for readmission prediction

Model	Accuracy			Sensitivity			Specificity		
	N	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval	*N*	Pooled AUC	Confidence interval
Random forest	5	0.736	0.573–0.898	4	0.801	0.689–0.912	4	0.797	0.527–1.066
Logistic regression	4	0.590	0.532–0.649	4	0.560	0.406–0.713	4	0.615	0.541–0.689
Support vector machine	3	0.833	0.696–0.970	2	0.835	0.767–0.903	2	0.950	0.852–1.049
Gradient boosting	6	0.782	0.683–0.882	4	0.756	0.669–0.842	3	0.885	0.755–1.015
K-nearest neighbors	2	0.668	0.379–0.957	2	0.656	0.370–0.942	1	NA	NA

potential solutions. These methods aim to make AI systems more transparent and understandable by elucidating how input features influence predictions. Techniques such as SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), and Grad-CAM (Gradient-weighted Class Activation Mapping) provide visual or quantitative insights into model behavior [91]. Additionally, inherently interpretable models, such as decision trees and linear regression, or hybrid approaches combining black-box and white-box methods, offer a trade-off between accuracy and explainability [91]. Implementing XAI techniques, from data preprocessing to post-modeling explainability, ensures a more robust and communicable system. Such advancements are essential for fostering clinician trust, improving patient outcomes, and ensuring ethical compliance in clinical decision support systems (CDSS) [91].Furthermore, the generalizability of these models is limited due to variations in etiologies and clinical presentations [90]. Despite advancements in machine learning and artificial intelligence, the performance of these models often remains modest, with C-statistics rarely exceeding 0.8 [92]. Additionally, several critical concerns, such as the roles of physicians and patients in decision-making,

issues of reliability, transparency, accountability, liability, data privacy, biases, monitoring of AI-related adverse events, cybersecurity, and system updates have raised skepticism about adopting AI algorithms in clinical practice [86]. Addressing these barriers in implementing ML in clinical practice requires a multifaceted approach. Robust data governance frameworks, such as compliance with General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA), should be enforced to ensure patient data is securely stored, anonymized, and used only with informed consent [93, 94]. Transparency in algorithm design, including documentation of training data sources and model assumptions, can help mitigate biases stemming from unrepresentative or skewed datasets [95]. Encouraging diverse, multicenter collaborations for data collection can enhance inclusivity and reduce disparities in ML outcomes [96]. Additionally, implementing XAI methodologies will foster trust and accountability in clinical applications [94]. Therefore, implementing ML algorithms in clinical practice is a complex process that necessitates a comprehensive regulatory framework for their research, development, and adoption in medicine [86]. These limitations underscore the need for more robust, interpretable, and dynamic predictive models to improve the prediction of HF prognosis in clinical practice.

Limitations

Our study, while comprehensive, has several limitations that should be acknowledged. Firstly, the reliance on retrospective data in many of the included studies introduces potential biases, as these datasets may not fully capture the dynamic nature of HF progression. Future research should prioritize prospective designs and longitudinal data to enhance model validity and generalizability. Furthermore, HF is a dynamic chronic condition characterized by periods of exacerbation and relative stability. This non-linear progression complicates risk prediction, as symptom exacerbation may not solely indicate disease progression but can also be influenced by social and economic determinants of health. Factors such as socioeconomic status have been shown to significantly impact readmission and mortality rates in HF patients, adding complexity to predictive modeling in this population [97].

Additionally, the heterogeneity in study designs, patient populations, and machine learning models complicates the generalizability of our findings. The heterogeneity of the data included in our meta-analysis reflects the inherent variability in study designs, patient populations, and predictive variables across the selected studies. Such diversity can introduce potential biases and impact the generalizability of the findings. Differences in study methodologies, such as variations in inclusion criteria, sample sizes, and follow-up periods, contribute to inconsistencies in reported outcomes. Furthermore, the use of distinct machine learning algorithms and predictors, ranging from clinical variables to imaging and laboratory data, increases the variability in model performance. This heterogeneity was quantified using the I^2 statistic, which highlighted substantial variability in some analyses, necessitating the use of a random-effects model to account for inter-study differences. Addressing these disparities is crucial for improving the robustness and applicability of meta-analytic conclusions. Furthermore, the variability in the reporting of predictive variables and outcomes across studies posed challenges in data extraction and pooling, which may affect the robustness of our meta-analysis. Additionally, specific variables incorporated within the models were not systematically reviewed, which may influence the interpretability and generalizability of the findings. Future studies should focus on a detailed evaluation of the predictors used in these models to identify key variables driving performance and to ensure alignment with clinical priorities. While we focused on AUC as the primary outcome measure, other important metrics such as sensitivity, specificity, and clinical utility were not extensively analyzed, which could provide a more holistic view of model performance. Importantly, we did not assess the individual contributions of variables within the models in our meta-analysis, which limits our understanding of the specific factors driving model predictions. These limitations highlight the need for standardized methodologies and prospective validation to enhance the reliability and applicability of machine learning models in HF prognosis.

Conclusions

In conclusion, the current systematic review highlights the substantial potential of ML models in predicting readmission and mortality outcomes in HF. These findings underscore the promise of ML algorithms in enhancing prognostic accuracy, leading to more personalized and effective patient management strategies in HF. Despite this promise, challenges such as model interpretability, generalizability, and integration into clinical practice remain. Addressing these issues requires a comprehensive regulatory framework and continued refinement of ML techniques to fully harness their potential in improving HF prognosis and patient outcomes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12872-025-04700-0.

Supplementary Material 1

Acknowledgements
The authors would like acknowledge the clinical research development unit of Imam Ali Hospital Karaj, Iran.

Author contributions
H.H., M.J.A.: Conceptualization, Project Administration, Data curation, Writing- Original Draft, Writing – Review & Editing, Visualization. D.K., A.T.: Validation, Resources, Methodology, Software, Formal analysis, Writing – Original Draft. E.S.: Writing- Original Draft, Data curation. M.T: Writing- Original Draft.

Funding
This study did not receive funding, grant, or sponsorship from any individuals or organizations.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethical
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Conflict of interest
The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript.

Author details
[1]Student Research Committee, Tehran Medical Sciences Branch, Islamic Azad University, Tehran, Iran
[2]Student Research Committee, Isfahan University of Medical Sciences, Isfahan, Iran
[3]Student Research Committee, Faculty of Medicine, Shahed University, Tehran, Iran
[4]Student Research Committee, School of Medicine, Alborz University of Medical Sciences, Karaj, Iran
[5]School of Medicine, Tehran University of Medical Sciences, Tehran, Iran
[6]Universal Scientific Education and Research Network (USERN), Tehran, Iran
[7]Student Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[8]Faculaty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Received: 1 November 2024 / Accepted: 24 March 2025
Published online: 07 April 2025

References

1. Heart failure. Centers for disease control and prevention; 2024-09-13. https://www.cdc.gov/heartdisease/heart_failure.htm
2. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke Statistics-2022 update: a report from the American heart association. Circulation. 2022;22(8):e153–639. https://doi.org/10.1161/cir.0000000000001052.
3. Mozaffarian D, Benjamin EJ, Go AS, et al. Executive summary: heart disease and stroke Statistics–2016 update: A report from the American heart association. Circulation. 2016;26(4):447–54. https://doi.org/10.1161/cir.0000000000000366.
4. Rahman MS, Rahman HR, Prithula J, et al. Heart failure emergency readmission prediction using stacking machine learning model. Diagnostics (Basel). 2023;2(11). https://doi.org/10.3390/diagnostics13111948.
5. Giamouzis G, Kalogeropoulos A, Georgiopoulou V, et al. Hospitalization epidemic in patients with heart failure: risk factors, risk prediction, knowledge gaps, and future directions. J Card Fail. 2011;17(1):54–75. https://doi.org/10.1016/j.cardfail.2010.08.010.
6. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. JACC Heart Fail. 2014;2(5):429–36. https://doi.org/10.1016/j.jchf.2014.04.006.
7. Shin S, Austin PC, Ross HJ, et al. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. ESC Heart Fail. 2021;8(1):106–15. https://doi.org/10.1002/ehf2.13073.
8. Ian Goodfellow YB. Aaron Courville. Deep Learning. 2016.
9. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Bmj. 2009;21:339:b2700. https://doi.org/10.1136/bmj.b2700.
10. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;1(1):W1–33. https://doi.org/10.7326/m18-1377.
11. Adler ED, Voors AA, Klein L, et al. Improving risk prediction in heart failure using machine learning. Eur J Heart Fail. 2020;22(1):139–47. https://doi.org/10.1002/ejhf.1628.
12. Ali MM, Al-Doori VS, Mirzah N, et al. A machine learning approach for risk factors analysis and survival prediction of heart failure patients. Article. Healthc Analytics. 2023;3100182. https://doi.org/10.1016/j.health.2023.100182.
13. Austin DE, Lee DS, Wang CX, et al. Comparison of machine learning and the regression-based EHMRG model for predicting early mortality in acute heart failure. Int J Cardiol. 2022;15:365:78–84. https://doi.org/10.1016/j.ijcard.2022.07.035.
14. Austin PC, Lee DS, Steyerberg EW, Tu JV. Regression trees for predicting mortality in patients with cardiovascular disease: what improvement is achieved by using ensemble-based methods? Biom J. 2012;54(5):657–73. https://doi.org/10.1002/bimj.201100251.
15. Bani Hani S, Ahmad M. Effective prediction of mortality by heart disease among women in Jordan using the Chi-Squared automatic interaction detection model: retrospective validation study. JMIR Cardio. 2023;20:7:e48795. https://doi.org/10.2196/48795.
16. Bollepalli SC, Sahani AK, Aslam N, et al. An optimized machine learning model accurately predicts In-Hospital outcomes at admission to a cardiac unit. Diagnostics (Basel). 2022;19(2). https://doi.org/10.3390/diagnostics12020241.
17. Cai A, Chen R, Pang C, et al. Machine learning model for predicting 1-year and 3-year all-cause mortality in ischemic heart failure patients. Postgrad Med. 2022;134(8):810–9. https://doi.org/10.1080/00325481.2022.2115735.
18. Chen Z, Li T, Guo S, Zeng D, Wang K. Machine learning-based in-hospital mortality risk prediction tool for intensive care unit patients with heart failure. Front Cardiovasc Med. 2023;10:1119699. https://doi.org/10.3389/fcvm.2023.1119699.
19. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med Inf Decis Mak. 2020;3(1):16. https://doi.org/10.1186/s12911-020-1023-5.
20. Dai Q, Sherif AA, Jin C, Chen Y, Cai P, Li P. Machine learning predicting mortality in sarcoidosis patients admitted for acute heart failure. Cardiovasc Digit Health J. 2022;3(6):297–304. https://doi.org/10.1016/j.cvdhj.2022.08.001.
21. Gao Y, Bai X, Lu J, et al. Prognostic value of multiple Circulating biomarkers for 2-year death in acute heart failure with preserved ejection fraction. Front Cardiovasc Med. 2021;8:779282. https://doi.org/10.3389/fcvm.2021.779282.
22. Gao Y, Zhou Z, Zhang B, et al. Deep learning-based prognostic model using non-enhanced cardiac cine MRI for outcome prediction in patients with heart failure. Eur Radiol. 2023;33(11):8203–13. https://doi.org/10.1007/s00330-023-09785-9.
23. Guo AX, Pasque M, Loh F, Mann DL, Payne PRO. Heart failure diagnosis, readmission, and mortality prediction using machine learning and artificial intelligence models. Curr Epidemiol Rep. 2020;7(4):212–9. https://doi.org/10.1007/s40471-020-00259-w.
24. Gutman R, Aronson D, Caspi O, Shalit U. What drives performance in machine learning models for predicting heart failure outcome? Eur Heart J Digit Health. 2023;4(3):175–87. https://doi.org/10.1093/ehjdh/ztac054.
25. Huang Y, Wang M, Zheng Z, et al. Representation of time-varying and time-invariant EMR data and its application in modeling outcome prediction for heart failure patients. J Biomed Inf. 2023;143:104427. https://doi.org/10.1016/j.jbi.2023.104427.
26. Jentzer JC, Kashou AH, Lopez-Jimenez F, et al. Mortality risk stratification using artificial intelligence-augmented electrocardiogram in cardiac

intensive care unit patients. Article. Eur Heart J Acute Cardiovasc Care. 2021;10(5):532–41. https://doi.org/10.1093/ehjacc/zuaa021.
27. Ju C, Zhou J, Lee S, et al. Derivation of an electronic frailty index for predicting short-term mortality in heart failure: a machine learning approach. ESC Heart Fail. 2021;8(4):2837–45. https://doi.org/10.1002/ehf2.13358.
28. Kim W, Park JJ, Lee HY, et al. Predicting survival in heart failure: a risk score based on machine-learning and change point algorithm. Clin Res Cardiol. 2021;110(8):1321–33. https://doi.org/10.1007/s00392-021-01870-7.
29. König S, Pellissier V, Hohenstein S, et al. Machine learning algorithms for claims data-based prediction of in-hospital mortality in patients with heart failure. ESC Heart Fail. 2021;8(4):3026–36. https://doi.org/10.1002/ehf2.13398.
30. Kwon JM, Kim KH, Jeon KH, et al. Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. PLoS ONE. 2019;14(7):e0219302. https://doi.org/10.1371/journal.pone.0219302.
31. Li D, Fu J, Zhao J, Qin J, Zhang L. A deep learning system for heart failure mortality prediction. PLoS ONE. 2023;18(2):e0276835. https://doi.org/10.1371/journal.pone.0276835.
32. Li F, Xin H, Zhang J, Fu M, Zhou J, Lian Z. Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database. BMJ Open. 2021;23(7):e044779. https://doi.org/10.1136/bmjopen-2020-044779.
33. Li J, Liu S, Hu Y, Zhu L, Mao Y, Liu J. Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: retrospective cohort study. J Med Internet Res. 2022;9(8):e38082. https://doi.org/10.2196/38082.
34. Li Y, Wang H, Luo Y. Improving fairness in the prediction of heart failure length of stay and mortality by integrating social determinants of health. Circ Heart Fail. 2022;15(11):e009473. https://doi.org/10.1161/circheartfailure.122.009473.
35. Luo C, Zhu Y, Zhu Z, Li R, Chen G, Wang Z. A machine learning-based risk stratification tool for in-hospital mortality of intensive care unit patients with heart failure. J Transl Med. 2022;20(1):136. https://doi.org/10.1186/s12967-022-03340-8.
36. McGilvray MMO, Heaton J, Guo A, et al. Electronic health Record-Based deep learning prediction of death or severe decompensation in heart failure patients. JACC Heart Fail. 2022;10(9):637–47. https://doi.org/10.1016/j.jchf.2022.05.010.
37. Moreno-Sánchez PA. Improvement of a prediction model for heart failure survival through explainable artificial intelligence. Front Cardiovasc Med. 2023;10:1219586. https://doi.org/10.3389/fcvm.2023.1219586.
38. Mpanya D, Celik T, Klug E, Ntsinjana H. Predicting in-hospital all-cause mortality in heart failure using machine learning. Front Cardiovasc Med. 2022;9:1032524. https://doi.org/10.3389/fcvm.2022.1032524.
39. Newaz A, Ahmed N, Shahriyar Haq F. Survival prediction of heart failure patients using machine learning techniques. Inf Med Unlocked. 2021;26doi. https://doi.org/10.1016/j.imu.2021.100772.
40. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and machine learning for heart failure survival analysis. pp. 40–44.
41. Park J, Hwang IC, Yoon YE, Park JB, Park JH, Cho GY. Predicting Long-Term mortality in patients with acute heart failure by using machine learning. J Card Fail. 2022;28(7):1078–87. https://doi.org/10.1016/j.cardfail.2022.02.012.
42. Radhach ranA, Garikipati A, et al. Prediction of short-term mortality in acute heart failure patients using minimal electronic health record data. BioData Min. 2021;14(1):23. https://doi.org/10.1186/s13040-021-00255-w.
43. Sun R, Wang X, Jiang H, et al. Prediction of 30-day mortality in heart failure patients with hypoxic hepatitis: development and external validation of an interpretable machine learning model. Front Cardiovasc Med. 2022;9:1035675. https://doi.org/10.3389/fcvm.2022.1035675.
44. Tasnim N, Al Mamun S, Islam MS, Kaiser MS, Mahmud M. Explainable mortality prediction model for congestive heart failure with Nature-Based feature selection method. Appl Sciences-Basel. 2023;13(10). https://doi.org/10.3390/app13106138.
45. Tian P, Liang L, Zhao X, et al. Machine learning for mortality prediction in patients with heart failure with mildly reduced ejection fraction. J Am Heart Assoc. 2023;12(12):e029124. https://doi.org/10.1161/jaha.122.029124.
46. Tohyama T, Ide T, Ikeda M, et al. Machine learning-based model for predicting 1 year mortality of hospitalized patients with heart failure. ESC Heart Fail. 2021;8(5):4077–85. https://doi.org/10.1002/ehf2.13556.
47. Tokodi M, Behon A, Merkel ED, et al. Sex-Specific patterns of mortality predictors among patients undergoing cardiac resynchronization therapy: a machine learning approach. Front Cardiovasc Med. 2021;8:611055. https://doi.org/10.3389/fcvm.2021.611055.
48. Wang BH, Ma X, Wang YF, et al. In-Hospital mortality prediction for heart failure patients using electronic health records and an improved bagging algorithm. J Med Imaging Health Inf. 2020;10(5):998–1004. https://doi.org/10.1166/jmihi.2020.3007.
49. Wang Z, Wang B, Zhou Y, Li D, Yin Y. Weight-based multiple empirical kernel learning with neighbor discriminant constraint for heart failure mortality prediction. J Biomed Inf. 2020;101:103340. https://doi.org/10.1016/j.jbi.2019.103340.
50. Yang J, Yan J, Pei Z, Hu A, Zhang Y, Prediction model for in-hospital mortality, of patients with heart failure based on OPTUNA and light gradient boosting machine. J Mech Med Biology. 2022;22(9). https://doi.org/10.1142/S0219519422400590.
51. Angraal S, Mortazavi BJ, Gupta A, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC Heart Fail Jan. 2020;8(1):12–21. https://doi.org/10.1016/j.jchf.2019.06.013.
52. Bat-Erdene BI, Zheng H, Son SH, Lee JY. Deep learning-based prediction of heart failure rehospitalization during 6, 12, 24-month follow-ups in patients with acute myocardial infarction. Health Inf J. 2022;28(2):14604582221101529. https://doi.org/10.1177/14604582221101529.
53. Beecy AN, Gummalla M, Sholle E, et al. Utilizing electronic health data and machine learning for the prediction of 30-day unplanned readmission or all-cause mortality in heart failure. Cardiovasc Digit Health J. 2020;1(2):71–9. https://doi.org/10.1016/j.cvdhj.2020.07.004.
54. Ben-Assuli O, Heart T, Klempfner R, Padman R. Human-machine collaboration for feature selection and integration to improve congestive heart failure risk prediction. Article. Decis Support Syst. 2023;172113982. https://doi.org/10.1016/j.dss.2023.113982.
55. Cornhill AK, Dykstra S, Satriano A, et al. Machine learning Patient-Specific prediction of heart failure hospitalization using cardiac MRI-Based phenotype and electronic health information. Front Cardiovasc Med. 2022;9890904. https://doi.org/10.3389/fcvm.2022.890904.
56. Friz HP, Esposito V, Marano G, et al. Machine learning and LACE index for predicting 30-day readmissions after heart failure hospitalization in elderly patients. Intern Emerg Med. 2022;17(6):1727–37. https://doi.org/10.1007/s11739-022-02996-w.
57. Frizzell JD, Liang L, Schulte PJ, et al. Prediction of 30-Day All-Cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches. JAMA Cardiol. 2017;1(2):204–9. https://doi.org/10.1001/jamacardio.2016.3956.
58. Golas SB, Shibahara T, Agboola S, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inf Decis Mak. 2018;22(1):44. https://doi.org/10.1186/s12911-018-0620-z.
59. Landicho JA, Esichaikul V, Sasil RM. Comparison of predictive models for hospital readmission of heart failure patients with cost-sensitive approach. Article. Int J Healthc Manag. 2020;1–6. https://doi.org/10.1080/20479700.2020.1797334.
60. Lorenzoni G, Sabato SS, Lanera C, et al. Comparison of machine learning techniques for prediction of hospitalization in heart failure patients. J Clin Med. 2019;8(9). https://doi.org/10.3390/jcm8091298.
61. Mortazavi BJ, Downing NS, Bucholz EM, et al. Analysis of machine learning techniques for heart failure readmissions. Circ Cardiovasc Qual Outcomes. 2016;9(6):629–40. https://doi.org/10.1161/circoutcomes.116.003039.
62. Park J, Sarijaloo FB, Canha C, Zhong X, Wokhlu A. A high-performance machine learning model, to predict 90-day acute heart failure readmission and death in heart failure with preserved ejection fraction. J Am Coll Cardiol. 2021;77(18):783–783.
63. Pishgar M, Theis J, Del Rios M, Ardati A, Anahideh H, Darabi H. Prediction of unplanned 30-day readmission for ICU patients with heart failure. BMC Med Inf Decis Mak. 2022;22(1):117. https://doi.org/10.1186/s12911-022-01857-y.
64. Rizinde T, Ngaruye I, Cahill ND. Comparing machine learning classifiers for predicting hospital readmission of heart failure patients in Rwanda. J Pers Med. 2023;13(9). https://doi.org/10.3390/jpm13091393.
65. Ru B, Tan X, Liu Y, et al. Comparison of machine learning algorithms for predicting hospital readmissions and worsening heart failure events in patients with heart failure with reduced ejection fraction: modeling study. JMIR Form Res. 2023;7:e41775. https://doi.org/10.2196/41775.
66. Sharma V, Kulkarni V, McAlister F, et al. Predicting 30-Day readmissions in patients with heart failure using administrative data: A machine learning approach. J Card Fail. 2022;28(5):710–22. https://doi.org/10.1016/j.cardfail.2021.12.004.

67. Wang Z, Chen X, Tan X, et al. Using deep learning to identify High-Risk patients with heart failure with reduced ejection fraction. J Health Econ Outcomes Res. 2021;8(2):6–13. https://doi.org/10.36469/jheor.2021.25753.
68. Chen S, Hu W, Yang Y, et al. Predicting six-month re-admission risk in heart failure patients using multiple machine learning methods: a study based on the Chinese heart failure population database. J Clin Med. 2023;21(3). https://doi.org/10.3390/jcm12030870.
69. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. JAMA Netw Open. 2020;3(1):e1918962. https://doi.org/10.1001/jamanetworkopen.2019.18962
70. Lv H, Yang X, Wang B, et al. Machine Learning-Driven models to predict prognostic outcomes in patients hospitalized with heart failure using electronic health records: retrospective study. J Med Internet Res. 2021;23(4):e24996. https://doi.org/10.2196/24996.
71. Sabouri M, Rajabi AB, Hajianfar G, et al. Machine learning based readmission and mortality prediction in heart failure patients. Sci Rep. 2023;13(1):18671. https://doi.org/10.1038/s41598-023-45925-3.
72. Tian J, Yan J, Han G, et al. Machine learning prognosis model based on patient-reported outcomes for chronic heart failure patients after discharge. Health Qual Life Outcomes. 2023;21(1):31. https://doi.org/10.1186/s12955-023-02109-x.
73. van der Galiën OP, Hoekstra RC, Gürgöze MT, et al. Prediction of long-term hospitalisation and all-cause mortality in patients with chronic heart failure on Dutch claims data: a machine learning approach. BMC Med Inf Decis Mak. 2021;21(1):303. https://doi.org/10.1186/s12911-021-01657-w.
74. Zhao H, Li P, Zhong G, et al. Machine learning models in heart failure with mildly reduced ejection fraction patients. Front Cardiovasc Med. 2022;9:1042139. https://doi.org/10.3389/fcvm.2022.1042139.
75. Prescott E. Prognostic factors and risk scores in heart failure. In: Dorobanţu M, Ruschitzka F, Metra M, editors. Current approach to heart failure. Springer International Publishing; 2016. pp. 575–602.
76. Tsai MF, Hwang SL, Tsay SL, et al. Predicting trends in dyspnea and fatigue in heart failure patients' outcomes. Acta Cardiol Sin Nov. 2013;29(6):488–95.
77. Mentz RJ, Mi X, Sharma PP, et al. Relation of dyspnea severity on admission for acute heart failure with outcomes and costs. Am J Cardiol. Jan 2015;1(1):75–81. https://doi.org/10.1016/j.amjcard.2014.09.048.
78. Jorge AJL, Mesquita ET. Pathophysiology, diagnosis, and management of heart failure. In: Mesquita CT, Rezende MF, editors. Nuclear cardiology: basic and advanced concepts in clinical practice. Springer International Publishing; 2021. pp. 383–97.
79. Singh MS, Thongam K, Choudhary P. Congestive heart failure prediction using artificial intelligence. Springer Nature Singapore; 2024. pp. 355–65.
80. Antman EM, Cohen M, Bernink PJ, et al. The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making. Jama. 2000;16(7):835–42. https://doi.org/10.1001/jama.284.7.835.
81. Eagle KA, Lim MJ, Dabbous OH, et al. A validated prediction model for all forms of acute coronary syndrome: estimating the risk of 6-month postdischarge death in an international registry. Jama. 2004;9(22):2727–33. https://doi.org/10.1001/jama.291.22.2727.
82. McAllister DA, Halbesma N, Carruthers K, Denvir M, Fox KA. GRACE score predicts heart failure admission following acute coronary syndrome. Eur Heart J Acute Cardiovasc Care. 2015;4(2):165–71. https://doi.org/10.1177/2048872614542724.
83. Lee W, Lee J, Woo SI, et al. Machine learning enhances the performance of short and long-term mortality prediction model in non-ST-segment elevation myocardial infarction. Sci Rep. 2021;18(1):12886. https://doi.org/10.1038/s41598-021-92362-1.
84. Greenberg B, Brann A, Campagnari C, Adler E, Yagil A. Machine learning applications in heart failure disease management: hype or hope? Curr Treatment Options Cardiovascular Med. 2021;23(6):35. https://doi.org/10.1007/s11936-021-00912-7
85. Ceccarelli F, Natalucci F, Picciariello L, et al. Application of machine learning models in systemic lupus erythematosus. Int J Mol Sci. 2023;24(5). https://doi.org/10.3390/ijms24054514.
86. Bazoukis G, Stavrakis S, Zhou J, et al. Machine learning versus conventional clinical methods in guiding management of heart failure patients-a systematic review. Heart Fail Rev. 2021;26(1):23–34. https://doi.org/10.1007/s10741-020-10007-3.
87. Teshale AB, Htun HL, Vered M, Owen AJ, Freak-Poli R. A systematic review of artificial intelligence models for Time-to-Event outcome applied in cardiovascular disease risk prediction. J Med Syst. 2024;19(1):68. https://doi.org/10.1007/s10916-024-02087-7.
88. Huang Y, Li J, Li M, Aparasu RR. Application of machine learning in predicting survival outcomes involving real-world data: a scoping review. BMC Med Res Methodol. 2023;13(1):268. https://doi.org/10.1186/s12874-023-02078-1.
89. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. Bioeng (Basel). 2024;29(4). https://doi.org/10.3390/bioengineering11040337.
90. Mpanya D, Celik T, Klug E, Ntsinjana H. Machine learning and statistical methods for predicting mortality in heart failure. Heart Fail Rev. 2021;26(3):545–52. https://doi.org/10.1007/s10741-020-10052-y.
91. Nasarian E, Alizadehsani R, Acharya UR, Tsui K-L. Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. Inform Fusion. 2024. https://doi.org/10.1016/j.inffus.2024.102412/08/01/2024;108:102412.
92. Aaronson KD, Cowger J. Heart failure prognostic models: why bother? Circ Heart Fail. 2012;5(1):6–9. https://doi.org/10.1161/circheartfailure.111.965848.
93. Farhud DD, Zokaei S. Ethical issues of artificial intelligence in medicine and healthcare. Iran J Public Health. 2021;50(11):i–v. https://doi.org/10.18502/ijph.v50i11.7600.
94. Jeyaraman M, Balaji S, Jeyaraman N, Yadav S. Unraveling the ethical enigma: artificial intelligence in healthcare. Cureus. 2023;15(8):e43262. https://doi.org/10.7759/cureus.43262.
95. Hanna M, Pantanowitz L, Jackson B, et al. Ethical and bias considerations in artificial intelligence (AI)/machine learning. Mod Pathol. 2024;16:100686. https://doi.org/10.1016/j.modpat.2024.100686.
96. Charpignon ML, Celi LA, Cobanaj M, et al. Diversity and inclusion: a hidden additional benefit of open data. PLOS Digit Health. 2024;3(7):e0000486. https://doi.org/10.1371/journal.pdig.0000486.
97. Mathews L, Ding N, Mok Y, et al. Impact of socioeconomic status on mortality and readmission in patients with heart failure with reduced ejection fraction: the ARIC study. J Am Heart Assoc. 2022;20(18):e024057. https://doi.org/10.1161/jaha.121.024057.

Publisher's note