Comparative Performance of Machine Learning Algorithms and Logistic Regression for Predicting In-Hospital Cardiac Arrest

Wei-Shan Chang, Kai-Yuan Hsiao, Lian-Yu Lin, MingChih Chen, Ben-Chang Shia, Chung-Yu Lin

Submitted to: JMIR Cardio
on: August 28, 2025

Table of Contents

Original Manuscript **5**
Supplementary Files **25**
Figures **26**
Figure 1 ..... **27**
Figure 2 ..... **28**
Figure 3 ..... **29**

Comparative Performance of Machine Learning Algorithms and Logistic Regression for Predicting In-Hospital Cardiac Arrest

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

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

Corresponding Author:
Chung-Yu Lin

Department of Cardiology
Fu Jen Catholic University Hospital
No. 69, Guizi Road, Taishan District
New Taipei City
TW

Abstract

Background: In-hospital cardiac arrest (IHCA) remains a catastrophic event with persistently low survival, even amid advances in resuscitation and critical care. Conventional early warning scores provide only limited predictive accuracy, often failing to identify patients at highest risk. To overcome these limitations, we evaluated a multimodal prediction framework that integrates traditional logistic regression with advanced machine learning (ML) approaches to optimize risk stratification.

Objective: This study aimed to develop and compare the performance of multiple machine learning algorithms against traditional logistic regression for predicting IHCA using a comprehensive set of electronic health record (EHR) data.

Methods: We conducted a retrospective case–control study including 800 IHCA cases and 3,464 matched controls from a large tertiary medical center. Candidate predictors comprised demographics, comorbidities, laboratory values, and vital signs. Five models—logistic regression, decision tree, random forest, XGBoost, and multivariate adaptive regression splines (MARS)—were trained and validated. Model performance was assessed using accuracy, sensitivity, specificity, F1 score, and area under the receiver operating characteristic curve.

Results: Among all models, XGBoost demonstrated the best predictive performance (AUC 0.909; accuracy 0.883), followed by random forest (AUC 0.910; accuracy 0.876). Logistic regression achieved robust but comparatively lower performance (AUC 0.895; accuracy 0.876). Importantly, ML models highlighted clinically relevant predictors—such as blood urea nitrogen, heart rate, and presence of heart failure—providing novel insights beyond traditional regression.

Conclusions: Integrating ML-based approaches with conventional regression substantially improved IHCA risk prediction. While logistic regression offers transparency and interpretability, ML models capture complex, non-linear interactions that enhance accuracy. This multimodal framework holds potential to strengthen hospital early warning systems, enabling earlier detection, timely intervention, and ultimately improved patient outcomes. Clinical Trial: The study protocol was approved by the Institutional Review Board of National Taiwan University Hospital (IRB No. 201807063RINC). Due to the retrospective nature of this study, which involved the analysis of pre-existing data, trial registration was not required.

(JMIR Preprints 28/08/2025:83120)
DOI: https://doi.org/10.2196/preprints.83120

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Comparative Performance of Machine Learning Algorithms and Logistic Regression for Predicting In-Hospital Cardiac Arrest

Author
Wei-Shan Chang, PhD[1]; ORCID ID: 0009-0004-7196-9633
Kai-Yuan Hsiao, PhD[2]; ORCID ID: 0009-0002-3611-2007
Lian-Yu Lin MD, PhD[3,4]; ORCID ID: 0000-0001-7505-6429
MingChih Chen, PhD[2,5]; ORCID ID: 0000-0002-8278-0033
Ben-Chang Shia, PhD[2,5]; ORCID ID: 0000-0003-2854-8361
*Chung-Yu Lin, MD[5,6]; ORCID ID: 0009-0003-1209-3603 - **Corresponding Author**
Email: a02076@mail.fjuh.fju.edu.tw
Cell phone: +886988322193
Adress: No. 179, Minquan West Road, 3F-3, Taipei 103, Taiwan

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

Key word: in-hospital cardiac arrest, machine learning, logistic regression, risk prediction, XGBoost, decision tree, random forest, multivariate adaptive regression splines

Abstract

Background:

In-hospital cardiac arrest (IHCA) remains a catastrophic event with persistently low survival, even amid advances in resuscitation and critical care. Conventional early warning scores provide only

limited predictive accuracy, often failing to identify patients at highest risk. To overcome these limitations, we evaluated a multimodal prediction framework that integrates traditional logistic regression with advanced machine learning (ML) approaches to optimize risk stratification.

Objective:

This study aimed to develop and compare the performance of multiple machine learning algorithms against traditional logistic regression for predicting IHCA using a comprehensive set of electronic health record (EHR) data.

Methods:

We conducted a retrospective case–control study including 800 IHCA cases and 3,464 matched controls from a large tertiary medical center. Candidate predictors comprised demographics, comorbidities, laboratory values, and vital signs. Five models—logistic regression, decision tree, random forest, XGBoost, and multivariate adaptive regression splines (MARS)—were trained and validated. Model performance was assessed using accuracy, sensitivity, specificity, F1 score, and area under the receiver operating characteristic curve.

Results:

Among all models, XGBoost demonstrated the best predictive performance (AUC 0.909; accuracy 0.883), followed by random forest (AUC 0.910; accuracy 0.876). Logistic regression achieved robust but comparatively lower performance (AUC 0.895; accuracy 0.876). Importantly, ML models highlighted clinically relevant predictors—such as blood urea nitrogen, heart rate, and presence of heart failure—providing novel insights beyond traditional regression.

Conclusions:

Integrating ML-based approaches with conventional regression substantially improved IHCA risk prediction. While logistic regression offers transparency and interpretability, ML models capture complex, non-linear interactions that enhance accuracy. This multimodal framework holds potential to strengthen hospital early warning systems, enabling earlier detection, timely intervention, and ultimately improved patient outcomes.

1. 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 [2,3]. Reported survival rates vary by region, with recent U.S. data indicating a survival-to-discharge rate of about 25.8% [4,5], whereas a Taiwanese study showed a return of spontaneous circulation (ROSC) in 66% of cases but survival-to-discharge of only 11.8% [6].

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 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 [9,10]. 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].

The widespread adoption of electronic health records (EHRs) 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 [12–14]. 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 out-of-hospital cardiac arrest (OHCA), leaving a critical gap in pre-arrest risk stratification for IHCA. Only a limited number of studies have begun to explore IHCA prediction, primarily by evaluating traditional risk factors with conventional statistical methods [16–20].

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.

2. Method

2.1 Study Design and Participants

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 cases required at least one electrocardiogram (ECG) performed prior to the IHCA event.

A total of 800 patients meeting these criteria were included in the IHCA group. For the control group, 3,464 patients were randomly selected from 205,999 hospitalized individuals during the same period. Patients were excluded if they were younger than 18 years, pregnant, had a documented do-not-resuscitate (DNR) order, or lacked complete laboratory data, including complete blood count (WBC, hemoglobin, platelet count) and basic chemistry tests (blood urea nitrogen, creatinine, sodium, and potassium). Patients could be included more than once if eligibility criteria were met on separate admissions.

The primary outcome was IHCA, defined as the absence of a palpable pulse with attempted resuscitation during hospitalization. Data collected included demographics (age, sex, body mass index [BMI]), comorbidities, vital signs (systolic blood pressure [SBP], diastolic blood pressure [DBP], mean arterial pressure [MAP], heart rate [HR], respiratory rate [RR], and body temperature), and laboratory results (e.g., creatinine, sodium, potassium, hemoglobin, platelet count, aspartate aminotransferase [AST], 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. The study protocol was approved by National Taiwan University Hospital's Institutional Review Board (IRB No. 201807063RINC).

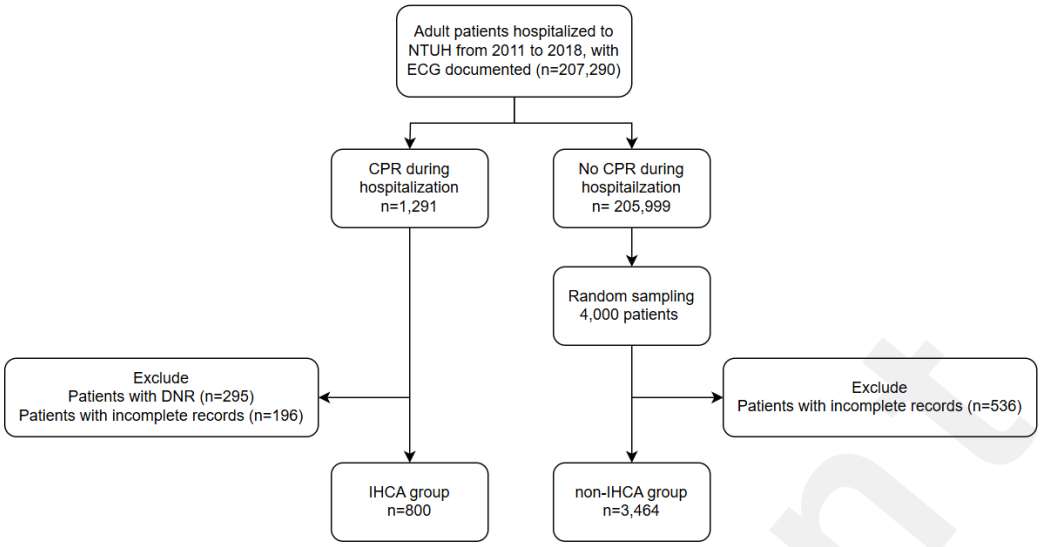


Figure 1. A flow diagram of patient selection and exclusion.

Abbreviations:

CPR, cardiopulmonary resuscitation; DNR, do-not-resuscitate; ECG, electrocardiogram; IHCA, in-hospital cardiac arrest; NTUH, National Taiwan University Hospital.

2.2 Model development and statistical analyses

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.

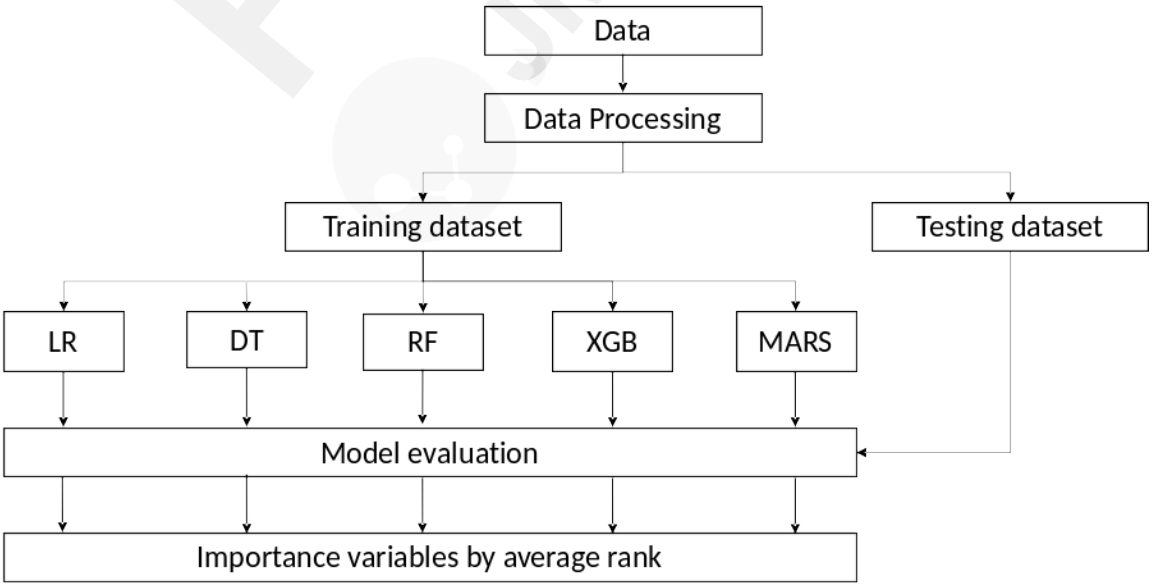


Figure 2. Machine learning (ML) analytical workflow in our study.

Abbreviations: DT: Decision Tree; LR: Logistic Regression; MARS: Multivariate Adaptive Regression Splines RF: Random Forest; XGB: Extreme Gradient Boosting;

2.2.1 Logistic regression

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 machine learning (ML) algorithms.

2.2.2 Decision Tree

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.

2.2.3 Random Forest

2.2.4 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.

2.2.5 XGBOOST

Extreme Gradient Boosting (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.

2.2.6 MARS

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.

2.2.7 Machine Learning Workflow

Model performance was assessed using accuracy, sensitivity, specificity, precision, F1 score, and area under the receiver operating characteristic curve (AUC). Feature importance was calculated separately for each algorithm and integrated using the *caret* package to provide a model-agnostic ranking of predictors.

All analyses were performed using R software (version 4.0.3) within RStudio (version 1.4.1103), with dedicated R packages supporting each machine learning 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). Multivariate Adaptive Regression Splines (MARS) were 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.

3 Result

A total of 800 patients with in-hospital cardiac arrest (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 (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%, $p < 0.001$), reflecting the high baseline cancer prevalence in this tertiary referral population.

Laboratory findings indicated greater systemic inflammation and renal dysfunction in IHCA

patients, with significantly higher white blood cell counts (11.63 vs. 7.29 ×10³/μL), blood urea nitrogen (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 ×10³/μL) (both $p < 0.001$). Serum potassium did not differ significantly. Electrocardiographic intervals were consistently prolonged, with longer PR (151 vs. 127 ms), QRS (100 vs. 90 ms), and QTc (471 vs. 431 ms) intervals (all $p < 0.001$).

IIn terms of vital signs, IHCA patients demonstrated higher heart rates (92.9 vs. 79.7 bpm) and respiratory rates (20.2 vs. 18.4 breaths/min) (both $p < 0.001$), whereas systolic and diastolic blood pressures were modestly lower (127.2 vs. 130.2 mmHg, and 72.3 vs. 77.2 mmHg, respectively; both $p < 0.001$). Mean arterial pressure did not differ significantly between groups.

Table 1 Characteristics of the experimental groups (IHCA) and the control groups(non-IHCA).

	Non-IHCA Control group	IHCA Experimental group	p
n	3464	800	
Male (%)	1956 (56.5)	483 (60.4)	0.048
Age (mean (SD))	57.00 (16.64)	64.64 (15.92)	<0.001
BMI (mean (SD))	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
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 (mean (SD)) 10^3/uL	7.29 (4.03)	11.63 (18.10)	<0.001
Hb (mean (SD)) g/dL	13.06 (2.13)	10.98 (2.55)	<0.001
PLT (mean (SD)) 10^9 /L	239.95 (85.96)	198.61 (115.89)	<0.001
BUN (mean (SD)) mg/dL	17.83 (13.53)	37.80 (31.37)	<0.001
Cre (mean (SD)) mg/dL	1.08 (1.34)	2.31 (2.61)	<0.001
Na (mean (SD)) mmol/L	138.63 (3.47)	136.14 (6.52)	<0.001
K (mean (SD)) mmol/L	4.19 (0.50)	4.17 (0.85)	0.344
ECG_PR (mean (SD)) ms	151.11 (41.41)	126.72 (74.29)	<0.001

ECG_QRS (mean (SD)) ms	90.38 (15.53)	100.30 (29.58)	<0.001
ECG_QTc (mean (SD)) ms	430.78 (32.09)	470.76 (57.82)	<0.001
Body temperature (mean (SD)) (Celsius degree)	36.40 (0.55)	36.46 (0.79)	0.006
Pulse rate (mean (SD)) / min	79.67 (14.81)	92.85 (21.09)	<0.001
Respiratory rate (mean (SD)) / min	18.36 (2.14)	20.24 (4.50)	<0.001
SBP (mean (SD)) mmHg	130.23 (20.51)	127.19 (24.08)	<0.001
DBP (mean (SD)) mmHg	77.22 (13.01)	72.33 (15.10)	<0.001
MBP (mean (SD)) 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-PR, PR interval; ECG-QRS, QRS duration; ECG-QTc, corrected QT interval; ESRD, end-stage renal disease; GI, gastrointestinal; Hb, hemoglobin; HF, heart failure; HR, heart rate; HTN, hypertension; K, 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.

Across predictive models, discrimination ranged from moderate to excellent (AUC 0.739–0.910). XGBoost achieved the highest overall accuracy (0.883) and F1 score (0.675), with balanced sensitivity (0.615) and specificity (0.949), providing reliable performance across risk categories. Random forest produced the highest AUC (0.910) and the best PPV (0.749), reflecting strong precision in detecting true IHCA cases, though its lower sensitivity (0.544) suggests a higher likelihood of missed events. MARS demonstrated consistently balanced metrics (accuracy 0.881; sensitivity 0.580; specificity 0.952; AUC 0.897), ranking second in F1 score (0.667) and providing stable predictive performance across measures.

Logistic regression, while a conventional statistical approach, remained competitive with more advanced algorithms. It achieved an AUC of 0.895, accuracy of 0.876, PPV of 0.724, and NPV of 0.907. Although its sensitivity (0.580) was only moderate, logistic regression offered strong interpretability and clinical relevance, supporting its role as a practical benchmark model.

By contrast, the decision tree showed the weakest performance overall (AUC 0.739; sensitivity 0.331), despite excellent specificity (0.965) and acceptable PPV (0.686). This pattern highlights its tendency to miss a substantial proportion of IHCA cases, limiting its utility in high-stakes clinical prediction where sensitivity is critical.

From a clinical perspective, models emphasizing higher sensitivity and balanced predictive values—such as XGBoost and MARS—appear more suitable for early IHCA risk stratification, whereas models prioritizing specificity, such as random forest, may serve as confirmatory tools to reduce false alarms. Logistic regression, though less powerful, provides valuable interpretability and

remains useful for clinical translation and comparison.

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.

Feature importance analysis (Table 3) revealed both overlapping and distinct patterns between logistic regression and the machine learning models. In the logistic regression model, the leading predictors were hemoglobin, pulse rate, acute coronary syndrome (ACS), heart failure, and platelet count. These reflect well-established cardiovascular risk factors and hemodynamic disturbances, highlighting the ability of logistic regression to capture linear associations between traditional clinical variables and IHCA risk.

By contrast, the machine learning models (decision tree, random forest, XGBoost, and MARS) consistently identified blood urea nitrogen (BUN) and corrected QT interval (QTc) as the most influential predictors, frequently ranking them first or second. When integrating results across all models, the most consistently high-ranking variables included BUN, QTc interval, hemoglobin, heart failure, pulse rate, platelet count, ACS, white blood cell (WBC) count, respiratory rate, and serum sodium.

Together, these predictors represent a multidimensional risk profile that encompasses chronic cardiovascular disease (heart failure, ACS), acute physiological stress (pulse rate, respiratory rate, platelet count), systemic inflammation (WBC count), renal and metabolic dysfunction (BUN, serum sodium), and electrical instability (QTc prolongation). The convergence of these features across models underscores the multifactorial nature of IHCA risk and supports the need to integrate both acute and chronic variables into predictive frameworks for earlier and more accurate identification of high-risk patients.

Table 3. Comparative variable importance rankings and average ranks across five predictive models

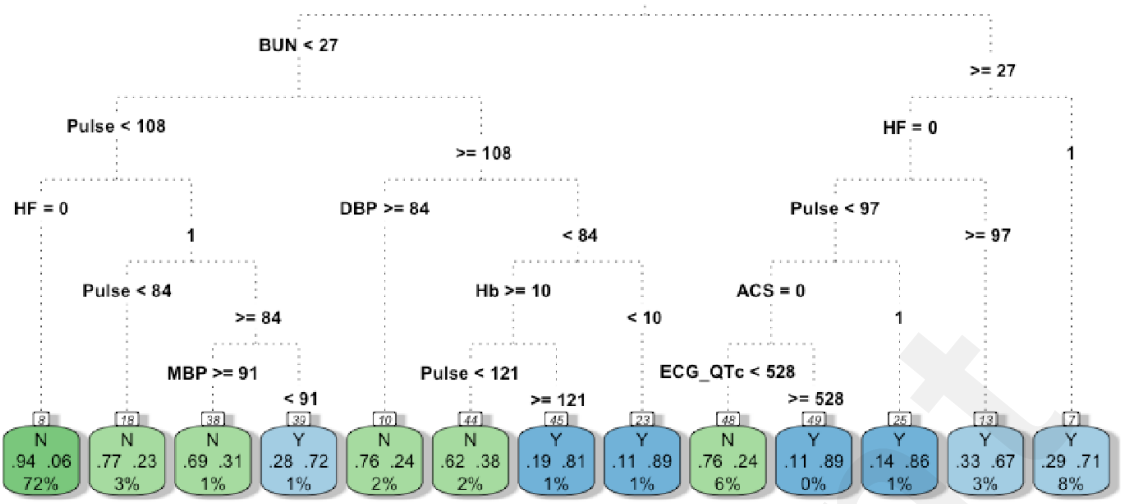
Variable	Logistic regression	Decision tree	Random forest	XGB	MARS	Average Rank

BUN	6	1	1	1	1	2
ECG_QTc interval	7	3	2	3	2	3.4
Hb	1	4	3	4	9	4.2
HF	4	2	9	2	5	4.4
Pulse 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 duration	12	27	13	17	31	20
ECG_PR interval	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
Dyslipidemia	14	18	28	24	31	23
MBP	19	31	18	16	31	23
SEX	28	11	26	23	31	23.8
ESRD	17	21	27	25	31	24.2
Body temperature	26	28	19	18	31	24.4
PAD	21	15	30	26	31	24.6
COPD/ Asthma	18	19	29	28	31	25
HTN	30	16	24	27	31	25.6
CKD	27	20	23	29	31	26
Malignancy	24	22	25	30	31	26.4
GI Bleeding	25	23	31	31	31	28.2

Abbreviations:

ACS, acute coronary syndrome; BMI, body mass index; 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; ESRD, end-stage renal disease; GI, gastrointestinal; Hb, hemoglobin; HF, heart failure; HTN, hypertension; K, potassium; MARS, multivariate adaptive regression splines; MBP, mean blood pressure; Na, sodium; PAD, peripheral arterial disease; PLT, platelet count; PR, PR interval on electrocardiogram; QTc, corrected QT interval; SBP, systolic blood pressure; SEX, biological sex; WBC, white blood cell count; XGB, extreme gradient boosting.

Picture 2 algorithm of decision tree



Abbreviations:

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

In addition to variable importance rankings, the decision tree model (Figure 2) illustrates how a small subset of variables—such as BUN, pulse rate, heart failure status, diastolic blood pressure, hemoglobin, ACS, and QTc interval—can be sequentially applied to stratify IHCA risk. For instance, a pathway beginning with BUN < 27, followed by pulse rate and heart failure status, effectively distinguished low- from high-risk patients with minimal computational steps.

Compared with logistic regression, which relies on simultaneous modeling of multiple variables and interpretation of regression coefficients, decision trees offer intuitive, rule-based pathways that are easily interpretable and directly applicable at the bedside. This transparency, combined with reasonable predictive performance, suggests that such models may be particularly valuable for real-time clinical decision support, especially in resource-limited settings or when rapid risk stratification is required. By operationalizing a handful of clinical and laboratory parameters into actionable rules, decision trees highlight the potential for machine learning–driven tools to complement or, in some scenarios, serve as practical alternatives to traditional statistical models.

4 Discussion

In this large-scale, nationwide study using data from the Taiwan National Health Insurance Research Database, we developed and validated machine learning–based prediction models for IHCA. Our findings highlight that combining traditional statistical approaches with modern ML methods provides complementary strengths in risk prediction. Logistic regression identified

established clinical predictors, while ensemble models such as random forests and XGBoost delivered superior overall performance. These results reinforce 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 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, QTc, 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.

Previous studies applying ML to IHCA prediction have achieved AUCs of 0.80–0.93 [28–30], similar to our results. Lu et al. [23] found that gradient boosting outperformed logistic regression in emergency patients, while Ding et al. [27] identified laboratory markers such as platelet count and sodium as powerful predictors, consistent with our findings. Do et al. [28] and Kwon et al. [22] emphasized the predictive value of ECG-derived features such as QTc, 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 [20,33-35]. Logistic regression and conventional models depend on predefined cutoffs (e.g., sodium <135 mmol/L, QTc >440 ms in men), which may obscure risk gradients within reference ranges. 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 [36], acute coronary syndromes [37], and arrhythmia prediction [27,30]. For example, in Picture 2, 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 machine learning 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, machine learning models have been widely used for disease prediction across medicine. Decision trees are simple and transparent but often lack sensitivity in high-risk settings [40]. Random forest, by combining multiple trees, improves stability and has shown strong performance in predicting sepsis, acute coronary syndromes, and heart failure [41]. XGBoost, an advanced gradient boosting method, consistently outperforms other algorithms in structured healthcare datasets by capturing complex nonlinear relationships with high efficiency [42]. Although less frequently applied, MARS offers flexibility for modeling both linear and nonlinear effects [43]. 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 [41–43]. Our findings echo prior evidence of XGBoost's superiority and further support the robustness of ML models across diverse patient populations and healthcare systems.

5 Limitation

This study has several limitations. First, its retrospective case–control and cross-sectional design precludes causal inference, and the single-center cohort from a Taiwanese tertiary hospital may limit generalizability to other populations. Second, to preserve the heterogeneity of a real-world inpatient population, propensity score matching was not applied; instead, random sampling was used for the control group, which maintained representativeness but introduced baseline imbalances between groups. Nevertheless, the relatively large sample size provided adequate statistical power. Third, only internal validation was performed, and external validation in independent, multicenter cohorts is essential to confirm the robustness and applicability of the models. Fourth, the dataset lacked important clinical variables such as echocardiographic findings, Holter monitoring, and ventricular premature complex origins, which may further influence IHCA risk assessment and the indications or outcomes of ablation. As this work serves as a pilot study, future research incorporating more granular, multimodal data and prospective designs—ideally comparing model predictions with physicians' real-time judgment—will be necessary to fully establish clinical utility and impact on patient care.

6 Conclusion

In this study, we compared traditional logistic regression with several ML models for predicting

IHCA. Logistic regression remained useful and easy to interpret, but advanced ML methods—especially XGBoost and random forest—showed superior accuracy. Across models, blood urea nitrogen, QTc interval, and hemoglobin consistently emerged as the strongest predictors, reflecting the complex and multifactorial nature of IHCA. Importantly, ML models identified new risk thresholds beyond conventional “normal ranges,” offering practical insights for earlier detection. These findings suggest that ML-based prediction tools, when integrated with hospital electronic records, could help clinicians recognize high-risk patients sooner and intervene in time. Future multicenter prospective studies are needed to confirm these results and to evaluate how such models may improve patient outcomes in real-world settings.

Acknowledgements

The authors gratefully acknowledge the data support from the Integrated Medical Database of National Taiwan University Hospital and the assistance provided by National Taiwan University Hospital. We also thank the Artificial Intelligence Development Center, Fu Jen Catholic University, New Taipei City, Taiwan, for their technical support. Two large language models, ChatGPT (version 5.0) and Gemini Pro (version 2.5), were utilized to refine the English grammar of this manuscript.

Funding

The authors received no funding for this research.

Conflicts of Interest

The authors declare no competing interests.

Data Availability

The data that support the findings of this study are not publicly available due to ethical restrictions and the protection of patient privacy. However, access to a de-identified dataset may be granted by the corresponding author upon reasonable request and in accordance with the data-sharing policies of National Taiwan University Hospital. The study protocol was approved by National Taiwan University Hospital's Institutional Review Board (IRB No. 201807063RINC).

Reference

1. Liu CT, Lai CY, Wang JC, Chung CH, Chien WC, Tsai CS. 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.

2. Peberdy MA, Kaye W, Ornato JP, et al. Cardiopulmonary resuscitation of adults in the hospital: a report of 14,720 cardiac arrests from the National Registry of Cardiopulmonary Resuscitation. *Resuscitation*. 2003;58(3):297-308.
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.
4. Lin YR, Chien YC, Wang CH, et al. Predicting outcomes after in-hospital cardiac arrest in Taiwan. *Am J Emerg Med*. 2018;36(5):825-829.
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.
6. Nolan JP, Soar J, Smith GB, et al. Incidence and outcome of in-hospital cardiac arrest in the United Kingdom National Cardiac Arrest Audit. *Resuscitation*. 2014;85(8):987-992.
7. Hravnak M, DeVita MA, Clontz A, et al. Cardiac arrest risk triage in a cardiac specialty hospital. *Crit Care Med*. 2007;35(1):81-86.
8. Royal College of Physicians. *National Early Warning Score (NEWS): Standardising the Assessment of Acute-Illness Severity in the NHS*. RCP; 2012.
9. Subbe CP, Kruger M, Rutherford P, et al. Validation of a modified Early Warning Score in medical admissions. *QJM*. 2001;94(10):521-526.
10. 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.
11. Smith GB, Prytherch DR, Meredith P, et al. 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.
12. Metsker O, Syachina I, Kovalchuk SV, et al. Predictive modeling of mortality risk in electronic health records: A review of the state-of-the-art. *Proc 10th Int Conf Mach Learn Comput*. 2018:376-382.
13. Chen Y, Liu C, Huang Y, et al. Machine learning for risk stratification of patients with chest pain in the emergency department. *J Am Coll Cardiol*. 2020;75(11 Suppl 1):81.
14. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning for electronic health records. *NPJ Digit Med*. 2018;1:18.
15. Weng SF, Reps J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction

using routine clinical data? *PLoS One*. 2017;12(4):e0174944.

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.
17. Andersen LW, Holmberg MJ, Berg KM, et al. In-hospital cardiac arrest: a review. *JAMA*. 2019;321(12):1200-1210.
18. Andersen LW, Granfeldt A, Callaway CW, 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.
19. Mitsunaga T, Hagiwara Y, Oshima K, 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. *PeerJ*. 2019;7:e6947.
20. Steele AJ, Denaxas SC, Shah AD, et al. Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. *PLoS One*. 2018;13(8):e0202344.
21. Holmstrom L, Bednarski B, Chugh H, et al. Artificial intelligence model predicts sudden cardiac arrest manifesting with pulseless electrical activity versus ventricular fibrillation. *Circ Arrhythm Electrophysiol*. 2024;17(2):e012338.
22. Kwon JM, Kim KH, Jeon KH, et al. Artificial intelligence algorithm for predicting cardiac arrest using electrocardiography. *Scand J Trauma Resusc Emerg Med*. 2020;28(1):98.
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.
24. Weng SF, Vaz L, Qureshi N, et al. 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.
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.
26. Soloperto R, Magni F, Farinella A, et al. A comparison of prognostic factors in a large cohort of in-hospital and out-of-hospital cardiac arrest patients. *Life (Basel)*. 2024;14(3):403.
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.
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.

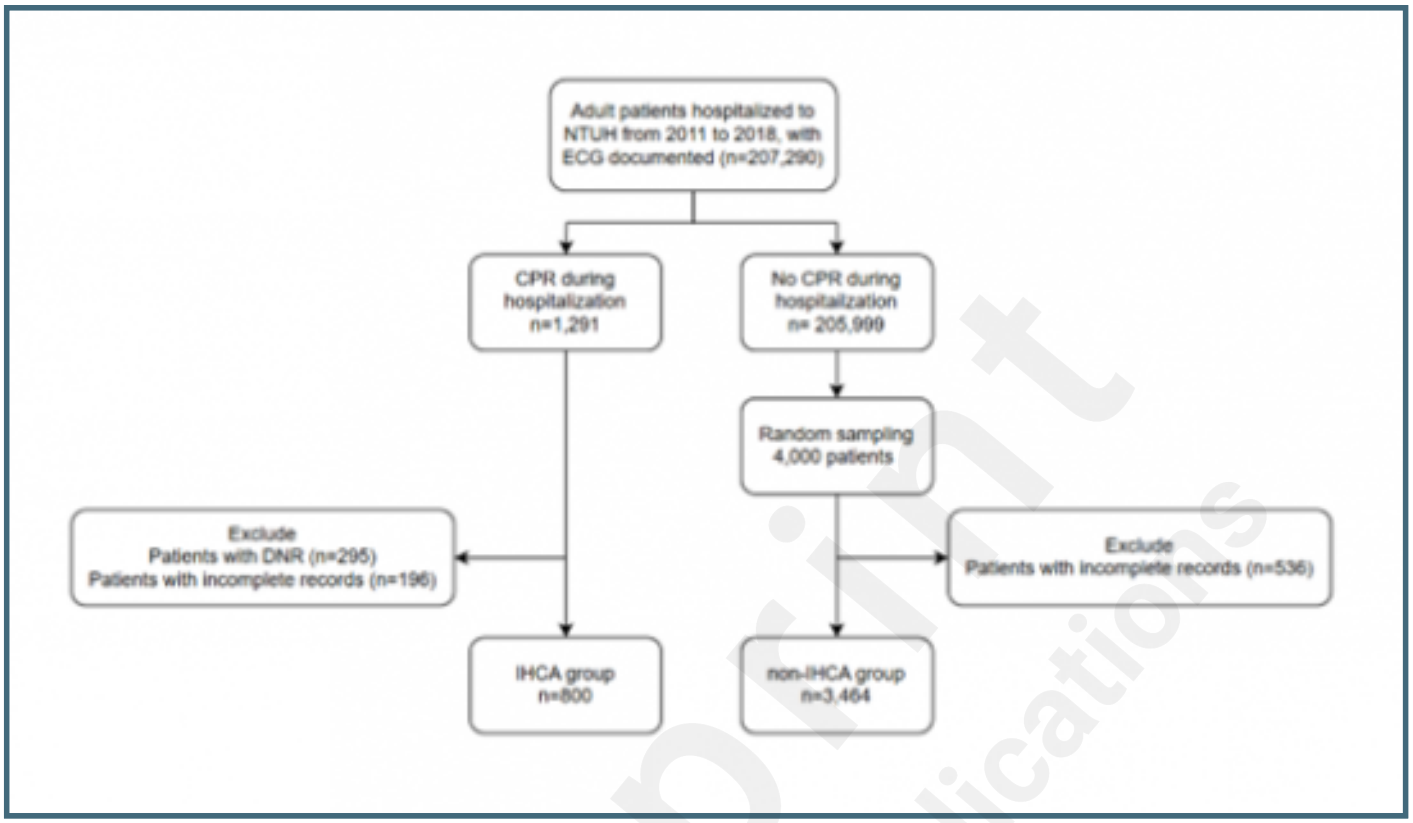
29. Straus SM, Kors JA, De Bruin ML, et al. Prolonged QTc interval and risk of sudden cardiac death in older adults. *J Am Coll Cardiol*. 2006;47(2):362-367.
30. Al-Khatib SM, LaPointe NM, Kramer JM, et al. What clinicians should know about the QT interval. *JAMA*. 2003;289(16):2120-2127.
31. Shoemaker MB, Mozaffarian D, Ndumele CE, et al. Anemia and cardiovascular risk in patients with heart failure. *JACC Heart Fail*. 2019;7(4):351-361.
32. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-1358.
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.
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.
35. Chen L, Zhang Y, Sun L, et al. Association between sodium levels and mortality in critically ill patients: a retrospective study. *Crit Care*. 2020;24(1):604.
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.
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. Chi CY, Tsai MS, Kuo LK, et al. Post-resuscitation diastolic blood pressure is a prognostic factor for outcomes of cardiac arrest patients: a multicenter retrospective registry-based analysis. *J Intensive Care*. 2022;10(1):39.
39. 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.
40. Rokach L, Maimon O. Top-down induction of decision trees classifiers—a survey. *IEEE Trans Syst Man Cybern C Appl Rev*. 2005;35(4):476-487.
41. 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.
42. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min*. 2016:785-794.

43. Friedman JH. Multivariate adaptive regression splines. *Ann Stat*. 1991;19(1):1-67.

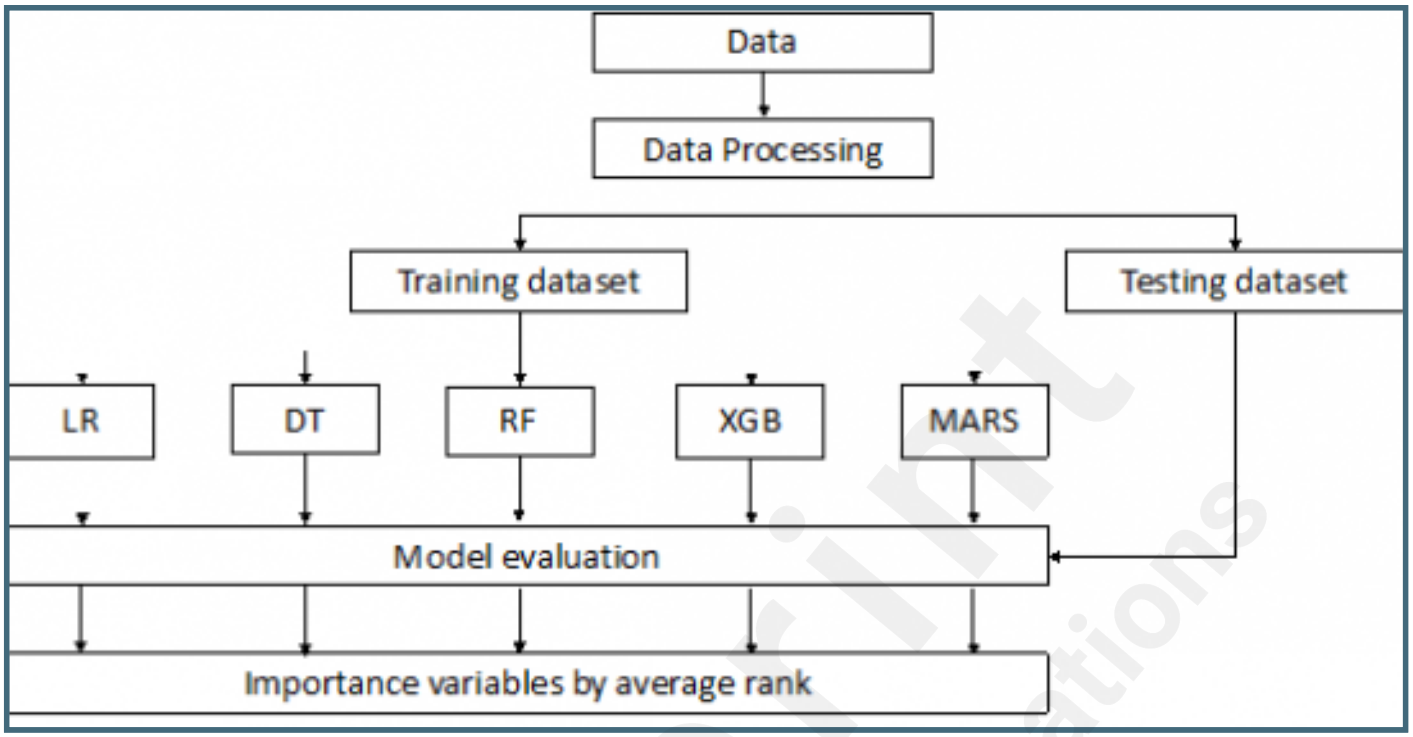
Supplementary Files

Figures

A flow diagram illustrating the selection of patients for the IHCA and non-IHCA groups.



A flow diagram of the machine learning analytical process used in this study.



Structure of the decision tree model illustrating the key predictive pathways for in-hospital cardiac arrest.