International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Machine Learning Models for Predicting in-Hospital Cardiac Arrest: A Comparative Analysis with Logistic Regression

Wei-Shan Chang[1], Kai-Yuan Hsiao[2], Lian-Yu Lin[3,4], MingChih Chen[2,5], Ben-Chang Shia[2,5], Chung-Yu Lin[5,6]

[1]Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; [2]Artificial Intelligence Development Center, Fu Jen Catholic University, New Taipei City, Taiwan; [3]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; [4]Smart Medicine and Health Informatics Program, National Taiwan University, Taipei, Taiwan; [5]Graduate Institute of Business Administration, Fu Jen Catholic University, New Taipei City, Taiwan; [6]Department of Cardiology, Fu Jen Catholic University Hospital, New Taipei City, Taiwan

Correspondence: Chung-Yu Lin, Department of Cardiology, Fu Jen Catholic University Hospital, No. 69, Guizi Road, Taishan District, New Taipei City, 243, Taiwan, Tel +886-988-322-193, Email ChungyuLin0727@gmail.com

---

Purpose: To develop and compare multiple machine learning (ML) algorithms with traditional logistic regression for predicting in-hospital cardiac arrest (IHCA) using comprehensive electronic health record data, with the goal of improving early risk stratification beyond conventional early-warning scores and providing potential integration into hospital early warning systems for timely clinical intervention.

Patients and Methods: We performed a retrospective case-control study at a large tertiary medical center, including 800 IHCA cases and 3,464 controls. Candidate predictors comprised demographics, comorbidities, vital signs, and laboratory measurements. Five models-logistic regression, decision tree, random forest, XGBoost, and multivariate adaptive regression splines (MARS)-were trained and validated. Performance was evaluated using area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, and F1 score.

Results: XGBoost yielded strong discrimination and the highest accuracy (AUC 0.909; accuracy 0.883), while random forest showed comparable discrimination (AUC 0.910) with slightly lower accuracy (0.876). Logistic regression performed robustly but lower than ML models (AUC 0.895; accuracy 0.876). ML models consistently identified clinically meaningful predictors-including blood urea nitrogen, heart rate, and pre-existing heart failure-offering insights beyond traditional regression.

Conclusion: Integrating ML approaches with conventional regression enhances IHCA risk prediction by capturing non-linear relationships and interactions while retaining the interpretability of regression. These approaches could strengthen hospital early-warning systems, enabling earlier detection and intervention, and ultimately improving patient outcomes.

Keywords: in-hospital cardiac arrest, machine learning, logistic regression, risk prediction, predictive analytics, early warning systems

---

Introduction

In-hospital cardiac arrest (IHCA) remains a frequent and critical event that places a substantial emotional and operational burden on healthcare teams. Once IHCA occurs, the prognosis is poor: more than half of patients do not survive despite resuscitation, and nearly 90% of survivors suffer significant neurological impairment.[1] The sudden onset of IHCA, often following rapid but under-recognized clinical deterioration, makes early detection particularly challenging. This is especially true in general wards, where approximately 72% of IHCAs occur.[2–4] Reported survival rates vary by region, with recent US data indicating a survival-to-discharge rate of about 25.8%,[5,6] whereas a Taiwanese study showed a return of spontaneous circulation (ROSC) in 66% of cases but survival-to-discharge of only 11.8%.[4]

Although IHCA management strategies are often adapted from out-of-hospital cardiac arrest (OHCA) research, important differences exist in epidemiology and underlying pathophysiology.[7] Conventional risk assessment methods

Received: 24 September 2025
Accepted: 11 October 2025
Published: 21 October 2025

typically rely on medical history, trends in vital signs, laboratory values, and procedural data to estimate clinical deterioration or mortality risk.[8] However, relatively few studies have specifically focused on identifying predictors of unexpected IHCA before the event, rather than outcomes after resuscitation.

To improve early recognition, clinical scoring systems such as the National Early Warning Score (NEWS) and the Modified Early Warning Score (MEWS) are widely used, particularly in the United Kingdom.[9] Other early warning systems, such as the Cardiac Arrest Risk Triage (CART) score,[10] have also been implemented in general wards in the United States. These scores depend mainly on vital signs to identify patients at risk of acute deterioration, including cardiac arrest. Their predictive performance, however, is modest, with reported areas under the receiver operating characteristic curve (AUC) ranging from 0.65 to 0.79.[11]

Once the high-risk patient group is identified, high-intensity care should be initiated, such as frequent vital sign monitoring, activation of rapid response teams or ICU admission for the most severe cases. According to a systematic review by Hogan et al, the implementation of the National Early Warning Score (NEWS) in daily practice-accompanied by the use of different algorithms-was associated with a 6.4% annual reduction in in-hospital cardiac arrest (IHCA) incidence and a 5% annual improvement in survival rates.[12]

The widespread adoption of electronic health records and digital healthcare systems has created opportunities for advanced predictive analytics. By leveraging dynamic, longitudinal patient data, predictive models may detect clinical deterioration earlier and with greater accuracy. Prior studies have shown that machine learning (ML) methods-such as random forest, XGBoost, decision trees, and multivariate adaptive regression splines (MARS)-often outperform traditional statistical models in predicting mortality and major cardiovascular events.[13,14] Ensemble ML approaches, which combine multiple algorithms, have demonstrated even stronger accuracy and calibration in clinical applications.[15]

Despite these advances, most existing studies have focused on post-arrest outcomes or on predicting OHCA, leaving a critical gap in pre-arrest risk stratification for IHCA.[16,17] Only a limited number of studies have begun to explore IHCA prediction, primarily by evaluating traditional risk factors with conventional statistical methods.[18,19]

To address this, the present study compares the predictive performance of conventional logistic regression with four ML algorithms-random forest, XGBoost, decision tree, and MARS-for forecasting IHCA among hospitalized patients. By incorporating comprehensive clinical variables, this study aims to enhance early risk stratification and support proactive interventions to reduce IHCA incidence and improve patient outcomes.

Materials and Methods

We conducted a retrospective, single-center, case-control study at National Taiwan University Hospital (NTUH), including adult patients (≥18 years) who experienced unexpected in-hospital cardiac arrest (IHCA) between 2011 and 2018. Eligible patients were required to have at least one documented electrocardiogram (ECG) prior to the IHCA event. The study protocol was approved by the Institutional Review Board of NTUH (IRB No. 201807063RINC). This study was conducted in accordance with the principles of the Declaration of Helsinki. Given the retrospective design and the use of de-identified data, the need for informed consent was waived.

For the control cohort, 4,000 patients were randomly selected from 205,999 hospitalized individuals without CPR events during the study period. Patients with do-not-resuscitate (DNR) orders at admission (n = 65) or with incomplete clinical records (n = 471) were excluded, resulting in 3,464 patients in the non-IHCA group. The selection and exclusion process is shown in Figure 1. Incomplete clinical records were defined as the absence of essential demographic information (eg, age, sex, comorbidities) or more than 30% missing vital sign or laboratory variables. For the remaining dataset, variables with ≤30% missing data were imputed using multiple imputation by chained equations (MICE). The percentage of missing data for each variable is summarized in Table S1.

The primary outcome was IHCA, defined as the absence of a palpable pulse with attempted resuscitation during hospitalization. The dataset included four major domains of variables. Demographic information comprised age, sex, and body mass index (BMI). Comorbidities were identified from medical records and coded using the International Classification of Diseases, Ninth and Tenth Revisions (ICD-9/10-CM). Vital signs included systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), pulse rate, respiratory rate, and body temperature. Laboratory parameters included serum creatinine, serum sodium, serum potassium, hemoglobin, platelet count, aspartate

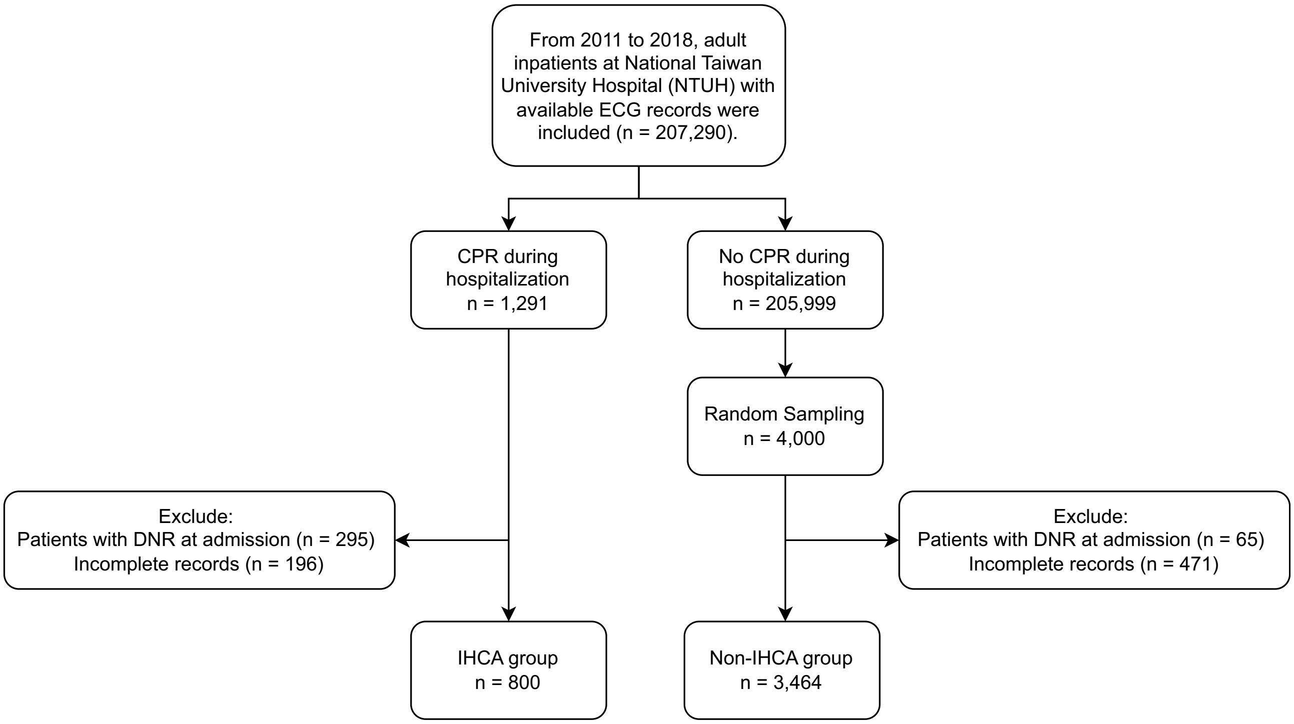

Figure 1 Flow diagram of study population selection. Adult inpatients at NTUH (2011–2018) with documented ECG (n = 207,290) were classified according to in-hospital CPR status. After exclusions, the IHCA group (with in-hospital CPR) comprised 800 patients and the non-IHCA group (without in-hospital CPR) comprised 3,464 patients.
Abbreviations: CPR, cardiopulmonary resuscitation; DNR, do-not-resuscitate; ECG, electrocardiogram; IHCA, in-hospital cardiac arrest; NTUH, National Taiwan University Hospital.

aminotransferase (AST), and alanine aminotransferase (ALT). Diagnoses were coded using the International Classification of Diseases, Ninth and Tenth Revisions (ICD-9-CM/ICD-10-CM), and procedural codes were obtained from Taiwan's National Health Insurance execution code system.

Five predictive models were developed: logistic regression, decision tree, random forest, extreme gradient boosting (XGBoost), and multivariate adaptive regression splines (MARS). Data preprocessing included quality checks and imputation of missing values to ensure integrity. The dataset was randomly divided into training (80%) and testing (20%) subsets. Model training used 10-fold cross-validation for hyperparameter optimization and to minimize overfitting. Figure 2 illustrates the ML analytical workflow used in our study.

Logistic regression was used as a benchmark model for binary classification, estimating the probability of IHCA based on clinical predictors. It remains widely applied in medical research and serves as a reference for comparing the performance of more advanced ML algorithms.

Decision trees are supervised learning models that classify outcomes by sequentially splitting data into subgroups based on predictor variables. Each branch represents a decision rule, and terminal nodes represent predicted outcomes. Their hierarchical, rule-based structure makes them intuitive and interpretable for both technical and clinical applications.

Random forest is an ensemble method that improves the stability and accuracy of decision trees. It generates multiple trees using bootstrap samples with randomized feature selection and aggregates their results by majority voting. Out-of-bag samples are used to estimate generalization error and feature importance, reducing overfitting and enhancing predictive reliability.

XGBoost is an optimized gradient boosting algorithm that combines multiple weak learners, typically decision trees, into a strong predictive model. It incorporates parallel processing, automated handling of missing data, and regularization to reduce overfitting. XGBoost has demonstrated state-of-the-art performance on structured clinical datasets and is widely applied in healthcare risk prediction.

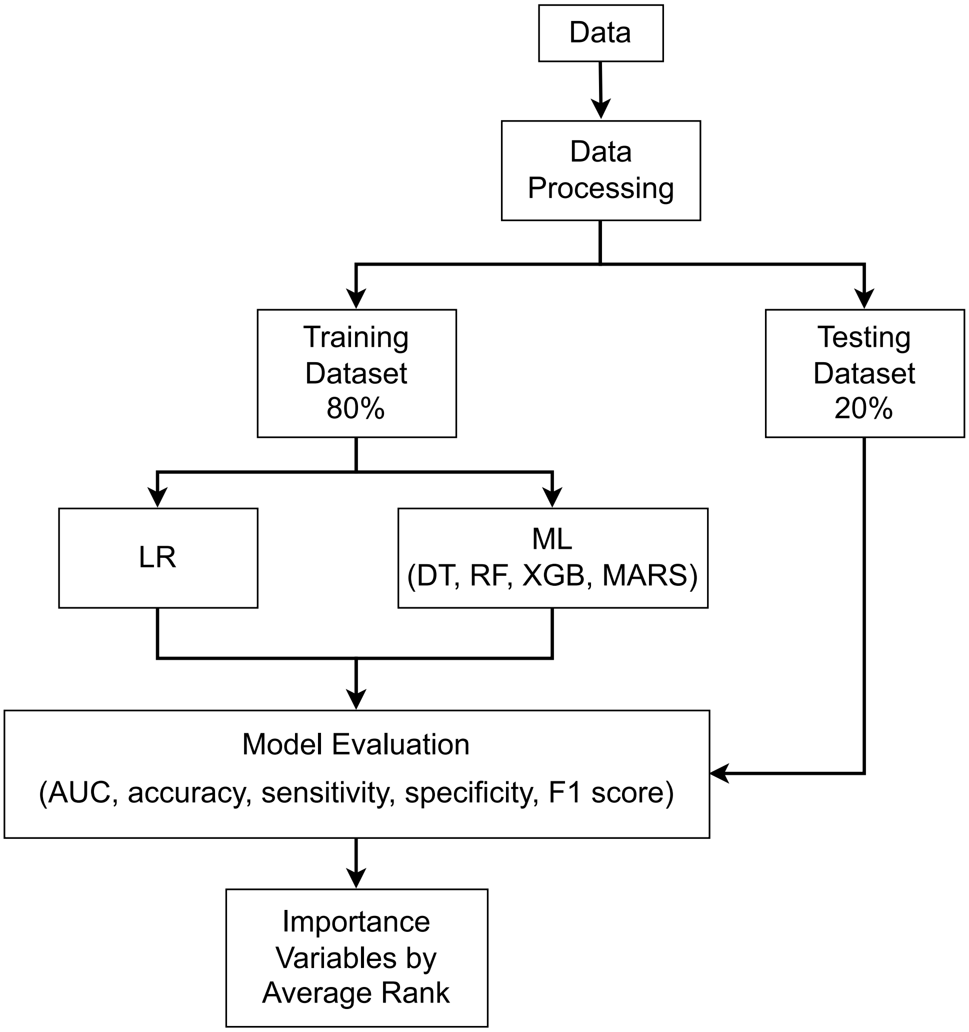

Figure 2 Modeling workflow. Data were processed and split into training and testing datasets. Five algorithms (LR, DT, RF, XGB, MARS) were trained on the training dataset, evaluated on standard metrics (AUC, accuracy, sensitivity, specificity, F1 score), and variable importance was summarized by average rank across models.
Abbreviations: AUC, area under the curve; DT, Decision Tree; LR, Logistic Regression; MARS, Multivariate Adaptive Regression Splines; RF, Random Forest; XGB, Extreme Gradient Boosting.

Multivariate Adaptive Regression Splines (MARS) is a non-linear regression technique that models complex relationships using adaptive spline functions. It builds models through forward selection of candidate basis functions followed by backward elimination to control complexity. This flexibility allows MARS to capture both linear and non-linear effects, making it suitable for identifying subtle patterns in clinical data.

While a concise overview of each model is presented here, detailed algorithmic descriptions and hyperparameter specifications are provided in Supplementary Material 1.

To minimize the impact of potential multicollinearity among predictors (eg, renal markers, ECG intervals), we applied L1 regularization when constructing logistic regression models, which performs variable selection and shrinks the coefficients of less informative or collinear variables. For the machine learning approaches, we primarily employed tree-based models (eg, random forest, XGBoost), which are inherently less sensitive to multicollinearity due to their recursive partitioning mechanisms. Together, these strategies reduced the influence of collinearity and enhanced the robustness of our analyses.

Model performance was evaluated using standard classification metrics. Accuracy was defined as the proportion of correct predictions among all cases. Sensitivity (recall, true positive rate) represented the proportion of actual positives correctly identified, whereas specificity (true negative rate) represented the proportion of actual negatives correctly identified. Positive predictive value (PPV, precision) indicated the proportion of predicted positives that were truly positive, and negative predictive value (NPV) indicated the proportion of predicted negatives that were truly negative. The F1 score, calculated as the harmonic mean of precision and recall, provides a single measure balancing false positives and false negatives, as shown in Equation (1). Finally, the AUC summarized overall discrimination across all decision thresholds, reflecting the probability that a randomly selected positive case would be ranked higher than a randomly selected negative case (0.5 = no discrimination; 1.0 = perfect discrimination).

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{1}$$

Equation (1). Formula for calculating the F1 score.

All analyses were performed using R software (version 4.0.3) within RStudio (version 1.4.1103), with dedicated R packages supporting each ML algorithm. Logistic regression was implemented using the glmnet package (version 4.1–1), decision trees with the rpart package (version 4.1–15), random forests with the randomForest package (version 4.6–14), and XGBoost with the xgboost package (version 1.5.0.1). MARS was conducted using the earth package (version 5.3.2). The caret package (version 6.0–90) was used for model training, hyperparameter tuning, and the evaluation of variable importance across methods.

An advanced language model (ChatGPT 5, OpenAI, San Francisco, CA, USA) was employed to enhance the grammar, phrasing, and readability of the manuscript. The model did not contribute to scientific content, data analyses, or interpretation. All generated text was thoroughly examined and edited by the authors, who assume full responsibility for the accuracy and conclusions of the manuscript.

Results

As summarized in Table 1, a total of 800 patients with IHCA and 3,464 randomly selected hospitalized controls were analyzed. Compared with controls, the IHCA group was significantly older (64.6 ± 15.9 vs 57.0 ± 16.6 years, $p < 0.001$), had a slightly higher proportion of males (60.4% vs 56.5%, $p = 0.048$), and a lower mean body mass index (23.6 ± 5.0 vs 24.3 ± 4.2 kg/m², $p < 0.001$).

Cardiovascular comorbidities were markedly more prevalent in the IHCA group, including heart failure (43.2% vs 7.7%), acute coronary syndrome (ACS) (23.8% vs 3.0%), chronic coronary syndrome (42.8% vs 16.7%), peripheral artery disease (13.9% vs 4.2%), and hypertension (59.2% vs 41.1%) (all $p < 0.001$). Non-cardiovascular conditions such as diabetes mellitus (41.2% vs 20.5%), chronic kidney disease (32.9% vs 10.2%), and end-stage renal disease (20.4% vs 5.3%) were also more frequent (all $p < 0.001$). In contrast, malignancy was less common among IHCA patients (43.0% vs 50.9%, $p < 0.001$), although both groups demonstrated a high prevalence of malignancy.

Table 1 Comparison of Baseline Characteristics Between IHCA and Non-IHCA Groups

Mean (SD)	**Non-IHCA Control Group**	**IHCA Experimental Group**	**p**
n	3464	800	
Male (%)	1956 (56.5)	483 (60.4)	0.048
Age	57.00 (16.64)	64.64 (15.92)	<0.001
BMI	24.33 (4.21)	23.56 (4.99)	<0.001
HF (%)	266 (7.7)	346 (43.2)	<0.001
ACS (%)	103 (3.0)	190 (23.8)	<0.001
CCS (%)	579 (16.7)	342 (42.8)	<0.001

(*Continued*)

Table 1 (Continued).

Mean (SD)	Non-IHCA Control Group	IHCA Experimental Group	p
PAD (%)	144 (4.2)	111 (13.9)	<0.001
HTN (%)	1425 (41.1)	474 (59.2)	<0.001
DM (%)	709 (20.5)	330 (41.2)	<0.001
Dyslipidemia (%)	714 (20.6)	230 (28.7)	<0.001
COPD / Asthma (%)	298 (8.6)	114 (14.2)	<0.001
CKD (%)	353 (10.2)	263 (32.9)	<0.001
ESRD (%)	182 (5.3)	163 (20.4)	<0.001
Malignancy (%)	1763 (50.9)	344 (43.0)	<0.001
GI Bleeding (%)	340 (9.8)	119 (14.9)	<0.001
WBC 10^3/uL	7.29 (4.03)	11.63 (18.10)	<0.001
Hb g/dL	13.06 (2.13)	10.98 (2.55)	<0.001
PLT 10^9 /L	239.95 (85.96)	198.61 (115.89)	<0.001
BUN mg/dL	17.83 (13.53)	37.80 (31.37)	<0.001
Cre mg/dL	1.08 (1.34)	2.31 (2.61)	<0.001
Na mmol/L	138.63 (3.47)	136.14 (6.52)	<0.001
K mmol/L	4.19 (0.50)	4.17 (0.85)	0.344
ECG PR ms	151.11 (41.41)	126.72 (74.29)	<0.001
ECG QRS ms	90.38 (15.53)	100.30 (29.58)	<0.001
ECG QTc ms	430.78 (32.09)	470.76 (57.82)	<0.001
Body temperature (Celsius degree)	36.40 (0.55)	36.46 (0.79)	0.006
Heart rate / min	79.67 (14.81)	92.85 (21.09)	<0.001
Respiratory rate / min	18.36 (2.14)	20.24 (4.50)	<0.001
SBP mmHg	130.23 (20.51)	127.19 (24.08)	<0.001
DBP mmHg	77.22 (13.01)	72.33 (15.10)	<0.001
MBP mmHg	94.43 (13.94)	90.10 (15.86)	<0.001

Abbreviations: ACS, acute coronary syndrome; BMI, body mass index; BT, body temperature; BUN, blood urea nitrogen; CCS, chronic coronary syndrome; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; Cre, creatinine; DBP, diastolic blood pressure; DM, diabetes mellitus; ECG, electrocardiogram; ECG PR, PR interval on ECG; ECG QRS, QRS duration on ECG; ECG QTc, corrected QT interval on ECG; ESRD, end-stage renal disease; GI, gastrointestinal; Hb, hemoglobin; HF, heart failure; HR, heart rate; HTN, hypertension; K, serum potassium; MBP, mean blood pressure; Na, sodium; PAD, peripheral arterial disease; PLT, platelet count; RR, respiratory rate; SBP, systolic blood pressure; WBC, white blood cell.

Laboratory findings indicated greater systemic inflammation and renal dysfunction in IHCA patients, with significantly higher white blood cell counts (11.63 vs 7.29 × $10^3/\mu L$), blood urea nitrogen (BUN) (37.8 vs 17.8 mg/dL), and creatinine (2.31 vs 1.08 mg/dL) (all $p < 0.001$). However, liver function markers such as AST and ALT were not further analyzed because a high proportion of missing data was detected. This was likely due to local clinical practice patterns, where physicians often order only one of these tests rather than both, partly influenced by insurance-related considerations. IHCA patients also exhibited more pronounced anemia (hemoglobin 11.0 vs 13.1 g/dL) and thrombocytopenia (198.6 vs 239.9 × $10^3/\mu L$) (both $p < 0.001$). Serum potassium did not differ significantly. Electrocardiographic intervals were consistently prolonged, with longer ECG PR interval (151 vs 127 ms), ECG QRS duration (100 vs 90 ms), and corrected QT interval on ECG (471 vs 431 ms) (all $p < 0.001$).

Vital sign comparisons revealed higher pulse rates (92.9 vs 79.7 bpm, $p < 0.001$) and respiratory rates (20.2 vs 18.4 breaths/min, $p < 0.001$) among IHCA patients. Blood pressure values were slightly lower, including systolic (127.2 vs 130.2 mmHg, $p < 0.001$), diastolic (72.3 vs 77.2 mmHg, $p < 0.001$), and mean blood pressure (90.1 vs 94.4 mmHg, $p < 0.001$). Body temperature was minimally higher (36.46 vs 36.40°C, $p = 0.006$). These findings collectively indicated a profile of advanced comorbidity burden, systemic inflammation, renal dysfunction, anemia, and hemodynamic compromise in the IHCA group.

As shown in Table 2, model discrimination ranged from moderate to excellent (AUC 0.739–0.910). The decision tree performed weakest overall, with an AUC of 0.739, sensitivity of 0.331, and the lowest F1 score of 0.450, despite excellent specificity (0.965). By comparison, ensemble approaches achieved superior discrimination. Random forest yielded the highest AUC (0.910) and the strongest positive predictive value (0.749), but this improvement in precision was accompanied by reduced sensitivity (0.544). XGBoost provided the most balanced performance, with an AUC of 0.909, accuracy of 0.883, sensitivity of 0.615, specificity of 0.949, NPV of 0.914, and F1 score of 0.675, representing the highest sensitivity among all models while maintaining excellent overall accuracy. MARS also showed consistent performance across metrics (AUC 0.897; accuracy 0.881; sensitivity 0.580; specificity 0.952; F1 score 0.667), highlighting its stability and calibration.

Logistic regression, although a conventional statistical approach, remained competitive. It achieved an AUC of 0.895 and accuracy of 0.876, with PPV 0.724 and NPV 0.907. However, sensitivity was only moderate (0.580). Overall, these results indicate that ensemble machine learning methods (XGBoost and random forest) outperformed single decision trees and conventional regression in terms of discriminatory power. XGBoost was the only model to achieve both high sensitivity and strong overall accuracy, while MARS provided well-balanced performance with interpretable nonlinear modeling.

Variable importance rankings are summarized in Table 3. Despite differences in methodology, there was strong convergence across models on several key predictors. Logistic regression prioritized hemoglobin, pulse rate, ACS, heart failure, and platelet count. In contrast, the machine learning models consistently ranked BUN and corrected QT interval on ECG among the top predictors, followed by hemoglobin, heart failure, and pulse rate.

When averaged across all five models, the top predictors were BUN, corrected QT interval on ECG, hemoglobin, heart failure, pulse rate, platelet count, ACS, white blood cell count, respiratory rate, and serum sodium. These features represented multiple domains: renal dysfunction and metabolic derangement (BUN, serum creatinine, serum sodium), chronic cardiovascular comorbidities (heart failure, ACS), hematologic impairment (hemoglobin, platelet count), systemic stress and inflammation (pulse rate, respiratory rate, white blood cell count), and electrophysiological abnormalities (corrected QT interval on ECG, ECG QRS duration).

Table 2 Performance of the LR, Decision Tree, Random Forest, XGBoost and MARS Methods

Method	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score	AUC
Logistic Regression	0.876	0.58	0.949	0.724	0.907	0.649	0.895
Decision Tree	0.839	0.331	0.965	0.686	0.862	0.45	0.739
Random Forest	0.876	0.544	0.958	0.749	0.901	0.634	0.91
XGBoost	0.883	0.615	0.949	0.736	0.914	0.675	0.909
MARS	0.881	0.58	0.952	0.736	0.908	0.667	0.897

Abbreviations: AUC, area under the curve; F1, F1 score; LR, logistic regression; MARS, multivariate adaptive regression splines; NPV, negative predictive value; PPV, positive predictive value; XGBoost, extreme gradient boosting.

Table 3 Comparative Variable Importance Rankings and Average Ranks Across Five Predictive Models

Variable	Logistic Regression	Decision Tree	Random Forest	XGB	MARS	Average Rank
BUN	6	1	1	1	1	2
ECG QTc	7	3	2	3	2	3.4
Hb	1	4	3	4	9	4.2
HF	4	2	9	2	5	4.4
Heart rate	2	8	4	5	4	4.6
PLT	5	7	6	7	6	6.2
ACS	3	6	20	6	3	7.6
WBC	9	9	5	8	8	7.8
Respiratory rate	10	10	10	10	11	10.2
Na	8	24	8	9	7	11.2
BMI	22	13	14	11	12	14.4
SBP	15	29	16	15	10	17
Cre	23	5	7	21	31	17.4
CCS	11	14	21	19	31	19.2
ECG QRS	12	27	13	17	31	20
ECG PR	20	26	11	14	31	20.4
AGE	31	12	17	12	31	20.6
DM	13	17	22	20	31	20.6
K	29	25	12	13	31	22
DBP	16	30	15	22	31	22.8

Abbreviations: ACS, acute coronary syndrome; AGE, age; BMI, body mass index; BUN, blood urea nitrogen; CCS, chronic coronary syndrome; Cre, serum creatinine; DBP, diastolic blood pressure; DM, diabetes mellitus; ECG, electrocardiogram; Hb, hemoglobin; HF, heart failure; K, serum potassium; MARS, multivariate adaptive regression splines; Na, serum sodium; PLT, platelet count; PR, PR interval on ECG; QRS, QRS duration on ECG; ECG QTc, corrected QT interval on ECG; SBP, systolic blood pressure; WBC, white blood cell count; XGB, Extreme Gradient Boosting (XGBoost).

The decision tree model presented in Figure 3 further demonstrates how a limited set of key predictors can effectively stratify IHCA risk. For example, pathways incorporating thresholds for BUN (<27 mg/dL), pulse rate, and heart failure status effectively separated patients into high- and low-risk subgroups with minimal computational steps. This simplified structure underscored the consistency of these variables across different modeling approaches.

Together, these results demonstrate that IHCA was associated with a multifactorial risk profile characterized by advanced age, cardiovascular comorbidities, renal dysfunction, hematologic abnormalities, and electrophysiological instability. Among the predictive models, ensemble machine learning approaches, particularly XGBoost and random forest, provided the highest discriminatory power, whereas MARS delivered stable and well-balanced performance. Logistic regression, although less powerful, remained a robust and interpretable benchmark. The convergence of predictors across methods highlights the reliability of these findings and supports the integration of both acute physiological variables and chronic disease burden into early risk stratification frameworks.

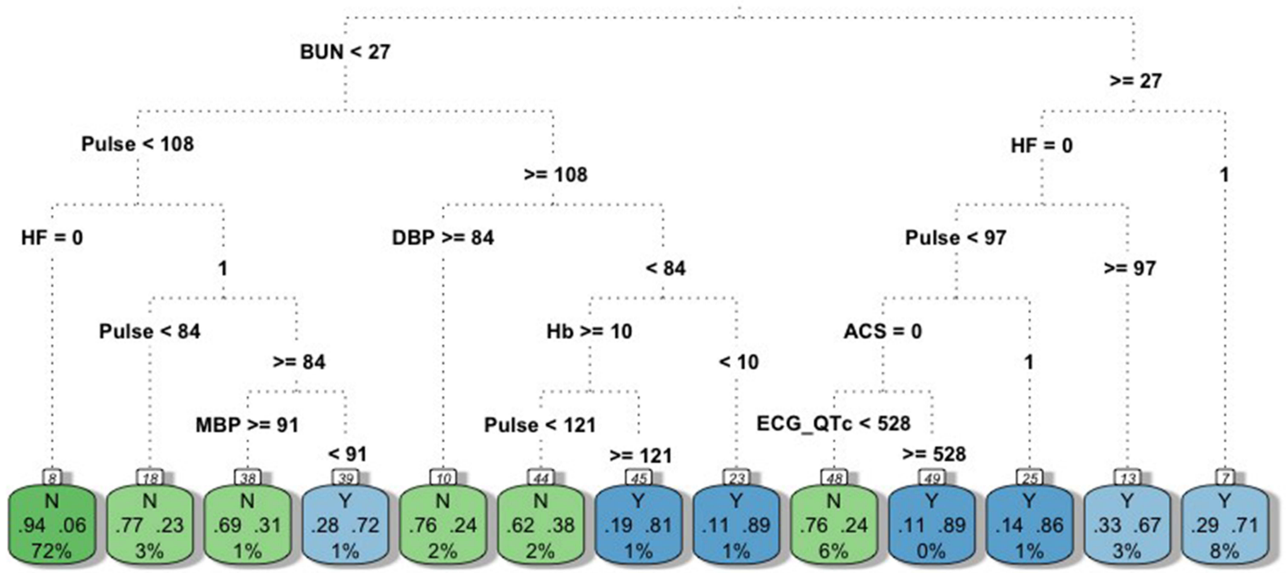

Figure 3 Decision tree model for IHCA prediction. The model stratified IHCA risk using key variables including BUN, HF, pulse rate, DBP, Hb, ACS, and ECG QTc, with terminal nodes showing predicted probabilities.

Abbreviations: ACS, acute coronary syndrome; BUN, blood urea nitrogen; DBP, diastolic blood pressure; ECG QTc, corrected QT interval on ECG; Hb, hemoglobin; HF, heart failure; MBP, mean blood pressure.

Discussion

In this single-center, retrospective case–control study based on NTUH electronic health records, we developed and validated machine-learning models for predicting IHCA. To ensure comparability with the general inpatient population rather than a high-acuity subgroup at imminent risk of IHCA, random sampling was adopted for the control cohort. This strategy enabled us to construct a prediction model representative of routine hospitalized patients and to assess its performance in that context. Notably, malignancy was less common in the IHCA group-a paradoxical finding that may be explained by the higher prevalence of DNR orders among terminal cancer patients, thereby reducing their likelihood of unexpected IHCA.[20]

Our findings highlight that combining traditional statistical approaches with modern ML methods provides complementary strengths in risk prediction. Logistic regression identified established clinical predictors, whereas ensemble models such as random forest and XGBoost achieved superior overall performance. These results underscore the value of integrating conventional regression with advanced ML in clinical prognostication.[21]

Feature importance analysis revealed complementary strengths. Logistic regression prioritized established predictors such as hemoglobin, pulse rate, ACS, heart failure, and platelet count, consistent with traditional cardiovascular frameworks.[5–7] In contrast, ML models consistently ranked BUN and corrected QT interval on ECG among the top variables, reflecting their ability to capture nonlinear relationships and complex interactions often overlooked by conventional approaches.[22,23] Together, these predictors, including BUN, corrected QT interval on ECG, hemoglobin, ACS, heart failure, platelet count, and inflammatory markers, illustrate the multifactorial nature of IHCA risk and underscore the value of integrating both chronic comorbidities and acute stressors into predictive models.[24,25]

In this study, we adopted random sampling to construct the control group. This approach allowed us to better represent the heterogeneity of the general inpatient population and to identify the subgroup truly at risk of IHCA who might benefit from early intervention. In contrast, propensity score matching, while effective in reducing baseline imbalances, would restrict the analysis to patients already similar to the IHCA cohort based on predefined risk factors. Such restriction could limit generalizability and potentially overlook the broader at-risk population that our prediction models aim to capture.[26]

Previous studies applying ML to IHCA prediction have reported AUCs of 0.80–0.93,[22,23,27] which are comparable to our results. One study demonstrated that gradient boosting outperformed logistic regression in emergency patients,[23] while another identified laboratory markers such as platelet count and serum sodium as powerful predictors,[27] aligning with our findings. Other investigations highlighted the predictive value of ECG-derived features such as corrected QT interval on ECG,[28–30] which was also confirmed in our analysis.

A conceptual strength of ML is its ability to move beyond binary "normal/abnormal" thresholds traditionally used in clinical medicine.[31–33] Logistic regression and conventional models depend on predefined cutoffs (eg, serum sodium <135 mmol/L) which may obscure risk gradients within reference ranges.[34] In contrast, ML derives optimal cut points directly from data. In our decision tree, BUN at 27 mg/dL emerged as a critical threshold for IHCA risk, despite lying near the conventional upper limit of normal. Similar data-driven thresholds were identified for hemoglobin (10 g/dL) and pulse rate (84 or 121 bpm). Such findings illustrate how ML can uncover hidden nonlinear risk profiles, as demonstrated in sepsis,[35,36] ACS,[37] and arrhythmia prediction.[27,30] For example, in Figure 3, the decision tree identified a diastolic blood pressure (DBP) threshold of 84 mmHg, which is not a commonly used clinical cut-off in daily practice. Nevertheless, prior studies have demonstrated that DBP is indeed an independent predictor of cardiac arrest, albeit with different threshold values.[38,39] This finding underscores the potential of ML models to uncover clinically relevant yet unconventional patterns that may be overlooked by traditional approaches. While such thresholds may not immediately translate into bedside decision rules, they highlight physiological parameters that warrant closer monitoring and further validation in prospective studies.

Beyond IHCA, ML models have been widely used for disease prediction across medicine. Decision trees are simple and transparent but often lack sensitivity in high-risk settings.[39] Random forest, by combining multiple trees, improves stability and has shown strong performance in predicting sepsis, ACS, and heart failure.[40] XGBoost, an advanced gradient boosting method, consistently outperforms other algorithms in structured healthcare datasets by capturing complex nonlinear relationships with high efficiency.[41] Although less commonly used, MARS provides flexibility in modeling both linear and nonlinear effects. A previous study demonstrated its predictive value by developing a model for summed stress score in Taiwanese women with type 2 diabetes mellitus using the MARS approach.[42]

Comparative studies confirm that ensemble methods, particularly random forest and XGBoost, provide the best overall accuracy and calibration, while decision trees and MARS contribute interpretability in selected scenarios.[40–42] Our findings echo prior evidence of XGBoost's superiority and further support the robustness of ML models across diverse patient populations and healthcare systems. Importantly, when integrated into electronic health records, ML-based prediction tools could be embedded within hospital early warning systems to deliver real-time alerts and facilitate timely clinical intervention.[14]

A key challenge for implementing ML in clinical practice is interpretability, as advanced models often act as "black boxes" compared with the transparency of logistic regression.[32] In addition, successful adoption requires seamless integration into electronic health record systems, with real-time outputs that are clinically actionable.[43] Overcoming these barriers will be crucial for translating predictive accuracy into meaningful patient outcomes.

We believe our study makes two main contributions. First, we systematically compared the performance of multiple machine learning models against traditional logistic regression, highlighting their relative strengths in predicting IHCA. Second, by applying multiple predictive tools, we were able to identify novel risk factors that are not typically captured by conventional approaches, and to establish an early warning framework that may help deliver intensive care to high-risk patients and thereby reduce mortality.

This study has several limitations. First, its retrospective, single-center design precludes causal inference and may limit generalizability. Second, we adopted random sampling rather than propensity score matching to ensure representativeness of the general inpatient population. This approach introduced baseline imbalances, but machine learning methods, with their ability to model multicollinearity and interactions, may have mitigated some of these differences. Third, only internal validation was performed; external, multicenter validation is needed to confirm robustness. Fourth, certain relevant variables (eg, echocardiography, Holter monitoring, imaging) were unavailable, which may influence risk assessment. Finally, as a pilot study, future research should incorporate multimodal data and prospective designs, ideally comparing model predictions with physicians' real-time judgment, to establish clinical utility.

Conclusion

In this study, we directly compared logistic regression with multiple machine learning models for predicting in-hospital cardiac arrest. While logistic regression provided interpretability, advanced models-particularly XGBoost and random forest-achieved superior discrimination and calibration. Key predictors consistently included BUN, corrected QT interval, and hemoglobin. These results suggest that ML-based tools can enhance early risk stratification beyond conventional approaches, and their integration into hospital electronic health records and early warning systems may facilitate earlier recognition and timely intervention. Prospective multicenter validation will be essential to confirm these findings and determine their clinical impact.

Acknowledgments

The authors sincerely appreciate the data resources made available through the Integrated Medical Database of National Taiwan University Hospital, as well as the kind support offered by its staff. We are also indebted to the Artificial Intelligence Development Center at Fu Jen Catholic University, New Taipei City, Taiwan, for their valuable technical assistance.

This paper was previously uploaded to ResearchGate as a preprint [https://www.researchgate.net/publication/395063593_Comparative_Performance_of_Machine_Learning_Algorithms_and_Logistic_Regression_for_Predicting_In-Hospital_Cardiac_Arrest_Preprint]. It was initially submitted to *JMIR Cardio* but was formally withdrawn prior to its current submission.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Liu C-T, Lai C-Y, Wang J-C, Chung C-H, Chien W-C, Tsai C-S. A population-based retrospective analysis of post-in-hospital cardiac arrest survival after modification of the chain of survival. *J Emerg Med*. 2020;59(2):246–253. doi:10.1016/j.jemermed.2020.04.045
2. Peberdy MA, Kaye W, Ornato JP, et al. Cardiopulmonary resuscitation of adults in the hospital: a report of 14720 cardiac arrests from the national registry of cardiopulmonary resuscitation. *Resuscitation*. 2003;58(3):297–308. doi:10.1016/s0300-9572(03)00215-6
3. Merchant RM, Yang L, Becker LB, et al. Incidence of treated cardiac arrest in hospitalized patients in the United States. *Crit Care Med*. 2011;39(11):2401–2406. doi:10.1097/CCM.0b013e3182257459
4. Wang CH, Tay J, Wu CY, et al. External validation and comparison of statistical and machine learning-based models in predicting outcomes following out-of-hospital cardiac arrest: a multicenter retrospective analysis. *J Am Heart Assoc*. 2024;13(20):e037088. doi:10.1161/JAHA.124.037088
5. Girotra S, Nallamothu BK, Spertus JA, et al. Trends in survival after in-hospital cardiac arrest. *N Engl J Med*. 2012;367(20):1912–1920. doi:10.1056/NEJMoa1109148
6. Nolan JP, Soar J, Smith GB, et al. National cardiac arrest audit. Incidence and outcome of in-hospital cardiac arrest in the United Kingdom national cardiac arrest audit. *Resuscitation*. 2014;85(8):987–992. doi:10.1016/j.resuscitation.2014.04.002
7. Guan G, Lee CMY, Begg S, Crombie A, Mnatzaganian G. The use of early warning system scores in prehospital and emergency department settings to predict clinical deterioration: a systematic review and meta-analysis. *PLoS One*. 2022;17(3):e0265559. doi:10.1371/journal.pone.0265559
8. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation*. 2013;84(4):465–470. doi:10.1016/j.resuscitation.2012.12.016
9. Smith ME, Chiovaro JC, O'Neil M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. *Ann Am Thorac Soc*. 2014;11(9):1454–1465. doi:10.1513/AnnalsATS.201403-102OC
10. Churpek MM, Yuen TC, Edelson DP. Risk stratification of hospitalized patients on the wards. *Chest*. 2013;143(6):1758–1765. doi:10.1378/chest.12-1605
11. Badriyah T, Briggs JS, Meredith P, et al. Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS). *Resuscitation*. 2014;85(3):418–423. doi:10.1016/j.resuscitation.2013.12.011
12. Hogan H, Hutchings A, Wulff J, et al. *Interventions to Reduce Mortality from in-Hospital Cardiac Arrest: A Mixed-Methods Study*. Southampton (UK): NIHR Journals Library; January 2019.
13. Shafiq M, Mazzotti DR, Gibson C. Risk stratification of patients who present with chest pain and have normal troponins using a machine learning model. *World J Cardiol*. 2022;14(11):565–575. doi:10.4330/wjc.v14.i11.565
14. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018;1:18. doi:10.1038/s41746-018-0029-1
15. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12(4):e0174944. doi:10.1371/journal.pone.0174944
16. Chen CT, Chiu PC, Tang CY, et al. Prognostic factors for survival outcome after in-hospital cardiac arrest: an observational study of the oriental population in Taiwan. *J Chin Med Assoc*. 2016;79(1):11–16. doi:10.1016/j.jcma.2015.07.011
17. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. In-hospital cardiac arrest: a review. *JAMA*. 2019;321(12):1200–1210. doi:10.1001/jama.2019.1696
18. Fernando SM, Tran A, Cheng W, et al. Pre-arrest and intra-arrest prognostic factors associated with survival after in-hospital cardiac arrest: systematic review and meta-analysis. *BMJ*. 2019:367:l6373. doi:10.1136/bmj.l6373

19. Mitsunaga T, Hasegawa I, Uzura M, et al. Comparison of the National Early Warning Score (NEWS) and the Modified Early Warning Score (MEWS) for predicting admission and in-hospital mortality in elderly patients in the prehospital setting and in the emergency department. *PeerJ*. 2019;7(e6947). doi:10.7717/peerj.6947
20. Giza DE, Graham J, Donisan T, et al. Impact of cardiopulmonary resuscitation on survival in cancer patients: do not resuscitate before or after CPR? *JACC CardioOncol*. 2020;2(2):359–362. doi:10.1016/j.jaccao.2020.03.003
21. Holmstrom L, Bednarski B, Chugh H, et al. Artificial intelligence model predicts sudden cardiac arrest manifesting with pulseless electric activity versus ventricular fibrillation. *Circ Arrhythm Electrophysiol*. 2024;17(2):e012338. doi:10.1161/CIRCEP.123.012338
22. Kwon JM, Kim KH, Jeon KH, Lee SY, Park J, Oh BH. Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography. *Scand J Trauma Resusc Emerg Med*. 2020;28(1):98. doi:10.1186/s13049-020-00791-0
23. Lu TC, Wang CH, Chou FY, et al. Machine learning to predict in-hospital cardiac arrest from patients presenting to the emergency department. *Intern Emerg Med*. 2023;18(2):595–605. doi:10.1007/s11739-022-03143-1
24. Weng SF, Vaz L, Qureshi N, Kai J. Prediction of premature all-cause mortality: a prospective general population cohort study comparing machine learning and standard epidemiological approaches. *PLoS One*. 2019;14(3):e0214365. doi:10.1371/journal.pone.0214365
25. Li H, Wu TT, Yang DL, et al. Decision tree model for predicting in-hospital cardiac arrest among patients admitted with acute coronary syndrome. *Clin Cardiol*. 2019;42(11):1087–1093. doi:10.1002/clc.23255
26. Stürmer T, Wyss R, Glynn RJ, Brookhart MA. Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs. *J Intern Med*. 2014;275(6):570–580. doi:10.1111/joim.12197
27. Ding X, Wang Y, Ma W, et al. Development of early prediction model of in-hospital cardiac arrest based on laboratory parameters. *Biomed Eng Online*. 2023;22(1):116. doi:10.1186/s12938-023-01178-9
28. Do DH, Kuo A, Lee ES, et al. Usefulness of trends in continuous electrocardiographic telemetry monitoring to predict in-hospital cardiac arrest. *Am J Cardiol*. 2019;124(7):1149–1158. doi:10.1016/j.amjcard.2019.06.032
29. Straus SM, Kors JA, De Bruin ML, et al. Prolonged QTc interval and risk of sudden cardiac death in a population of older adults. *J Am Coll Cardiol*. 2006;47(2):362–367. doi:10.1016/j.jacc.2005.08.067
30. Al-Khatib SM, LaPointe NM, Kramer JM, Califf RM. What clinicians should know about the QT interval. *JAMA*. 2003;289(16):2120–2127. doi:10.1001/jama.289.16.2120
31. Matsushita K, Ballew SH, Wang AY, et al. Epidemiology and risk of cardiovascular disease in populations with chronic kidney disease. *Nat Rev Nephrol*. 2022;18(11):696–707. doi:10.1038/s41581-022-00616-6
32. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–1358. doi:10.1056/NEJMra1814259
33. Misra D, Avula V, Wolk DM, et al. Early detection of septic shock onset using interpretable machine learners. *J Clin Med*. 2021;10(2):301. doi:10.3390/jcm10020301
34. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206–215. doi:10.1038/s42256-019-0048-x
35. Jin D, Jin S, Liu B, et al. Association between serum sodium and in-hospital mortality among critically ill patients with spontaneous subarachnoid hemorrhage. *Front Neurol*. 2022;13:1025808. doi:10.3389/fneur.2022.1025808
36. Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA*. 2019;321(20):2003–2017. doi:10.1001/jama.2019.5791
37. VanHouten JP, Starmer JM, Lorenzi NM, Maron DJ, Lasko TA. Machine learning for risk prediction of acute coronary syndrome. *AMIA Annu Symp Proc*. 2014;2014:1940–1949.
38. Aziz S, Barratt J, Starr Z, et al. The association between intra-arrest arterial blood pressure and return of spontaneous circulation in out-of-hospital cardiac arrest. *Resuscitation*. 2024;205:110426. doi:10.1016/j.resuscitation.2024.110426
39. Zelic I, Kononenko I, Lavrac N, Vuga V. Induction of decision trees and Bayesian classification applied to diagnosis of sport injuries. *J Med Syst*. 1997;21(6):429–444. doi:10.1023/A:1022880431298
40. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J*. 2017;38(23):1805–1814. doi:10.1093/eurheartj/ehw302
41. Wang Z, Gu Y, Huang L, et al. Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data. *Cardiovasc Diabetol*. 2024;23(1):351. doi:10.1186/s12933-024-02439-0
42. Yuan CH, Lee PC, Wu ST, Yang CC, Chu TW, Yeih DF. Using multivariate adaptive regression splines to estimate summed stress score on myocardial perfusion scintigraphy in Chinese women with type 2 diabetes: a comparative study with multiple linear regression. *Diagnostics*. 2025;15(17):2270. doi:10.3390/diagnostics15172270
43. Hofer IS, Burns M, Kendale S, Wanderer JP. Realistically integrating machine learning into clinical practice: a road map of opportunities, challenges, and a potential future. *Anesth Analg*. 2020;130(5):1115–1118. doi:10.1213/ANE.0000000000004575