Archives of Academic Emergency Medicine. 2026; 14(1): e2

OPEN ACCESS

REVIEW ARTICLE

Comparing Machine Learning Models for Predicting Mortality after Myocardial Infarction: A Systematic Review and Meta-analysis

Seyedhesamoddin Khatami[1], Mohammadsadegh Faghihi[2], Parsa Irajian[2], Aysouda Jafari-Nakhjavanlou[2], Hannanesadat Khatami[2], Reihanesadat Khatami[3], Arash Sarveazad[4,5]*, Mahmoud Yousefifard[2]†

1. Emergency Care Promotion Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
2. Physiology Research Center, Iran University of Medical Sciences, Tehran, Iran.
3. Technische Universität Berlin, Faculty of Electrical Engineering and Computer Science, Berlin, Germany.
4. Colorectal Research Center, Iran University of Medical Sciences, Tehran, Iran.
5. Nursing and Midwifery Care Research Center, Health Management Research Institute, Iran University of Medical Sciences, Tehran, Iran.

Received: July 2025; Accepted: August 2025; Published online: 5 October 2025

Abstract: **Introduction:** Accurate prediction of mortality following myocardial infarction (MI) is critical for timely identification of high-risk patients and optimization of interventions. Conventional statistical models are commonly used; however, advanced machine learning (ML) methods are being increasingly recognized. This meta-analysis aimed to systematically evaluate and compare the predictive performances of various ML models. **Methods:** A systematic search of the Medline (via PubMed), Embase, Scopus, and Web of Science databases was conducted up to January 9, 2025. A total of 14933 articles were identified, of which 330 underwent a full-text review and 69 met the inclusion criteria. The meta-analysis was conducted using a bivariate random-effects model in the 'midas' package of STATA 14. Subgroup analyses were conducted based on the follow-up duration and selected clinical features. The risk of bias was assessed using the QUAPAS. Publication bias and evidence certainty were assessed using Deeks' funnel plots and GRADE framework, respectively. **Results:** Gradient Boosting Machines (GBM), Single Decision Tree Models, and Random Forest models yielded similarly high predictive accuracies. Advanced GBMs, particularly XGBoost (AUC = 0.90, 95% CI: 0.87-0.92; sensitivity = 0.78, 95% CI: 0.74-0.82; specificity = 0.87, 95% CI: 0.83-0.89), showed the highest evidence certainty due to precision and minimal publication bias. Across advanced GBMs, adding echocardiographic parameters increased the sensitivity from 0.77 to 0.83 and specificity from 0.85 to 0.90, indicating a clinically meaningful yet resource-dependent gain in discrimination. **Conclusion:** Advanced Gradient Boosting Machines, particularly XGBoost, currently provide the most reliable mortality predictions in patients with MI. Future research should emphasize external validation, transparent reporting of feature selection, detailed data preprocessing, and dedicated studies on populations with NSTEMI.

Keywords: Machine Learning, Myocardial Infarction, Mortality, Boosting Machine Learning Algorithms, Decision Trees, Random Forest

Cite this article as: Khatami S, Faghihi M, Irajian P, et al. Comparing Machine Learning Models for Predicting Mortality after Myocardial Infarction: A Systematic Review and Meta-analysis. Arch Acad Emerg Med. 2026; 14(1): e2. https://doi.org/10.22037/aaem.v14i1.2783.

1. Introduction

Despite significant advancements in cardiovascular care, myocardial infarction (MI) continues to represent a substantial global health burden with persistently high mortality rates (1). Timely and accurate identification of high-risk patients remains critical, enabling targeted interventions and personalized management strategies to improve outcomes (2, 3).

Traditional scores such as TIMI were originally derived from logistic regression, whereas the latest GRACE 3.0 revision already employs a tree-based boosting algorithm; nonetheless, these tools are still implemented clinically as fixed point-based calculators rather than fully dynamic ML systems (4). However, recent advances in computational methodologies have enabled the analysis of extensive clinical datasets. This has prompted the exploration of advanced Machine Learning (ML) techniques. ML approaches, including Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting Machines, and Neural Networks, can handle complex, multidimensional clinical data more effectively. For example, ML models may more effectively integrate nonlinear relationships and interactions among clinical parameters,

* **Corresponding Author:** Arash Sarveazad; Colorectal Research Center, Iran University of Medical Sciences, Hazrat Rasul Akram educational, research and treatment complex; Niyayesh st., Tehran, Iran. Phone: +982186704771, Email: arashsarveazad@gmail.com, ORCID: https://orcid.org/0000-0001-9273-1940.

† **Corresponding Author:** Mahmoud Yousefifard; Physiology Research Center Iran University of Medical Sciences, Tehran, Iran. Phone: +982186704771, Email: yousefifard20@gmail.com, ORCID: https://orcid.org/0000-0001-5181-4985.

# S.K. and M.F. have equally contributed to this work.

such as biomarkers and imaging findings, thus potentially refining patient risk stratification (5). Despite their promise, the clinical adoption of ML-based predictive models remains limited owing to concerns regarding interpretability, external validity, and feasibility of integration into existing clinical workflows.

Given the rapid increase in studies evaluating ML applications in cardiology, a systematic synthesis of current evidence is essential to guide clinical practice. This systematic review and meta-analysis analyzed key performance metrics, including sensitivity, specificity, and essential classification outcomes such as true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Additionally, the impact of specific clinical features on predictive model accuracy was examined to optimize model selection and clinical applicability. Assessing model performance across diverse follow-up periods, including short-term and long-term mortality, may help tailor predictive strategies to clinical scenarios. Given the differences in clinical outcomes between ST-elevation and Non ST-elevation MI, evaluating the ML model performance separately or comparatively across these subpopulations is essential.

As research on ML-based predictive models expands, insights derived from comprehensive meta-analyses may enhance clinical decision-making by guiding the selection of optimal predictive models. Improved predictive accuracy could lead to more effective risk stratification, timely interventions, and better patient prognosis for MI management. Moreover, explainability techniques (e.g., SHAP) are now routinely paired with tree-based boosting and warrant dedicated scrutiny in cardiac prognostic modelling. This meta-analysis aimed to systematically evaluate and compare the predictive performances of various ML models.

2. Methods

2.1. Eligibility Criteria

This systematic review adhered to the PRISMA 2020 guidelines (6), and the protocol was prospectively registered in PROSPERO (CRD42025633287) (7). Studies were included if they reported prognostic accuracy metrics of ML models predicting mortality, specifically in patients with MI. No restrictions were placed on the language or publication dates.

The exclusion criteria were as follow: 1 studies mixing MI with other Acute Coronary Syndromes populations; 2 articles without mortality outcomes or population sizes; 3 studies lacking quantitative results for ML models; and 4 studies using stacking models owing to methodological inconsistencies and dataset-dependent variability, making reliable meta-analysis challenging.

We excluded mixed ACS cohorts to preserve pathophysiological homogeneity and outcome specificity for MI. Patients with unstable angina exhibit substantially lower event rates and different management algorithms, which can attenuate model signal for mortality prediction. Nevertheless, we recognize that seminal scores (e.g., GRACE, PRAISE) were derived in broader ACS populations; this methodological choice is therefore discussed as a limitation below.

2.2. Information Sources

A comprehensive search was conducted on January 9, 2025, across MEDLINE (via PubMed), Embase, Scopus, and Web of Science. The search strategies for each database are provided in Supplementary Material (Table S1).

2.3. Search Strategy and Selection Process

The search strategy combined Medical Subject Headings (MeSH) and free-text terms related to "myocardial infarction" and "machine learning". Two independent reviewers systematically screened the titles and abstracts and resolved discrepancies by consensus with a third reviewer. A PRISMA flow diagram summarizing this process is shown in Figure 1.

2.4. Data Collection Process and Data Items

Data extraction was independently performed by three reviewers using a standardized form and pilot-tested on ten studies for consistency. Discrepancies were resolved through consensus. Collected data included 1 bibliographic details; 2 study characteristics, such as MI subtype (STEMI, NSTEMI, or combined), MI confirmation method, population size, mortality events, mean age, and male percentage; 3 ML model details, specifically features including ECG, laboratory, angiographic, and echocardiographic findings; 4 clinically relevant follow-up durations (in-hospital, 30-day, 1-6 months, 1-year, and >1 year); and 5 ML model performance metrics.

Individual articles often reported multiple populations, such as training and validation cohorts, or analyses at different follow-up durations. After excluding shorter follow-up analyses within identical populations to avoid double-counting, 336 distinct analyses remained. When studies provided separate results for training, internal validation, or external validation cohorts we abstracted only the latest independent dataset to avoid optimistic bias.

2.5. Risk of Bias Assessment

Methodological quality was evaluated using the QUAPAS tool (8), assessing biases and applicability concerns across five domains (Participants, Index Test, Outcome, Flow and Timing, Analysis). Ratings ("Low", "High" or "Unclear" risk) were assigned independently by two reviewers.

2.6. Effect Measures

The primary effect measures were TP, TN, FP, and FN, which were derived from the sensitivity, specificity, sample size, and mortality counts. The sensitivity and specificity were directly reported or calculated algebraically. For ROC curves without explicit data, 'PlotDigitizer' software was used. Analyses reporting only AUC were excluded, as sensitivity and specificity cannot be precisely reconstructed from AUC alone, de-

spite the established utility of the binormal model. The detailed calculation methodologies have been previously published (9). Because most primary studies reported fixed-time binary mortality, we synthesized sensitivity and specificity at those horizons. Time-to-event metrics (e.g., C-index, time-dependent AUROC, calibration slope) were unavailable in a majority of reports; their future inclusion would enable a more comprehensive prognostic appraisal.

2.7. Synthesis Methods

Data were analyzed using the 'midas' package in Stata 14. Bivariate random-effects model pooled sensitivity, specificity, and prognostic odds ratio (POR). Summary receiver operating characteristic (SROC) curves were used to visualize the prognostic accuracy. Substantial heterogeneity (I^2 statistics) was expected due to variations in MI definitions, patient populations, clinical settings, feature selection, and ML methodologies. Subgroup analyses based on follow-up duration and clinical features reduced heterogeneity.

2.8. Reporting Bias Assessment

Publication bias was evaluated using Deeks' funnel plot asymmetry test (P-value <0.05 indicating significant bias) (10).

2.9. Certainty Assessment

Evidence certainty was assessed using the GRADE framework, considering the risk of bias, inconsistency, indirectness, imprecision, and publication bias (11).

3. Results

3.1. Study Selection

A total of 14933 records were retrieved through database search. After removing 5213 duplicates, 9720 articles remained. Two independent reviewers screened the titles and abstracts, and a third reviewer resolved any disagreement. Of the 330 articles reviewed in the full text, 261 were excluded for reasons outlined in Figure 1, including mixed populations without separate MI groups (n = 134), absence of mortality as a distinct outcome (n = 31), lack of relevance to the study question (n = 28), insufficient reporting of ML metrics (such as reporting only AUC) or mortality data (n = 28), and conference abstracts or incomplete manuscripts (n = 25). Some studies presented duplicate or overlapping data (n = 12) or used unique models incompatible with the meta-analysis (n = 3). Ultimately, 69 articles were included in the final review (Figure 1).

3.2. Study Characteristics

The included studies were published between 2012 and 2025, with a notable increase from 2023 onward. Patient populations primarily consisted of individuals diagnosed with myocardial infarction, including STEMI and NSTEMI, sometimes accompanied by complications such as cardiogenic shock, diabetes mellitus, or hyperuricemia. The study sample sizes varied considerably, ranging from 70 to 755402 patients, with a combined total of 1958806 individuals. The mean age of the patients ranged from 53.8 to 77 years, and the proportion of male participants ranged from 0% (studies exclusively involving female patients) to approximately 86%. MI diagnoses were confirmed through various combinations of clinical presentations, ECG findings, cardiac biomarkers, angiographic data, echocardiography, imaging, or administrative coding systems (e.g., ICD-9/10). Mortality predictions covered diverse follow-up intervals, including in-hospital, 30-day, 1-6 months, 1-year, and >1 year. The number of predictive features utilized in the ML models ranged from three to 53, with one study employing 103 features. Ten distinct ML model categories were assessed, with each study utilizing to one and eight model categories. The detailed characteristics of the individual studies are summarized in Table 1.

3.3. Features Used in ML Models

The number of predictors used in the included ML models ranged from three to 53, with one study reporting up to 103 features. All studies incorporated at least one of the following feature categories: demographic characteristics, past medical history, and vital signs. Additionally, we systematically assessed whether each model included four specific categories: 1 ECG findings (e.g., rhythm and ST-T changes), 2 laboratory results, 3 angiographic data (features obtained from angiography procedures beyond a history of angiography), and 4 echocardiographic findings. Inclusion or exclusion of these feature categories was recorded for each model.

3.4. Syntheses findings

A total of 336 analyses, encompassing ten distinct ML model categories, were included in this meta-analysis. Table 2 presents the pooled performance metrics for each model category, synthesizing area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, likelihood ratios, and POR to evaluate their effectiveness in predicting mortality in patients with MI. SROC curve and sensitivity, specificity, DLR and DLOR plots for each ML model are provided in Supplementary Material (Figures S1-S13).

3.5. Logistic Regression (LR)

Logistic Regression was examined in 72 analyses, with a pooled mortality rate of 0.06. Summary estimates demonstrated an AUROC of 0.86, pooled sensitivity of 0.77, and specificity of 0.81, indicating balanced prognostic performance (Table 2). Inclusion of echocardiographic features reduced sensitivity but increased specificity, implying improved discrimination with potential under-identification of high-risk cases. In 54 studies assessing pure LR models, sensitivity declined and specificity increased at 30-day follow-up. At 1-year follow-up, sensitivity further decreased, while long-term follow-up showed higher sensitivity but reduced specificity (Table 3).

3.6. Single Decision Tree (DT)

Single Decision Tree models were evaluated in 31 studies, with pooled AUROC of 0.90, sensitivity of 0.80, and specificity of 0.88 (Table 2). These models performed consistently across datasets due to minimal dependence on specific input features (Table 3). However, substantial heterogeneity and wide confidence intervals indicate that future evidence may alter current estimates.

3.7. Support Vector Machine (SVM)

Support Vector Machine models were evaluated in 39 studies, with pooled estimates showing an AUROC of 0.86, sensitivity of 0.78, and specificity of 0.84 (Table 2). Sensitivity appeared to decrease with the inclusion of echocardiographic and angiographic data, though evidence remains inconclusive due to limited sample size. Subgroup analysis revealed reduced sensitivity at 30-day and 1-year follow-ups, while both sensitivity and specificity declined in long-term follow-up (Table 3).

3.8. k-Nearest Neighbors (kNN) and Bayesian-based models

kNN models, assessed in nine analyses, showed an AUROC of 0.85, with low sensitivity (0.57) and acceptable specificity (0.87) (Table 2). Bayesian-based models, evaluated in nine observations, lacked sufficient data for robust conclusions and were affected by numerical overflow issues, limiting further analysis. Overall, both models demonstrated limited suitability for mortality prediction in this context.

3.9. Random Forest (RF)

Random Forest models, analyzed in 58 studies, exhibited an AUROC of 0.89, sensitivity of 0.76, and specificity of 0.88 (Table 2). While these models did not demonstrate strong dependencies on ECG or laboratory features, sensitivity increased with the inclusion of echocardiographic data, whereas specificity decreased. This suggests that echocardiography provides additional discriminatory power, but may introduce misclassification in specific scenarios. A similar pattern was observed in the RF model, with reduced sensitivity at the 30-day and 1-year follow-ups. However, in the long-term follow-up subgroup, a decrease in specificity was observed (Table 3).

3.10. Gradient Boosting Machine (GBM)

Classical GBM models, analyzed in 14 studies, reported an AUROC of 0.89, sensitivity of 0.79, and specificity of 0.88 (Table 2). Due to the limited number of studies, the effect of specific input features remains inconclusive. No significant variation in performance was observed across follow-up durations.

Advanced GBM models, evaluated in 60 studies, achieved an AUROC of 0.90, with sensitivity of 0.78 and specificity of 0.87, ranking them among the highest-performing algorithms (Table 2). Figure 2 shows summary receiver operating characteristic (SROC) curve sensitivity and specificity estimates for individual studies in advanced GBM models. Sensitivity increased when ECG data were included, while specificity decreased. Echocardiographic data enhanced both sensitivity and specificity. Subgroup analyses showed reduced sensitivity and specificity at 30-day follow-up, reduced sensitivity but improved specificity at 1-year, and decreased specificity in long-term follow-up (Table 3). In 46 studies focusing on XGBoost, sensitivity declined at 1-year and specificity decreased in long-term follow-up.

AdaBoost, examined in 11 studies, showed an AUROC of 0.82, sensitivity of 0.67, and specificity of 0.88 (Table 2). Although it appeared feature-independent, further studies are needed to validate this observation.

3.11. Neural Networks (NN)

Neural Networks were evaluated in 31 studies, with pooled AUROC of 0.86, sensitivity of 0.81, and specificity of 0.80 (Table 2). Sensitivity declined when additional features, particularly echocardiographic data, were included—suggesting a need for improved model architectures and feature engineering. However, due to the limited number of studies per feature type, these effects remain inconclusive. Subgroup analyses by follow-up duration were not feasible due to insufficient data.

Heterogeneity across models was primarily driven by variations in clinical definitions, follow-up periods, patient characteristics, feature inclusion, and ML methods. Stratification by follow-up duration (e.g., 30-day, 1-year, long-term) notably reduced heterogeneity (Table 3). Decreased performance linked to certain clinical features may reflect suboptimal integration methods rather than limitations of the features themselves. A planned comparison by MI subtype (STEMI vs. NSTEMI) was not performed due to a lack of NSTEMI-specific studies. Future research should aim for standardized MI definitions, consistent follow-up reporting, and transparent documentation of model configurations to improve generalizability and clinical relevance.

3.12. Risk of bias

Most studies showed low risk of bias in the Index Test, Outcome, and Flow and Timing domains, indicating well-defined models, consistent mortality measurement, and appropriate follow-up. Applicability concerns were minimal, supporting the relevance of included populations and methods to clinical settings.

However, issues were noted in Participants (domain 1) and Analysis (domain 5). In domain 1, 50 studies were rated low risk, while five were high risk and 14 unclear due to insufficient details on sampling and inclusion criteria, raising concerns about selection bias. Despite this, all but one study remained applicable to the review question. In domain 5, 63 studies were low risk, and six unclear due to limited information on missing data handling, highlighting a need for better statistical transparency.

Domains 2, 3, and 4 were consistently methodologically robust. Two signaling questions (S5.2 and S4.3) were excluded as not applicable, as noted in Supplementary Table S2. Overall, while most studies were methodologically sound, improved reporting on recruitment and missing data is needed to enhance future research. Risk and applicability assessments are presented in Figure 3.

3.13. Publication bias

Publication bias was evaluated through visual examination of funnel plots and statistically using Deeks' funnel plot asymmetry test for each ML model category. Funnel plots for each model are presented in the Supplementary Material (Figures S14 and S15).

Statistical analysis revealed significant evidence of publication bias (P-value < 0.05) in several ML categories, including the Logistic Regression (P = 0.042), pure LR (P = 0.017), kNN (P = 0.002), RF (P = 0.001), and AdaBoost (P = 0.039).

Conversely, no significant evidence of publication bias was detected in the Single DT-based models (P = 0.084), SVM (P = 0.259), Bayesian-based models (P = 0.836), Classical GBM (P = 0.404), Advanced GBM (P = 0.473), XGBoost (P = 0.469), NN overall (P = 0.286), specifically MLP (P = 0.207), and Simple NN (P = 0.250).

These results indicate that publication bias might have influenced the pooled accuracy estimates for certain ML categories, particularly the kNN, RF, and LR. The findings from these categories should be interpreted with caution.

3.14. Certainty of evidence

The overall certainty of evidence for each ML model was assessed using a structured grading approach, initially set at high certainty and downgraded according to the methodological concerns. LR, single DT-based models, SVM, RF, classical GBM, and Neural Networks have been downgraded primarily because of serious concerns regarding inconsistency (heterogeneity) or publication bias, resulting in a low level of certainty. Bayesian-based models and AdaBoost showed very serious inconsistencies and additional concerns regarding imprecision or publication bias, thus yielding very low certainty. Only the advanced Gradient Boosting Machines maintained moderate certainty, reflecting fewer methodological limitations. The comprehensive details for each model and the specific reasons for downgrading are presented in Table 4.

4. Discussion

This systematic review and meta-analysis synthesized 336 analyses from 69 studies evaluating machine learning (ML) models for predicting mortality following myocardial infarction (MI). Among the ten evaluated model types, advanced gradient boosting machines (particularly XGBoost) consistently demonstrated the highest predictive accuracy, supported by robust evidence and minimal publication bias. Single decision trees and random forests also performed well, offering interpretable and relatively feature-independent alternatives, particularly in resource-constrained settings.

Incorporation of echocardiographic data significantly enhanced both sensitivity and specificity, corroborating prior evidence on the incremental prognostic value of cardiac imaging. Modalities such as left-ventricular ejection fraction, CT-derived plaque burden, and late-gadolinium-enhanced cardiac MRI have been shown to improve risk stratification (12-14). These findings strengthen the biological plausibility of imaging-augmented ML models. However, their reliance on advanced imaging may limit scalability in routine care.

Model explainability remains essential for clinical adoption. Techniques such as SHAP and integrated gradients allow visualization of feature contributions at both the cohort and individual-patient level. Standardized reporting of these explainability measures—alongside performance metrics—will help translate ML predictions into actionable clinical insight. While ongoing advances may shift comparative model rankings, current evidence supports the superior prognostic performance of advanced gradient boosting machines. Future studies should prioritize robust external validation across diverse cohorts, clearly report feature engineering and confusion matrices, and develop clinician-facing interfaces to support practical deployment of ML in post-MI care.

5. Limitations

This review has several limitations. Despite subgroup analyses by follow-up duration and input features, substantial heterogeneity persisted. Contributing factors included inconsistent MI definitions, variable patient characteristics, and unstandardized model development pipelines—particularly in feature selection and preprocessing. Notably, few studies focused exclusively on NSTEMI populations, limiting our ability to compare performance between MI subtypes. This represents an important evidence gap, especially since many benchmark tools (e.g., GRACE) were derived from mixed ACS cohorts. Although we excluded mixed-ACS studies to maintain population homogeneity, this decision may reduce comparability with widely used clinical risk scores.

Our analysis treated mortality as a fixed binary outcome at specific time points, which, while aligned with most source literature, fails to account for censoring or time-to-event dynamics. This may misrepresent underlying hazards. Future studies should adopt survival modeling approaches (e.g., Cox-based GBM or time-dependent AUROC) and report concordance indices, calibration slopes, and decision-curve analyses to more comprehensively evaluate prognostic utility.

A major limitation across included studies was the lack of external validation. Models trained and tested within the same dataset—or using only internal cross-validation—are at risk of overfitting and may yield overly optimistic results. Adherence to TRIPOD-AI (15) and PROBAST-AI (16) guidelines is needed to improve methodological transparency, including clear reporting of dataset splitting strategies, missing data

handling, and model calibration.

6. Conclusions

Among machine learning approaches for post-myocardial infarction prognostication, advanced gradient boosting machines (particularly XGBoost) consistently achieve the highest accuracy, while single decision trees and random forests remain interpretable, lower-complexity alternatives. Incorporating echocardiographic features provides a clinically meaningful improvement, underscoring the prognostic value of structural cardiac data. Future studies should emphasize external validation across diverse STEMI and NSTEMI cohorts, employ survival-based performance metrics, and ensure transparent reporting of feature selection, preprocessing, and explainability methods. Such rigor will facilitate safe clinical adoption and enable personalized, timely interventions for high-risk patients.

7. Declarations

7.1. Acknowledgments

None.

7.2. Author contributions

MY conceived the idea for this systematic review and meta-analysis. MY and SK, developed the study design and methodological approach. SK, MF, PI, AJ, HK, and RK performed the literature search, screened studies, and extracted data independently. PI and AJ assessed the risk of bias. SK performed statistical analyses. SK wrote the initial draft of the manuscript with substantial input from MY. MF, PI, AJ, HK, RK, and AS critically revised the manuscript. SK, MF, and MY directly accessed and verified the underlying data reported in the manuscript. All authors had full access to all data, contributed substantially to interpreting the results, and approved the final manuscript submission. MY had the final responsibility for the decision to submit the manuscript for publication.

7.3. Funding

This study was funded and supported by Iran University of Medical Sciences.

7.4. Conflict of Interest

The authors declare that they have no conflicts of interest relevant to this manuscript. Mahmoud Yousefifard, one of the corresponding authors of this manuscript, is also an editor of the Archives of Academic Emergency Medicine. However, he was not involved in the peer-review or decision-making process for this article.

7.5. Using Artificial Intelligence Chatbots

The editing of this manuscript was assisted by ChatGPT for language and clarity improvement.

References

1. Johansson S, Rosengren A, Young K, Jennings EJBcd. Mortality and morbidity trends after the first year in survivors of acute myocardial infarction: a systematic review. 2017;17(1):53.
2. Sachdeva P, Kaur K, Fatima S, Mahak F, Noman M, Siddenthi SM, et al. Advancements in myocardial infarction management: exploring novel approaches and strategies. 2023;15(9).
3. Reed GW, Rossi JE, Cannon CPJTL. Acute myocardial infarction. 2017;389(10065):197-210.
4. Wenzl FA, Kraler S, Ambler G, Weston C, Herzog SA, Räber L, et al. Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation. 2022;400(10354):744-56.
5. Cho SM, Austin PC, Ross HJ, Abdel-Qadir H, Chicco D, Tomlinson G, et al. Machine learning compared with conventional statistical models for predicting myocardial infarction readmission and mortality: a systematic review. 2021;37(8):1207-14.
6. 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. 2021;372.
7. Schiavo JHJMrsq. PROSPERO: an international register of systematic review protocols. 2019;38(2):171-80.
8. Lee J, Mulder F, Leeflang M, Wolff R, Whiting P, Bossuyt PMJAoim. QUAPAS: an adaptation of the QUADAS-2 tool to assess prognostic accuracy studies. 2022;175(7):1010-8.
9. Khatami R, Faghihi M, Khatami H, Yousefifard M, Khatami SJAoAEM. Calculation of Sensitivity and Specificity from Partial Data for Meta-Analyses: Introducing Some Practical Methods. 2025;13(1):e56-e.
10. Deeks JJ, Macaskill P, Irwig LJJoce. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. 2005;58(9):882-93.
11. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. 2016;353.
12. Molenaar MA, Bouma BJ, Asselbergs FW, Verouden NJ, Selder JL, Chamuleau SA, et al. Explainable machine learning using echocardiography to improve risk prediction in patients with chronic coronary syndrome. 2024;5(2):170-82.
13. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. 2017;38(7):500-7.

14. Pezel T, Sanguineti F, Garot P, Unterseeh T, Champagne S, Toupin S, et al. Machine-learning score using stress CMR for death prediction in patients with suspected or known CAD. 2022;15(11):1900-13.
15. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. 2024;385.
16. Moons KG, Damen JA, Kaul T, Hooft L, Navarro CA, Dhiman P, et al. PROBAST+ AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. 2025;388.
17. Alcober GMI, Lagman AC, Revano TF, editors. Predicting the Mortality of Female Patients suffering from Myocardial Infarction using Data Mining Methods: A Comparison. 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM); 2020: IEEE.
18. Aziz F, Malek S, Ibrahim KS, Raja Shariff RE, Wan Ahmad WA, Ali RM, et al. Short-and long-term mortality prediction after an acute ST-elevation myocardial infarction (STEMI) in Asians: A machine learning approach. 2021;16(8):e0254894.
19. Bai Z, Lu J, Li T, Ma Y, Liu Z, Zhao R, et al. Clinical Feature-Based Machine Learning Model for 1-Year Mortality Risk Prediction of ST-Segment Elevation Myocardial Infarction in Patients with Hyperuricemia: A Retrospective Study. 2021;2021(1):7252280.
20. Barrett LA, Payrovnaziri SN, Bian J, He ZJASoTSP. Building computational models to predict one-year mortality in ICU patients with acute myocardial infarction and post myocardial infarction syndrome. 2019;2019:407.
21. Chen P, Wang B, Zhao L, Ma S, Wang Y, Zhu Y, et al. Machine learning for predicting intrahospital mortality in ST-elevation myocardial infarction patients with type 2 diabetes mellitus. 2023;23(1):585.
22. Chen Z, Zhang L, Li R, Wang J, Chen L, Jin Y, et al. The development and validation of a prognostic prediction modeling study in acute myocardial infarction patients after percutaneous coronary intervention: hemorrhea and major cardiovascular adverse events. 2024;16(9):6216.
23. Cheng C-Y, Chiu I-M, Lin C-HR, Lin X-H, Chen F-C, Hsu T-Y, editors. Stepwise regression machine learning models for in-hospital mortality prediction in patients after ST-segment slevation myocardial infarction (STEMI). 2022 IEEE/ACIS 23rd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD); 2022: IEEE.
24. Cosentino N, Resta ML, Somaschini A, Campodonico J, D'Aleo G, Di Stefano G, et al. ST-Segment elevation acute myocardial infarction complicated by cardiogenic shock: early predictors of very long-term mortality. 2021;10(11):2237.
25. Deng L, Zhao X, Su X, Zhou M, Huang D, Zeng XJBMI, et al. Machine learning to predict no reflow and in-hospital mortality in patients with ST-segment elevation myocardial infarction that underwent primary percutaneous coronary intervention. 2022;22(1):109.
26. Farah C, Adla YA, Awad M, editors. Can machine learning predict mortality in myocardial infarction patients within several hours of hospitalization? A comparative analysis. 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON); 2022: IEEE.
27. Geltser, B. I., et al. "Comparative Analysis of the Effectiveness of Riskometer Scales in Predicting the Risk of in-Hospital Mortality in Patients With ST-Segment Elevation Myocardial Infarction After Percutaneous Coronary Intervention." Kardiologiia 64.8 (2024): 48-55.
28. Ghafari R, Azar AS, Ghafari A, Aghdam FM, Valizadeh M, Khalili N, et al. Prediction of the fatal acute complications of myocardial infarction via machine learning algorithms. 2023;18(4):278.
29. Gong M, Liang D, Xu D, Jin Y, Wang G, Shan PJCiB, et al. Analyzing predictors of in-hospital mortality in patients with acute ST-segment elevation myocardial infarction using an evolved machine learning approach. 2024;170:107950.
30. Gupta MD, Goyal D, Kunal S, Shetty MK, Girish M, Batra V, et al. Comparative evaluation of machine learning models versus TIMI score in ST-segment-elevation myocardial infarction patients. 2025.
31. Hadanny A, Shouval R, Wu J, Shlomo N, Unger R, Zahger D, et al. Predicting 30-day mortality after ST elevation myocardial infarction: Machine learning-based random forest and its external validation using two independent nationwide datasets. 2021;78(5):439-46.
32. Halim MHA, Yusoff YS, Yusuf MMJIJE. Predicting sudden deaths following myocardial infarction in Malaysia using machine learning classifiers. 2018;7(415):4-6.
33. Helgestad OK, Povlsen AL, Josiassen J, Möller S, Hassager C, Jensen LO, et al. Data-driven point-of-care risk model in patients with acute myocardial infarction and cardiogenic shock. 2021;10(6):668-75.
34. Hsieh M-H, Lin S-Y, Lin C-L, Hsieh M-J, Hsu W-H, Ju S-W, et al. A fitting machine learning prediction model for short-term mortality following percutaneous catheterization intervention: a nationwide population-based study. 2019;7(23):732.
35. Iftikhar A, Bond RR, Rjoob K, Knoery C, Leslie SJ, McShane A, et al., editors. Machine learning to predict 30 days and 1-year mortality in STEMI and turndown patients. 2020 Computing in Cardiology; 2020: IEEE.
36. Islam SA, Islam MM, Kahraman HA, Ciril MF, Mizrak A, Tayfur IJA. Developing and validating a new model to predict in-hospital mortality in patients with acute myocardial infarction. 2023;59:12.
37. Kashirina I, Firyulina M, editors. Building models for pre-

dicting mortality after myocardial infarction in conditions of unbalanced classes, including the influence of weather conditions. CEUR Workshop Proceedings; 2020.
38. Kashirina IL, Firyulina MA, Bondarenko YV, Desyatirikova EN, Efimova OE, Chernenkaya LV, editors. Identification of risk factors for mortality after myocardial infarction using machine learning methods. 2021 XXIV International Conference on Soft Computing and Measurements (SCM); 2021: IEEE.
39. Kasim S, Malek S, Cheen S, Safiruz MS, Ahmad WAW, Ibrahim KS, et al. In-hospital risk stratification algorithm of Asian elderly patients. 2022;12(1):17592.
40. Kasim S, Amir Rudin PNF, Malek S, Ibrahim KS, Wan Ahmad WA, Fong AYY, et al. Ensemble machine learning for predicting in-hospital mortality in Asian women with ST-elevation myocardial infarction (STEMI). 2024;14(1):12378.
41. Khera R, Haimovich J, Hurley NC, McNamara R, Spertus JA, Desai N, et al. Use of machine learning models to predict death after acute myocardial infarction. 2021;6(6):633-41.
42. Kononova Y, Abramyan L, Derevitskii I, Babenko AJJoPM. Predictors of Carbohydrate Metabolism Disorders and Lethal Outcome in Patients after Myocardial Infarction: A Place of Glucose Level. 2023;13(6):997.
43. Kwon J-m, Jeon K-H, Kim HM, Kim MJ, Lim S, Kim K-H, et al. Deep-learning-based risk stratification for mortality of patients with acute myocardial infarction. 2019;14(10):e0224502.
44. Lee W, Lee J, Woo S-I, Choi SH, Bae J-W, Jung S, et al. Machine learning enhances the performance of short and long-term mortality prediction model in non-ST-segment elevation myocardial infarction. 2021;11(1):12886.
45. Lee J, Lee S, Street WN, Polgreen LAJBMI, Making D. Machine learning approaches to predict the 1-year-after-initial-AMI survival of elderly patients. 2022;22(1):115.
46. Li Y-m, Jiang L-c, He J-j, Jia K-y, Peng Y, Chen MJT, et al. Machine learning to predict the 1-year mortality rate after acute anterior myocardial infarction in Chinese patients. 2020:1-6.
47. Li Y-W, Yang X-D, Cai Y-C, Kong X-Y, editors. Artificial intelligence-based prediction of acute myocardial infarction mortality risk. 2023 2nd International Conference on Health Big Data and Intelligent Healthcare (ICHIH); 2023: IEEE.
48. Li P, Yao W, Wu J, Gao Y, Zhang X, Hu WJC, et al. Development and validation of a nomogram model for predicting in-hospital mortality in non-diabetic patients with non-ST-segment elevation acute myocardial infarction. 2024;30:10760296241276524.
49. Lin X, Pan X, Yang Y, Yang W, Wang X, Zou K, et al. Machine learning models to predict 30-day mortality for critical patients with myocardial infarction: a retrospective analysis from MIMIC-IV database. 2024;11:1368022.
50. Mohammad MA, Olesen KK, Koul S, Gale CP, Rylance R, Jernberg T, et al. Development and validation of an artificial neural network algorithm to predict mortality and admission to hospital for heart failure after myocardial infarction: a nationwide population-based study. 2022;4(1):e37-e45.
51. Niedziela JT, Cieśla D, Wojakowski W, Gierlotka M, Dudek D, Witkowski A, et al. Is neural network better than logistic regression in death prediction in patients after ST-segment elevation myocardial infarction? 2021;79(12):1353-61.
52. Oliveira M, Seringa J, Pinto FJ, Henriques R, Magalhães TJBMI, Making D. Machine learning prediction of mortality in acute myocardial infarction. 2023;23(1):70.
53. Peng P-C, Chien H-C, Mitra P, Pai T-W, Wang C-H, Liu M-H, editors. Using machine learning approach to predict short-term mortality risk of acute myocardial infarction after emergency admission. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2021: IEEE.
54. Piros P, Ferenci T, Fleiner R, Andréka P, Fujita H, Főző L, et al. Comparing machine learning and regression models for mortality prediction based on the Hungarian Myocardial Infarction Registry. 2019;179:1-7.
55. Piros P, Fleiner R, Kovács L, editors. Random Forest-based predictive modelling on Hungarian Myocardial Infarction Registry. 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE); 2020: IEEE.
56. Razavi SR, Szun T, Zaremba AC, Shah AH, Moussavi ZJM. 1-Year mortality prediction through artificial intelligence using hemodynamic trace analysis among patients with ST elevation myocardial infarction. 2024;60(4):558.
57. Razavi SR, Zaremba AC, Szun T, Cheung S, Shah AH, Moussavi ZJCiB, et al. Comprehensive prediction of outcomes in patients with ST elevation myocardial infarction (STEMI) using tree-based machine learning algorithms. 2025;184:109439.
58. Romanov N, Popa IV, Burlacu A, Brinza C, Fotache M, editors. A Comparison of Interpretable Machine Learning Models to Predict In-Hospital Mortality After Myocardial Infarction: Analyzing Two Years Data from a High-Volume Interventional Center. Proceedings of Seventh International Congress on Information and Communication Technology: ICICT 2022, London, Volume 3; 2022: Springer.
59. Roudini B, Khajehpiri B, Moghaddam HA, Forouzanfar MJIM. Machine learning predicts long-term mortality after acute myocardial infarction using systolic time intervals and routinely collected clinical data. 2024;4(03):170-6.
60. Salman IJTJoEE, Sciences C. Heart attack mortality prediction: an application of machine learning methods. 2019;27(6):4378-89.
61. Shakhgeldyan K, Kuksin N, Domzhalov I, Geltser BJC.

Performance of the Models Predicting In-Hospital Mortality in Patients with ST-Segment Elevation Myocardial Infarction with Predictors in Categorical and Continuous Forms. 2024;16(1 (eng)):15-25.
62. Sherazi SWA, Zheng H, Lee JYJS. A machine learning-based applied prediction model for identification of acute coronary syndrome (ACS) outcomes and mortality in patients during the hospital stay. 2023;23(3):1351.
63. Shetty MK, Kunal S, Girish M, Qamar A, Arora S, Hendrickson M, et al. Machine learning based model for risk prediction after ST-Elevation myocardial infarction: Insights from the North India ST elevation myocardial infarction (NORIN-STEMI) registry. 2022;362:6-13.
64. Song T, Qu XF, Zhang YT, Cao W, Han BH, Li Y, et al. Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. 2014;14(1):59.
65. Sritharan HP, Nguyen H, Ciofani J, Bhindi R, Allahwala UKJFiCM. Machine-learning based risk prediction of in-hospital outcomes following STEMI: the STEMI-ML score. 2024;11:1454321.
66. Tang N, Liu S, Li K, Zhou Q, Dai Y, Sun H, et al. Prediction of in-hospital mortality risk for patients with acute ST-elevation myocardial infarction after primary PCI based on predictors selected by GRACE score and two feature selection methods. 2024;11:1419551.
67. Tarabanis C, Kalampokis E, Khalil M, Alviar CL, Chinitz LA, Jankelson LJCDHJ. Explainable SHAP-XGBoost models for in-hospital mortality after myocardial infarction. 2023;4(4):126-32.
68. Tindale A, Cretu I, Meng H, Panoulas VJEHJACC. Complete revascularization is associated with higher mortality in patients with ST-elevation myocardial infarction, multi-vessel disease and shock defined by hyperlactataemia: results from the Harefield Shock Registry incorporating explainable machine learning. 2023;12(9):615-23.
69. Tsarapatsani K-H, Sakellarios A, Tsakanikas VD, Kleber M, Marz W, Fotiadis DI, editors. Machine learning models predict fatal myocardial infarction within 10-years follow-up utilizing explainable ai. 2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE); 2023: IEEE.
70. Vazquez B, Fuentes-Pineda G, Garcia F, Borrayo G, Prohias JJIiMU. Risk markers by sex for in-hospital mortality in patients with acute coronary syndrome: a machine learning approach. 2021;27:100791.
71. Wallert J, Tomasoni M, Madison G, Held CJBmi, making d. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. 2017;17(1):99.
72. Wang N, Wang M, Zhou Y, Liu H, Wei L, Fei X, et al. Sequential data–based patient similarity framework for patient outcome prediction: algorithm development. 2022;24(1):e30720.
73. Wang Y, Li C, Yuan M, Ren B, Liu C, Zheng J, et al. Development of a complete blood count with differential—based prediction model for in-hospital mortality among patients with acute myocardial infarction in the coronary care unit. 2022;9:1001356.
74. Wang L, Zhang Y, Li F, Li C, Xu HJPo. Mortality prediction of inpatients with NSTEMI in a premier hospital in China based on stacking model. 2024;19(12):e0312448.
75. Xie P, Wang H, Xiao J, Xu F, Liu J, Chen Z, et al. Development and validation of an explainable deep learning model to predict in-hospital mortality for patients with acute myocardial infarction: algorithm development and validation study. 2024;26:e49848.
76. Yan M, Miao Y, Sheng S, Gan X, He B, Shen LJJoSJU. Ensemble learning-based mortality prediction after acute myocardial infarction. 2025;30(1):153-65.
77. Yang J, Li Y, Li X, Tao S, Zhang Y, Chen T, et al. A machine learning model for predicting In-Hospital mortality in Chinese patients with ST-Segment elevation myocardial infarction: findings from the China myocardial infarction registry. 2024;26:e50067.
78. Yu F, Xu Y, Peng JJSCJ. Evaluation of a nomogram model for predicting in-hospital mortality risk in patients with acute ST-elevation myocardial infarction and acute heart failure post-PCI. 2024;58(1):2387001.
79. Zhang Q, Xu L, Xie Z, He W, Huang XJFiCM. Machine learning-based prediction of mortality in acute myocardial infarction with cardiogenic shock. 2024;11:1402503.
80. Zhang Y, Wang L, Li F, Xu H, Pei S, editors. Stacking Ensemble Models Predict Mortality from Acute Myocardial Infarction. International Conference in Communications, Signal Processing, and Systems; 2023: Springer.
81. Zhao J, Zhao P, Li C, Hou YJT, management cr. Optimized machine learning models to predict in-hospital mortality for patients with ST-segment elevation myocardial infarction. 2021:951-61.
82. Zhao P, Liu C, Zhang C, Hou Y, Zhang X, Zhao J, et al. Using machine learning to predict the in-hospital mortality in women with ST-segment elevation myocardial infarction. 2023;24(5):126.
83. Zheng H, Waqar MM, Arif S, Sherazi SWA, Son SH, Lee JY, editors. An Explainable Machine Learning-based Prediction Model for In-hospital Mortality in Acute Myocardial Infarction Patients with Typical Chest Pain. Proceedings of the 2023 6th International Conference on Software Engineering and Information Management; 2023.
84. Zheng H, Sherazi SWA, Lee JYJFiCM. A cost-sensitive deep neural network-based prediction model for the mortality in acute myocardial infarction patients with hypertension on imbalanced data. 2024;11:1276608.
85. Zhu X, Xie B, Chen Y, Zeng H, Hu JJCCA. Machine learning in the prediction of in-hospital mortality in patients with first acute myocardial infarction. 2024;554:117776.

Table 1: Baseline Characteristics of Included Studies Evaluating Machine Learning Models for Post-Myocardial Infarction Mortality Prediction

Author	Population Type	Dataset & Confirmation method	Country	Sample Size	Mean Age	Male %	Mortality numbers	Features number	Features type	Mortality Prediction Time	ML models
Alcober, 2020 (17)	MI	MONICA Project CP + ECG + Cardiac marker	Philippines	509	NR	0	99	4, 8	None	In-hospital	LR, RF
Aziz, 2021 (18)	STEMI	ECG + Cardiac marker	Malaysia	2939, 3130, 6299	56.6, 55.8	86	423, 252, 338	6	LD	1-year, 30-day, In-hospital	LR, SVM, RF
Bai, 2021 (19)	STEMI + Hyperuricemia	CP + ECG + Cardiac marker	China	656	62.5	84	91	16	LD	1-year	LR, KNN, RF, Adv.GBM
Barrett, 2019 (20)	MI + ICU	MIMIC III ICD9 (CP + ECG + Cardiac marker)	USA	5436	72	61	1629	10	LD, TTE	1-year	LR, DT, SVM, Bayesian, RF, AdaBoost, NN
Chen, 2023 (21)	STEMI + T2DM	ESC 2017 CP + ECG + Cardiac marker	China	438	62	71	42	10	LD, TTE	In-hospital	LR, Bayesian, RF, GBM, Adv.GBM
Chen, 2024 (22)	MI	2015 CSC guidelines CP + ECG + Cardiac marker	China	7931	65.8	74	107	10	LD, TTE, CAG	In-hospital	LR, RF, Adv.GBM, NN
Cheng, 2022 (23)	STEMI	2013 ACCF/AHA guideline CP + ECG + Cardiac marker	Taiwan	1567	61.1	84	120	10	LD	In-hospital	LR, RF
Cosentino, 2021 (24)	STEMI + CS	CP + ECG + PCI	Italy	465	68	73	281	4	LD, TTE	Long-term	LR
Deng, 2022 (25)	STEMI	CP + ECG + Cardiac marker	China	854	60.3	83	47	11	LD, TTE	In-hospital	RF, DT, SVM, NN
Farah, 2022 (26)	MI	NR	Russia	1700	NR	NR	271	4	None	In-hospital	LR, SVM, NN
Geltser, 2024 (27)	STEMI	ECG + PCI	Russia	1402, 3273	63	68	95, 223	5	LD, TTE	In-hospital	LR, GBM, NN
Ghafari, 2023 (28)	MI	ECG + Lab data	Iran	1699	61.9	63	271	50	LD, ECG	In-hospital	LR, SVM, RF, Adv.GBM
Gong, 2024 (29)	STEMI	NR	China	3205	63.1	77	297	6	LD	In-hospital	SVM
Gupta, 2024 (30)	STEMI	4th Universal Definition of MI = Cardiac marker + {CP or ECG or Imaging or PCI}	India	1700	54.5	81	210	10	TTE	30-day	LR, DT, RF, Adv.GBM
Hadanny, 2021 (31)	STEMI	ACSIS & MINAP datasets = CP + ECG + Cardiac marker	Israel, United Kingdom	22963	63.5	73	989	10	LD, TTE	30-day	RF
Halim, 2018 (32)	MI	NCVD dataset (CP + ECG + Cardiac marker)	Malaysia	28420	NR	NR	3410	5	None	30-day	KNN
Helgestad, 2021 (33)	MI + CS	3rd Universal Definition of MI = Cardiac marker + {CP or ECG or imaging or PCI}	Denmark	220, 716, 1005	67.4, 67.8	74, 75	123, 340, 545	6	LD, TTE	30-day	LR
Hsieh, 2019 (34)	MI	NHIRD = ICD9CM (CP + ECG + Cardiac marker) + PCI	Taiwan	3421	64.8	75.9	125	4	LD, TTE	30-day	LR, DT, SVM, NN, Bayesian
Iftikhar, 2020 (35)	STEMI	ECG	Northern Ireland	980, 1464	63, 65	76, 68	68, 174	8, 7	None	30-day, 1-year	LR, DT, SVM
Islam, 2023 (36)	MI	ESC 2020 = Cardiac marker + {CP or ECG or Imaging or PCI}	Turkey	70, 209	61, 58	77, 78	5, 13	4	LD, TTE	In-hospital	LR

Table 1: Baseline Characteristics of Included Studies Evaluating Machine Learning Models for Post-Myocardial Infarction Mortality Prediction

Author	Population Type	Dataset & Confirmation method	Country	Sample Size	Mean Age	Male %	Mortality numbers	Features number	Features type	Mortality Prediction Time	ML models
Kashirina, 2020 (37)	MI	NR	Russia	11326	66		1947	6	None	1-year	Adv.GBM
Kashirina, 2021 (38)	MI	NR	Russia	11457	66	62	2025	4	None	30-day	LR, Adv.GBM
Kasim, 2022 (39)	STEMI	NCVD dataset (CP + ECG + Cardiac marker)	Malaysia	1407	72.15	75	505	6	LD	In-hospital	LR, SVM, RF, Adv.GBM, NN
Kasim, 2024 (40)	STEMI	NCVD dataset (CP + ECG + Cardiac marker)	Malaysia	871, 2003	62.2	0	73, 256	12	ECG, LD	In-hospital	DT, SVM, KNN, RF, Adv.GBM, AdaBoost
Khera, 2021 (41)	MI	CP + ECG + Cardiac marker	USA	755402	65	66	33237	9	ECG, LD	In-hospital	LR, Adv.GBM, NN
Kononova, 2023 (42)	MI	CP + ECG + Cardiac marker + Echo	Russia	177			46	5	LD, CAG	Long-term	LR, SVM , RF, Adv.GBM
Kwon, 2019 (43)	STEMI, NSTEMI	CP + ECG + Cardiac marker	South Korea	4882, 5841	66.4, 59.2	72, 78	214, 257	6	LD	In-hospital	LR, RF, NN
Lee, 2021 (44)	NSTEMI	CP + ECG + Cardiac marker	South Korea	4911, 5557, 7716, 8626	66.3	72	120, 273, 306, 281	6	LD	In-hospital, 1-year	LR, SVM, RF, Adv.GBM
Lee, 2022 (45)	MI	ICD9CM (CP + ECG + Cardiac marker)	USA	124031	77	48	19418	6	None	1-year	LR, RF
Li, 2020 (46)	STEMI	CP + ECG + Cardiac marker	China	1244	63.8	78	185	15	LD, TTE	1-year	LR, DT, KNN, Bayesian, RF, Adv.GBM
Li, 2023 (47)	MI	MIMIC IV=ICD9-10 (CP + ECG + Cardiac marker)	USA	4366	63.2	66	436	17	LD	In-hospital	RF, Adv.GBM
Li, 2024 (48)	NSTEMI	NR	China	774	76	69	40	6	LD	In-hospital	LR
Lin, 2024 (49)	MI	MIMIC IV=ICD9-10 (CP + ECG + Cardiac marker)	USA	595	69.4	67	66	6	LD	30-day	LR
Mohammad, 2022 (50)	MI	4th Universal Definition of MI =Cardiac marker+ {CP or ECG or imaging or PCI}	Sweden	27730, 30971, 111558	71, 66	65, 67	10709	21	ECG, LD, TTE	1-year	NN
Niedziela, 2021 (51)	STEMI	ECG+PCI	Poland	10675, 3559	64	57	629, 210	8	TTE	1-6 month	LR, NN
Oliveira, 2023 (52)	MI	ICD9CM (CP + ECG + Cardiac marker)	Portugal	445, 1749		65	32, 218	10	LD	In-hospital	LR, DT, SVM, KNN, RF, Bayesian, Adv.GBM, AdaBoost, NN
Peng, 2021 (53)	MI	NHIRD ICD9CM (CP + ECG + Cardiac marker)	Taiwan	6487	NR	NR	458	103	None	1-6 month	Adv. GBM
Piros, 2019 (54)	MI	HUMIR, ESC 2017 = CP + ECG + Cardiac marker	Hungary	33174	67	60	3703, 6548	12	ECG	30-day, 1-year	LR, DT, NN
Piros, 2020 (55)	MI	HUMIR, ESC 2017 = CP + ECG + Cardiac marker	Hungary	14217, 33174	67	60	2807, 6548	9	LD	1-year	RF
Razavi, 2024 (56)	STEMI	ECG + PCI	Canada	605	64.2	71	51	9	CAG	1-year	KNN, SVM
Razavi, 2025 (57)	STEMI	ECG + PCI	Canada	605	64.2	71	43	20	CAG	1-year	RF, Adv.GBM, AdaBoost
Romanov, 2023 (58)	MI	4th Universal Definition of MI = Cardiac marker + {CP or ECG or imaging or PCI}	Romania	2035	NR	NR	124	6	CAG, TTE	In-hospital	RF, Adv.GBM

Table 1: Baseline Characteristics of Included Studies Evaluating Machine Learning Models for Post-Myocardial Infarction Mortality Prediction

Author	Population Type	Dataset & Confirmation method	Country	Sample Size	Mean Age	Male %	Mortality numbers	Features number	Features type	Mortality Prediction Time	ML models
Roudini, 2024 (59)	MI	4th Universal Definition of MI = Cardiac marker + {CP or ECG or imaging or PCI}	Taiwan	139	63.87	74	87	6	ECG	Long-term	LR, RF, Adv.GBM, AdaBoost
Salman, 2019 (60)	MI	NR	Czech, Syria	787	NR	NR	86	8	LD	In-hospital	LR, DT, Bayesian
Shakhgeldyan, 2024 (61)	STEMI	ECG+PCI	Russia	934, 3740	63	68	64, 254	10	LD, TTE	30-day	LR, RF, Classical.GBM
Sherazi, 2023 (62)	STEMI, NSTEMI	KAMIR-NIH = CP + ECG + Cardiac marker	South Korea	6325, 6779	62.7, 65.1	78, 70	290, 135	39	ECG, LD	In-hospital	DT, RF, Classical.GBM
Shetty, 2022 (63)	STEMI	4th Universal Definition of MI = Cardiac marker + {CP or ECG or imaging or PCI}	India	523, 2668	53.8	84	41, 207	31	LD, CAG, TTE	30-day	LR, DT, RF, Adv.GBM, NN
Song, 2014 (64)	MI	at least 2 of {CP , ECG , Cardiac marker , Echo or MRI}	China	190, 208	60.86	73	9, 12	5	ECG	1-6 month	SVM
Sritharan, 2024 (65)	STEMI	ECG+PCI	Australia	1863	64.9	77	128	5	TTE	In-hospital	LR, DT,SVM, RF, Classical.GBM, AdaBoost
Tang, 2024 (66)	STEMI	ECG + PCI	Russia	4677	61.2	72	318	8,6	LD, TTE	In-hospital	Adv.GBM
Tarabanis, 2023 (67)	MI, STEMI, NSTEMI	ICD9CM (CP + ECG + Cardiac marker)	USA	42965, 457096		0, 68, 1	2362 - 22741	12	ECG	In-hospital	Adv.GBM
Tindale, 2023 (68)	STEMI	ECG + Cardiac marker + PCI	England	408	67.9	75	112	4	LD, CAG	30-day	LR, DT, SVM, KNN, RF, Classical.GBM, Adv.GBM
Tsarapatsani, 2023 (69)	MI	ECG+PCI	Germany	2427	60.6	70	106	12	LD	Long-term	DT, SVM, RF, Adv.GBM
Vázquez, 2021 (70)	STEMI, NSTEMI	MIMIC III = ICD9 (CP + ECG + Cardiac marker)	USA	1299, 2820	67.3, 72.3	65, 58	88, 260	20	LD	In-hospital	Adv.GBM
Wallert, 2017 (71)	MI	ICD10 = CP + ECG + Cardiac marker	Sweden	20777	68.8	65	2284	5	LD	Long-term	LR, SVM, RF
Wang N., 2022 (72)	MI	MIMIC III = ICD9,10(CP + ECG + Cardiac marker)	USA, China	1846, 3010		62	132, 245	3	LD	In-hospital	LR, KNN, RF, NN
Wang Y., 2022 (73)	MI	MIMIC IV = ICD9-10 (CP + ECG + Cardiac marker)	USA	2231	72.2	62	397	12	LD	In-hospital	LR, DT, SVM, KNN, Bayesian, RF, AdaBoost, NN
Wang L., 2024 (74)	NSTEMI	(CP + ECG + Cardiac marker)	China	3061	70	60	150	25 - 49	ECG, LD, TTE	In-hospital	LR, DT, SVM, RF, Classical.GBM, AdaBoost
Xie, 2024 (75)	MI	ICD9-10(CP + ECG + Cardiac marker)	USA, China	3287, 9355	66, 72	66, 68	287, 853	39	LD	In-hospital	LR, RF, Adv.GBM, NN
Yan, 2025 (76)	MI	NR	China	491, 600	63	84	22, 31	4	LD, TTE	Long-term	LR, DT, SVM, RF, Classical GBM, Adv.GBM
Yang, 2024 (77)	STEMI	3rd Universal Definition of MI =Cardiac marker + {CP or ECG or imaging or PCI}	China	9125	65	75	621	48	ECG, LD, TTE	In-hospital	LR, RF, Adv.GBM

Table 1: Baseline Characteristics of Included Studies Evaluating Machine Learning Models for Post-Myocardial Infarction Mortality Prediction

Author	Population Type	Dataset & Confirmation method	Country	Sample Size	Mean Age	Male %	Mortality numbers	Features number	Features type	Mortality Prediction Time	ML models
Yu, 2024 (78)	STEMI+ ADHF	ESC 2020 = Cardiac marker + {CP or ECG or imaging or PCI}	China	84, 196	73	51, 62	17, 42	5	LD, TTE	In-hospital	LR
Zhang Q., 2024 (79)	MI-CS	MIMIC IV = ICD9-10 (CP + ECG + Cardiac marker)	USA	391, 570	68, 72	62, 60	102, 215	13	LD	30-day	LR, Bayesian, Adv.GBM, AdaBoost
Zhang Y., 2024 (80)	NSTEMI	NR	China	3061	NR	NR	150	24 - 53	NR	In-hospital	DT, SVM, RF, Classical.GBM
Zhao, 2021 (81)	STEMI	CP + ECG + Cardiac marker	China	5708	63	74	154	16	LD, CAG	In-hospital	LR, DT, SVM, RF
Zhao, 2023 (82)	STEMI	ESC 2017 = CP + ECG + Cardiac marker	China	2074, 8158	70, 63	0, 75	102, 246	12	LD	In-hospital	DT, SVM, RF, Adv.GBM
Zheng, 2023 (83)	MI	KAMIR-NIH = CP + ECG + Cardiac marker	South Korea	11294	NR	NR	337	16	LD, TTE	In-hospital	LR, DT, SVM, RF, Adv.GBM
Zheng, 2024 (84)	MI	KAMIR-NIH = CP + ECG + Cardiac marker	South Korea	5602	66.2	67	200	27	LD, TTE	Long-term	LR, DT, SVM, RF, Adv.GBM, AdaBoost, NN
Zhu, 2024 (85)	MI	CP + ECG + Cardiac marker	China	1196	65	74	67	8	LD	In-hospital	LR, SVM, RF, Classical.GBM, Adv.GBM, NN

This table summarizes demographic, clinical, and methodological characteristics of the 69 studies included in this systematic review and meta-analysis. Details include population type, country, sample size, mean age, sex distribution, mortality counts, diagnostic criteria for myocardial infarction (MI), feature types used in modeling, prediction time frames, and applied machine learning (ML) algorithms. MI diagnosis was confirmed using combinations of clinical presentation (CP), electrocardiography (ECG), cardiac biomarkers, imaging modalities such as transthoracic echocardiography (TTE), coronary angiography (CAG), and administrative coding systems (ICD-9/10).

Abbreviations: MI, myocardial infarction; STEMI, ST-segment elevation MI; NSTEMI, non-ST-segment elevation MI; CS, cardiogenic shock; ICU, intensive care unit; T2DM, type 2 diabetes mellitus; CP, clinical presentation; ECG, electrocardiography; PCI, percutaneous coronary intervention; LD, laboratory data; CAG, coronary angiography; TTE, transthoracic echocardiography; LR, logistic regression; RF, random forest; DT, decision tree; SVM, support vector machine; KNN, k-nearest neighbors; GBM, gradient boosting machine; Adv.GBM, advanced gradient boosting machine (e.g., XGBoost); NN, neural network; AdaBoost, adaptive boosting; NR, not reported.

Table 2: Pooled Diagnostic Performance of Machine Learning Models in Predicting Mortality Following Myocardial Infarction

Model	N	Dead	Alive	Prevalence	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	+LR (95% CI)	-LR (95% CI)	DOR (95% CI)
1- Logistic regression	72	114848	1797740	0.06	0.86 [0.82 - 0.88]	0.77 [0.74 - 0.80]	0.81 [0.77 - 0.84]	4 [3.4 - 4.8]	0.28 [0.25 - 0.31]	14 [12 - 18]
Pure LR	54	59140	931328	0.06	0.86 [0.83 - 0.89]	0.77 [0.73 - 0.81]	0.82 [0.77 - 0.86]	4.2 [3.4 - 5.3]	0.28 [0.24 - 0.32]	15 [12 - 20]
LR +LASSO	7									
LR +Ridge	4									
MLR	7									
2- Single DT-based	31	16056	135577	0.11	0.90 [0.87 - 0.92]	0.80 [0.72 - 0.86]	0.88 [0.79 - 0.94]	6.9 [3.5 – 13.6]	0.23 [0.15 - 0.34]	30 [11 - 84]
3- SVM	39	10563	123981	0.08	0.86 [0.83 - 0.89]	0.78 [0.73 - 0.82]	0.84 [0.76 - 0.89]	4.7 [3.3 – 6.9]	0.26 [0.21 - 0.32]	18 [11 - 30]
4- KNN	11	4946	36539	0.12	0.85 [0.81 - 0.87]	0.57 [0.35 - 0.77]	0.87 [0.78 - 0.93]	4.5 [2.9 - 6.9]	0.49 [0.3 - 0.8]	9 [4 - 19]
5- Bayesian-based	9									
Naïve Bayes	5									
GNB	5									
6- Random forest	58	44248	350409	0.11	0.89 [0.86 - 0.91]	0.76 [0.71 - 0.81]	0.88 [0.84 - 0.91]	6.4 [4.6 – 8.9]	0.27 [0.22 - 0.33]	24 [15 - 38]
7- Classical-GBM	14	1763	31808	0.05	0.89 [0.86 - 0.91]	0.79 [0.71 - 0.85]	0.88 [0.80 - 0.93]	6.7 [3.9 - 11.4]	0.24 [0.18 - 0.32]	28 [15 - 52]
GBM	6									
SGB	4									
GBDT	4									
8- Advanced-GBM	60	141117	2628485	0.05	0.90 [0.87 - 0.92]	0.78 [0.74 - 0.82]	0.87 [0.83 - 0.89]	5.9 [4.7 - 7.4]	0.25 [0.21 - 0.30]	24 [17 - 33]
XGBoost	46	135318	2579483	0.05	0.89 [0.86 - 0.92]	0.79 [0.75 - 0.83]	0.86 [0.82 - 0.90]	5.8 [4.5 - 7.5]	0.24 [0.20 - 0.29]	24 [17 - 35]
LGBM	5									
CatBoost	9									
9- AdaBoost	11	3172	19400	0.14	0.82 [0.79 - 0.85]	0.67 [0.56 - 0.77]	0.88 [0.76 - 0.94]	5.6 [2.9 - 10.9]	0.37 [0.28 - 0.49]	15 [8 - 30]
10- Neural Network	31	67714	1019230	0.06	0.86 [0.83 - 0.89]	0.81 [0.77 - 0.83]	0.80 [0.74 - 0.86]	4.1 [3.1 – 5.5]	0.24 [0.21 - 0.28]	17 [12 - 24]
DNN	6									
MLP	10	2681	28672	0.09	0.88 [0.85 - 0.90]	0.81 [0.76 - 0.85]	0.82 [0.75 - 0.87]	4.5 [3.3 - 6.1]	0.24 [0.19 - 0.29]	19 [14 - 27]
Simple NN	16	61957	962023	0.06	0.85 [0.81 - 0.88]	0.79 [0.73 - 0.83]	0.81 [0.68 - 0.89]	4.1 [2.4 – 7]	0.26 [0.22 - 0.33]	16 [8 - 29]

This table presents pooled performance estimates from 336 model evaluations across 69 studies, stratified by machine learning (ML) algorithm type.
Metrics include area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, positive likelihood ratio (+LR), negative likelihood ratio (LR), and diagnostic odds ratio (DOR), all reported with 95% confidence intervals (CIs). Estimates were calculated using a bivariate random-effects model.
Abbreviations: AUROC, area under the receiver operating characteristic curve; +LR/LR, positive/negative likelihood ratio; DOR, diagnostic odds ratio; CI, confidence interval; LR, logistic regression; Ridge, ridge regression; LASSO, least absolute shrinkage and selection operator; MLR, multivariable logistic regression; DT, decision tree; SVM, support vector machine; KNN, k-nearest neighbors; GBM, gradient boosting machine; Adv.GBM, advanced GBM (e.g., XGBoost, LGBM, CatBoost); XGBoost, extreme gradient boosting; LGBM, light gradient boosting machine; CatBoost, categorical boosting; AdaBoost, adaptive boosting; NN, neural network; DNN, deep neural network; MLP, multilayer perceptron.

Table 3: Subgroup Analyses of Sensitivity and Specificity Stratified by Clinical Features and Follow-up Duration

Model/Subgroup	N	Sensitivity	Sens. p-value	Specificity	Spec. p-value	I^2	p-value
Logistic Regression						100	
ECG Data	N/A						
Lab Data	N/A						
Angiography Data +	6	0.75 [0.63 - 0.86]	<0.001	0.79 [0.66 - 0.92]	0.04	0	0.69
Angiography Data -	66	0.78 [0.75 - 0.81]		0.81 [0.77 - 0.85]			
Echocardiography Data +	31	0.77 [0.72 - 0.81]	<0.001	0.83 [0.78 - 0.88]	<0.001	0	0.66
Echocardiography Data -	41	0.78 [0.74 - 0.82]		0.79 [0.74 - 0.84]			
Pure LR							
ECG Data +	N/A						
Lab Data	N/A						
Angiography Data +	6	0.75 [0.63 - 0.87]	0.01	0.79 [0.64 - 0.94]	0.06	0	0.65
Angiography Data -	48	0.78 [0.74 - 0.81]		0.82 [0.78 - 0.87]			
Echocardiography Data +	21	0.76 [0.70 - 0.82]	<0.001	0.85 [0.79 - 0.91]	0.01	0	0.38
Echocardiography Data -	33	0.78 [0.74 - 0.83]		0.79 [0.73 - 0.85]			
30 days	11	0.76 [0.68–0.84]	<0.001	0.85 [0.76–0.93]	0.03	0	0.77
1–6 months	2	0.83 [0.67–0.98]	0.50	0.64 [0.30–0.99]	0.17	0	0.48
1 year	5	0.63 [0.48–0.77]	<0.001	0.90 [0.82–0.99]	0.56	62	0.07
Long term	7	0.78 [0.67–0.88]	0.01	0.69 [0.52–0.86]	<0.001	55	0.11
Single DT-based						100	
ECG Data +	8	0.90 [0.83 - 0.97]	0.95	0.94 [0.86 - 1]	0.57	93	<0.001
ECG Data -	22	0.74 [0.65 - 0.83]		0.86 [0.76 - 0.96]			
Lab Data +	25	0.80 [0.73 - 0.88]	0.55	0.91 [0.84 - 0.97]	0.18	92	<0.001
Lab Data -	5	0.76 [0.57 - 0.95]		0.71 [0.36 - 1]			
Angiography Data +	4	0.80 [0.62 - 0.99]	0.47	0.77 [0.43 - 1]	0.43	92	<0.001
Angiography Data -	26	0.80 [0.72 - 0.87]		0.90 [0.83 - 0.97]			
Echocardiography Data +	14	0.78 [0.68 - 0.89]	0.04	0.87 [0.75 - 0.99]	0.31	91	<0.001
Echocardiography Data -	16	0.81 [0.72 - 0.90]		0.90 [0.81 - 0.99]			
30 days	7	0.85 [0.75–0.96]	0.45	0.87 [0.71–1.00]	0.63	14	0.31
1–6 months	N/A						
1 year	4	0.72 [0.49–0.94]	0.15	0.87 [0.66–1.00]	0.99	0	0.57
Long term	4	0.70 [0.45–0.94]	0.14	0.69 [0.28–1.00]	0.17	0	0.38
SVM					100		
ECG Data +	7	0.86 [0.79 - 0.94]	0.16	0.79 [0.61 - 0.97]	0.14	92	<0.001
ECG Data -	31	0.76 [0.71 - 0.81]		0.84 [0.78 - 0.91]			
Lab Data +	32	0.78 [0.73 - 0.83]	0.12	0.85 [0.78 - 0.91]	0.89	91	<0.001
Lab Data -	6	0.74 [0.61 - 0.88]		0.76 [0.56 - 0.97]			
Angiography Data +	4	0.76 [0.61 - 0.90]	0.07	0.81 [0.59 - 1]	0.48	91	<0.001
Angiography Data -	34	0.78 [0.73 - 0.83]		0.84 [0.77 - 0.90]			
Echocardiography Data +	9	0.71 [0.61 - 0.82]	<0.001	0.91 [0.83 - 0.99]	0.75	92	<0.001
Echocardiography Data -	29	0.79 [0.75 - 0.84]		0.80 [0.72 - 0.88]			
30 days	4	0.73 [0.57–0.88]	0.03	0.93 [0.83–1.00]	0.64	14	0.31
1–6 months	2	0.92 [0.78–1.00]	0.28	0.81 [0.50–1.00]	0.90	0	0.43
1 year	6	0.66 [0.53–0.79]	<0.001	0.88 [0.75–1.00]	0.63	57	0.10
Long term	6	0.72 [0.59–0.85]	0.01	0.65 [0.40–0.89]	0.02	64	0.06
KNN						100	
ECG Data	N/A						
Lab Data	N/A						
Angiography Data	N/A						
Echocardiography Data	N/A						
30 days	2	0.13 [0.05–0.32]	0.01	0.91 [0.80–1.00]	0.93	74	0.02
1–6 months	N/A						
1 year	3	0.34 [0.03–0.72]	0.19	0.95 [0.89–1.00]	0.79	48	0.14
Long term	N/A						
Bayesian-based	N/A						
ECG Data							
Lab Data							
Angiography Data							
Echocardiography Data							
Subgroup by F/U duration							
Random Forest						100	
ECG Data +	7	0.92 [0.86 - 0.97]	0.92	0.96 [0.92 - 1]	0.91	95	<0.001
ECG Data -	50	0.73 [0.68 - 0.78]		0.86 [0.82 - 0.91]			

Table 3: Subgroup Analyses of Sensitivity and Specificity Stratified by Clinical Features and Follow-up Duration

Model/Subgroup	N	Sensitivity	Sens. p-value	Specificity	Spec. p-value	I^2	p-value
Lab Data +	50	0.76 [0.71 - 0.82]	0.24	0.89 [0.85 - 0.93]	0.43	91	<0.001
Lab Data -	7	0.73 [0.58 - 0.88]		0.84 [0.70 - 0.98]			
Angiography Data +	8	0.79 [0.67 - 0.91]	0.13	0.86 [0.75 - 0.98]	0.07	91	<0.001
Angiography Data -	49	0.76 [0.70 - 0.81]		0.88 [0.84 - 0.93]			
Echocardiography Data +	19	0.78 [0.70 - 0.86]	0.01	0.85 [0.77 - 0.93]	<0.001	92	<0.001
Echocardiography Data -	38	0.75 [0.69 - 0.81]		0.89 [0.85 - 0.94]			
30 days	8	0.71 [0.56–0.85]	0.03	0.86 [0.74–0.98]	0.06	0	0.64
1–6 months	N/A						
1 year 10	0.67 [0.53–0.81]	<0.001	0.89 [0.80–0.98]	0.06	19	0.29	
Long term	7	0.83 [0.71–0.94]	0.30	0.74 [0.54–0.94]	0.01	60	0.08
Classical GBM						100	
ECG Data +	3	0.77 [0.64 - 0.90]	0.13	0.98 [0.95 - 1]	0.31	92	<0.001
ECG Data -	10	0.77 [0.69 - 0.85]		0.82 [0.74 - 0.90]			
Lab Data N/A							
Angiography Data N/A							
Echocardiography Data +	9	0.79 [0.71 - 0.87]	0.23	0.81 [0.72 - 0.91]	<0.001	93	<0.001
Echocardiography Data -	4	0.74 [0.61 - 0.87]		0.96 [0.93 - 1]			
30 days	3	0.75 [0.60–0.91]	0.08	0.83 [0.65–1.00]	0.26	0	0.59
1–6 months N/A							
1 year N/A							
Long term	2	0.93 [0.85–1.00]	0.33	0.70 [0.38–1.00]	0.09	64	0.06
Advanced GBM						100	
ECG Data +	14	0.82 [0.74 - 0.89]	0.01	0.84 [0.77 - 0.91]	<0.001	0	0.48
ECG Data -	46	0.77 [0.72 - 0.82]		0.87 [0.84 - 0.91]			
Lab Data	N/A						
Angiography Data +	8	0.83 [0.73 - 0.93]	0.08	0.86 [0.77 - 0.95]	<0.001	0	0.66
Angiography Data -	52	0.78 [0.73 - 0.82]		0.87 [0.84 - 0.90]			
Echocardiography Data +	16	0.83 [0.76 - 0.89]	0.01	0.90 [0.86 - 0.95]	<0.001	59	0.09
Echocardiography Data -	44	0.77 [0.72 - 0.82]		0.85 [0.81 - 0.89]			
30 days	7	0.74 [0.61–0.88]	0.02	0.81 [0.70–0.93]	<0.001	0	0.42
1–6 months	1	0.56 [0.11–1.00]	0.25	0.83 [0.54–1.00]	0.82	0	0.49
1 year 8	0.66 [0.51–0.80]	<0.001	0.90 [0.84–0.97]	0.02	58	0.09	
Long term	6	0.79 [0.66–0.93]	0.11	0.72 [0.56–0.88]	<0.001	62	0.07
XGBoost						100	
ECG Data +	14	0.82 [0.75 - 0.88]	<0.001	0.84 [0.77 - 0.92]	<0.001	0	0.6
ECG Data -	32	0.78 [0.73 - 0.83]		0.87 [0.83 - 0.91]			
Lab Data	N/A						
Angiography Data +	5	0.86 [0.76 - 0.95]	0.21	0.89 [0.80 - 0.98]	0.13	4	0.35
Angiography Data -	41	0.78 [0.74 - 0.83]		0.86 [0.82 - 0.90]			
Echocardiography Data +	11	0.82 [0.74 - 0.90]	0.01	0.90 [0.84 - 0.95]	0.01	15	0.31
Echocardiography Data -	35	0.78 [0.73 - 0.83]		0.85 [0.81 - 0.89]			
30 days	3	0.76 [0.58–0.94]	0.13	0.79 [0.59–0.99]	0.10	0	0.58
1–6 months	1	0.56 [0.16–0.96]	0.16	0.83 [0.54–1.00]	0.88	0	0.37
1 year 6	0.68 [0.54–0.83]	<0.001	0.89 [0.81–0.97]	0.08	40	0.19	
Long term	5	0.81 [0.68–0.94]	0.12	0.73 [0.55–0.91]	<0.001	45	0.16
AdaBoost						100	
ECG Data +	3	0.74 [0.56 - 0.91]	0.92	0.88 [0.70 - 1]	0.93	0	0.65
ECG Data -	8	0.65 [0.52 - 0.77]		0.88 [0.78 - 0.98]			
Lab Data +	8	0.64 [0.52 - 0.76]	0.14	0.91 [0.84 - 0.99]	0.24	0	0.37
Lab Data -	3	0.75 [0.58 - 0.92]		0.74 [0.45 - 1]			
Angiography Data	N/A						
Echocardiography Data +	4	0.68 [0.51 - 0.85]	0.56	0.92 [0.82 - 1]	0.67	0	0.55
Echocardiography Data -	7	0.66 [0.53 - 0.80]		0.85 [0.72 - 0.98]			
Subgroup by F/U duration	N/A						
Neural Networks					100		
ECG Data +	6	0.78 [0.71 - 0.85]	<0.001	0.78 [0.63 - 0.93]	0.08	0	0.57
ECG Data -	25	0.81 [0.78 - 0.84]		0.81 [0.75 - 0.87]			
Lab Data +	26	0.80 [0.77 - 0.83]	<0.001	0.83 [0.78 - 0.88]	0.59	58	0.09
Lab Data -	5	0.83 [0.77 - 0.90]		0.62 [0.42 - 0.82]			

Table 3: Subgroup Analyses of Sensitivity and Specificity Stratified by Clinical Features and Follow-up Duration

Model/Subgroup	N	Sensitivity	Sens. p-value	Specificity	Spec. p-value	I^2	p-value
Angiography Data +	3	0.80 [0.70 - 0.91]	0.01	0.83 [0.66 - 1]	0.56	0	0.96
Angiography Data -	29	0.81 [0.77 - 0.84]		0.80 [0.74 - 0.86]			
Echocardiography Data +	14	0.80 [0.75 - 0.85]	<0.001	0.83 [0.75 - 0.91]	0.08	0	0.72
Echocardiography Data -	17	0.81 [0.77 - 0.85]		0.78 [0.70 - 0.87]			
Subgroup by F/U duration	N/A						
MLP	N/A					99	
ECG Data							
Lab Data							
Angiography Data							
Echocardiography Data							
Subgroup by F/U duration	N/A						
Simple NN						100	
ECG Data +	6	0.78 [0.71 - 0.86]	<0.001	0.78 [0.59 - 0.97]	0.37	0	0.88
ECG Data -	10	0.79 [0.73 - 0.85]		0.83 [0.70 - 0.95]			
Lab Data +	11	0.76 [0.70 - 0.82]	<0.001	0.87 [0.78 - 0.95]	0.26	53	0.12
Lab Data -	5	0.83 [0.76 - 0.90]		0.62 [0.37 - 0.88]			
Angiography Data	N/A						
Echocardiography Data	N/A						
Subgroup by F/U duration	N/A						

This table shows pooled sensitivity and specificity values for machine learning (ML) models stratified by the inclusion of clinical data types-electrocardiography (ECG), laboratory data (Lab), angiography, and echocardiography-and by follow-up durations (30 days, 1–6 months, 1 year, long-term). Estimates were generated using bivariate random-effects meta-analysis. Subgroup differences were assessed using p-values and heterogeneity was quantified using I^2 statistics.
Abbreviations: Sens, sensitivity; Spec, specificity; CI, confidence interval; I^2, inconsistency index; p-value, probability value; F/U, follow-up; ECG, electrocardiography; Lab, laboratory data; DT, decision tree; LR, logistic regression; SVM, support vector machine; KNN, k-nearest neighbors; GBM, gradient boosting machine; Adv.GBM, advanced gradient boosting machine; NN, neural network; MLP, multilayer perceptron; Simple NN, simple neural network; N/A, not applicable.

Table 4: GRADE Assessment of Certainty of Evidence for Diagnostic Accuracy of Machine Learning Models

ML Model	Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Overall Certainty
Logistic Regression	Not Serious (0)	Serious (1)	Not Serious (0)	Not Serious (0)	Serious (1)	Low
Single DT-based	Not Serious (0)	Serious (1)	Not Serious (0)	Serious (1)	Not Serious (0)	Low
SVM	Not Serious (0)	Serious (1)	Not Serious (0)	Not Serious (0)	Not Serious (0)	Moderate
KNN	Not Serious (0)	Very Serious (2)	Not Serious (0)	Serious (1)	Serious (1)	Very Low
Bayesian-based	Not Serious (0)	Very Serious (2)	Not Serious (0)	N/A	Not Serious (0)	Very Low
Random Forest	Not Serious (0)	Serious (1)	Not Serious (0)	Not Serious (0)	Serious (1)	Low
Classical GBM	Not Serious (0)	Serious (1)	Not Serious (0)	Serious (1)	Not Serious (0)	Low
Advanced GBM	Not Serious (0)	Serious (1)	Not Serious (0)	Not Serious (0)	Not Serious (0)	Moderate
AdaBoost	Not Serious (0)	Very Serious (2)	Not Serious (0)	Serious (1)	Serious (1)	Very Low
Neural Networks	Not Serious (0)	Very Serious (2)	Not Serious (0)	Not Serious (0)	Not Serious (0)	Low

Certainty of evidence for the diagnostic accuracy of machine learning (ML) models was appraised using the GRADE framework, assessing five domains: risk of bias, inconsistency, indirectness, imprecision, and publication bias. Each domain was graded as not serious (0), serious (1), or very serious (2), yielding an overall certainty rating (high, moderate, low, very low) for each model class.
Abbreviations: GRADE, Grading of Recommendations Assessment, Development and Evaluation; DT, decision tree; SVM, support vector machine; KNN, k-nearest neighbors; GBM, gradient boosting machine; Adv.GBM, advanced gradient boosting machine (e.g., XGBoost); AdaBoost, adaptive boosting; N/A, not applicable.

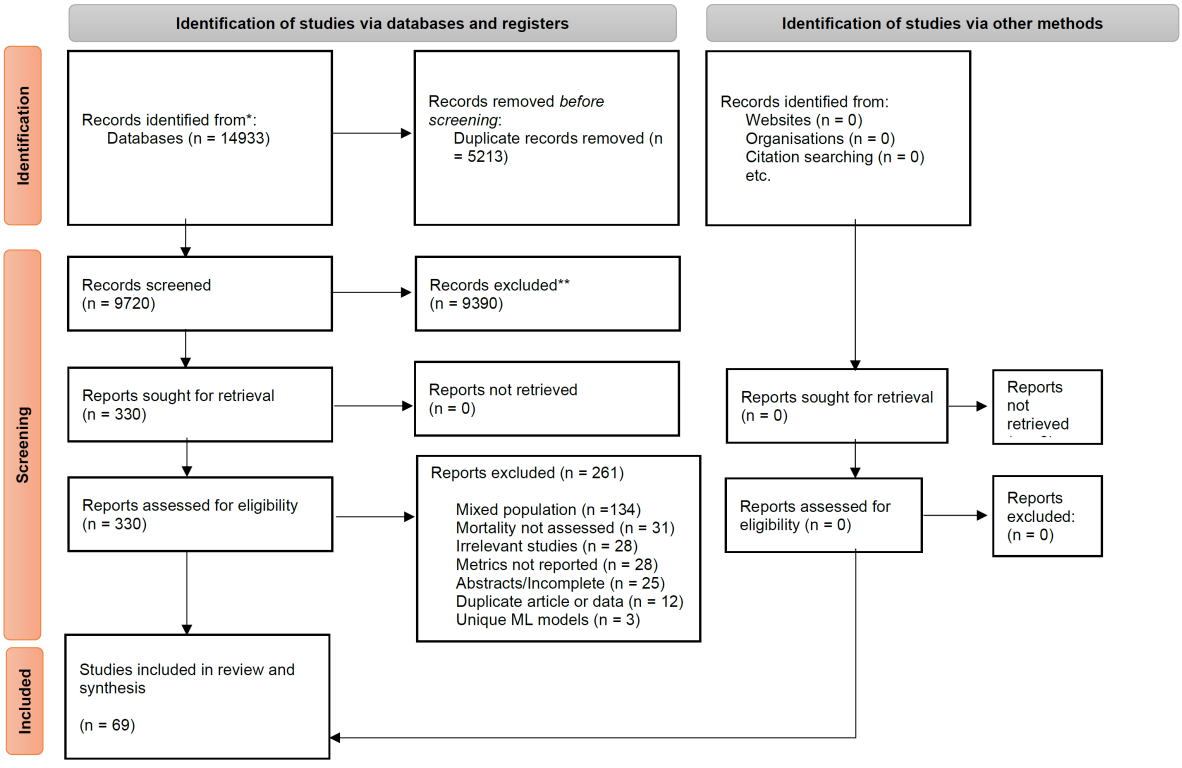

Figure 1: PRISMA flow diagram depicting study selection process. the diagram outlines the screening and selection process for the included studies. From 14,933 initial records, 5,213 duplicates were excluded. After screening titles and abstracts of 9,720 records, 330 full texts were assessed for eligibility, resulting in the inclusion of 69 studies in the final meta-analysis. Reasons for full-text exclusions included mixed populations (n=134), non-mortality outcomes (n=31), irrelevant topics (n=28), insufficient reporting (n=28), incomplete reports (n=25), duplicate data (n=12), and ineligible ML models (n=3). Abbreviations: PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; ML, machine learning.

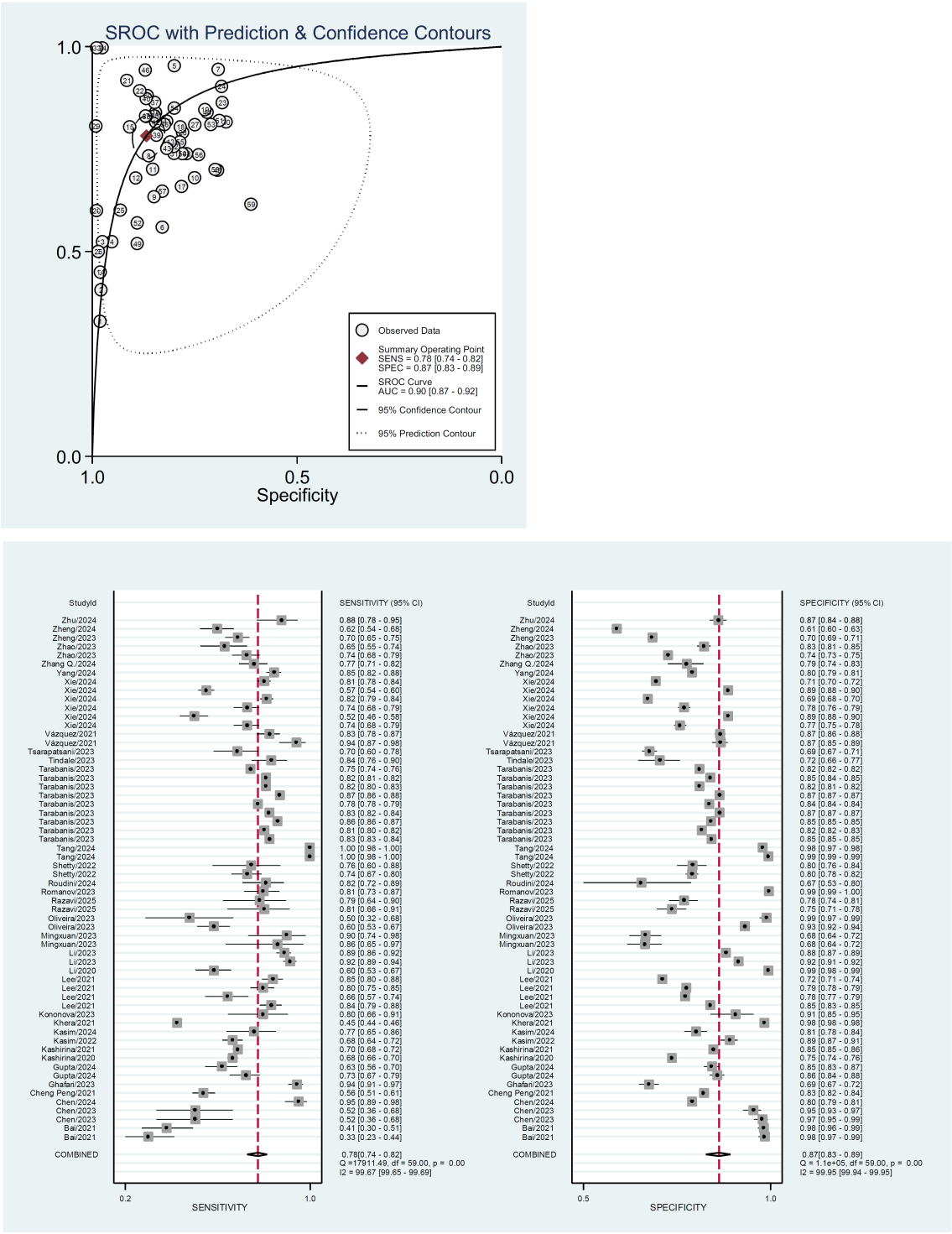

Figure 2: Prognostic accuracy of advanced gradient boosting machines. A: presents the summary receiver operating characteristic (SROC) curve for advanced gradient boosting machines (GBM), including XGBoost, showing pooled sensitivity of 0.78 (95% CI: 0.74–0.82), specificity of 0.87 (95% CI: 0.83–0.89), and an AUC of 0.90 (95% CI: 0.87–0.92). The solid contour represents the 95% confidence region; the dashed contour marks the 95% prediction region. B: Presents the sensitivity and specificity estimates for individual studies and their contribution to the pooled analysis, demonstrating substantial heterogeneity ($I^2 > 90\%$, $p < 0.001$).

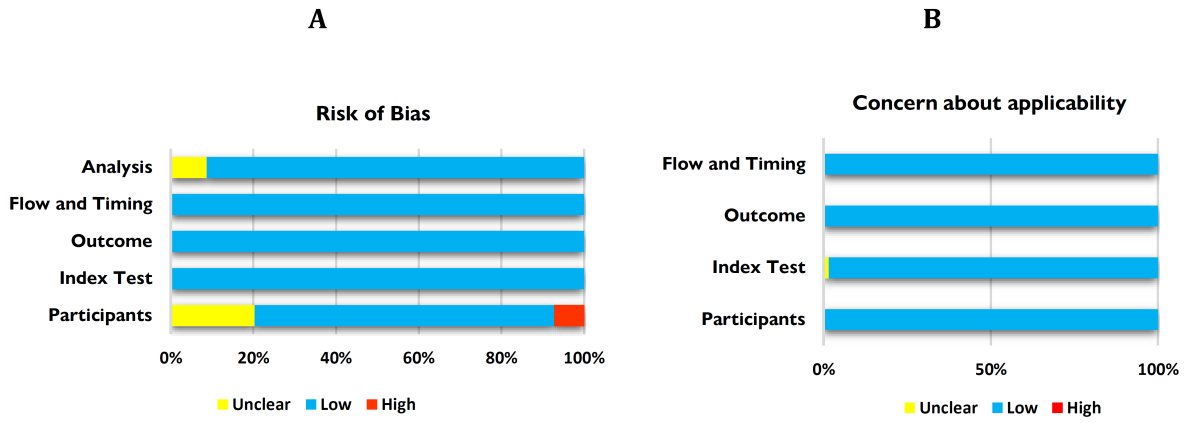

Figure 3: Risk of Bias and Applicability Concerns in Included Studies based on QUAPAS tool. A: shows risk of bias across five domains (Participants, Index Test, Outcome, Flow and Timing, and Analysis), with most studies rated as low risk except for concerns in the Participants domain. B: illustrates the applicability of each study to clinical practice, with minimal concerns observed across domains.