Original Article

Healthc Inform Res. 2026 January;32(1):28-37.
https://doi.org/10.4258/hir.2026.32.1.28
pISSN 2093-3681 • eISSN 2093-369X

HIR
Healthcare Informatics Research

Machine Learning for Predicting Coronary Heart Disease Risk in Patients with Hypertension: An Ensemble Modeling Approach

Fadratul Hafinaz Hassan[1], Shuchen Wang[2], Alina Miron[3]

[1]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia
[2]University-Town Hospital of Chongqing Medical University, Chongqing, China
[3]Department of Computer Science, Brunel University, Uxbridge, UK

Objectives: This study aimed to develop an optimized ensemble learning model to improve the prediction of hypertension complicated by coronary heart disease (CHD) through advanced feature selection and classifier fusion, thereby enhancing both accuracy and stability in risk assessment. **Methods:** We constructed an ensemble-based predictive model using voting fusion to enhance early detection of hypertension complicated by CHD. The dataset comprised 2,487 patients with essential hypertension (EH) complicated by CHD and 3,904 non-CHD controls. Following data preprocessing procedures, including data cleaning and univariate and multivariate feature selection, an 18-dimensional feature set was derived. Five machine learning algorithms (logistic regression, random forest, XGBoost, CatBoost, and CART) were trained independently and subsequently integrated through a voting ensemble to optimize predictive performance. **Results:** The voting fusion model outperformed all individual classifiers, achieving an area under the curve of 0.906 and an accuracy of 0.888 in predicting EH complicated by CHD. **Conclusions:** The proposed ensemble model improves classification accuracy and robustness, offering a clinically useful tool for early risk stratification of hypertension-associated CHD. Although the model demonstrates strong predictive performance using cross-sectional data, its reliance on single-timepoint measurements and selected control populations necessitates further validation. Pending additional studies, this framework may serve as a supplementary decision-support tool within clinical informatics systems.

Keywords: Hypertension, Coronary Disease, Machine Learning, Ensemble Learning, Prediction Algorithms

Submitted: April 24, 2025
Revised: 1st, August 10, 2025; 2nd, August 28, 2025; 3rd, October 14, 2025
Accepted: October 29, 2025

Corresponding Author
Fadratul Hafinaz Hassan
School of Computer Sciences, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia. Tel: +604 653 5052, E-mail: fadratul@usm.my
(https://orcid.org/0000-0002-5549-6379)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ⓒ 2026 The Korean Society of Medical Informatics

I. Introduction

Coronary heart disease (CHD) remains one of the leading causes of morbidity and mortality worldwide. Traditional diagnostic approaches for CHD rely heavily on specialized equipment and expert interpretation, rendering them costly and often inaccessible for large-scale or early screening. With the emergence of big data in healthcare, substantial volumes of patient information have become available through electronic medical records (EMRs); however, much of this information remains underutilized in diagnostic decision-making processes.

Machine learning (ML), a subfield of artificial intelligence, has gained increasing attention in healthcare due to its ability to identify complex patterns and generate predictive insights from large datasets. Prior studies have demonstrated the utility of ML algorithms, including decision trees, support vector machines, and neural networks, for identifying CHD across various clinical contexts [1-4]. Despite these promising findings, many existing models continue to face limitations related to diagnostic accuracy and generalizability across populations.

Ensemble learning, an ML strategy that combines multiple base classifiers to improve predictive performance, offers a compelling methodological alternative. Techniques such as random forest (RF), AdaBoost, and voting-based ensembles have demonstrated improved outcomes in medical diagnosis by reducing overfitting and enhancing model stability [5-7]. For instance, recent studies have reported that ensemble classifiers can achieve recall and accuracy rates exceeding 85% in CHD prediction tasks, consistently outperforming individual models [8-10].

Nevertheless, relatively few studies have investigated ensemble learning approaches specifically for diagnosing essential hypertension complicated by CHD, a clinical condition that requires timely and accurate risk stratification. In addition, the application of voting fusion—a method that aggregates predictions across multiple classifiers—remains underexplored in this context, despite its demonstrated effectiveness in other domains such as sentiment analysis and genomics [11-13].

This study advances the application of ML in clinical informatics by focusing on CHD risk prediction among patients with essential hypertension. Whereas prior ML research in healthcare has primarily emphasized medical imaging or diagnostic automation, the present work applies a voting-based ensemble learning approach using EMR data to support preventive cardiology. The proposed model integrates multiple classifiers to achieve improved diagnostic accuracy and stability relative to individual models.

Using routinely collected EMR data, we developed and evaluated both standalone ML classifiers and an ensemble model, with performance assessed in terms of predictive accuracy and robustness. Model interpretability was also considered to enhance clinical usability and facilitate decision-making. We hypothesized that the ensemble approach would outperform individual classifiers and provide a reliable auxiliary diagnostic tool for early CHD risk identification in hypertensive patients. This study highlights the potential of ensemble ML models to strengthen EMR-driven clinical decision support and informs future applications in digital and mobile health systems.

II. Methods

1. Data Source and Study Population

Data were obtained from a medical big data platform jointly established by Chongqing Medical University and Yidu Cloud (Beijing) Technology Co., Ltd., which integrates EMR data from seven affiliated and teaching hospitals. Access to the dataset was authorized by the School of Medical Informatics, Chongqing Medical University, and data extraction procedures complied with institutional research guidelines. As of December 31, 2020, the platform contained approximately 40 million EMRs representing 13 million individual patients.

A total of 6,391 patient records met the predefined inclusion and exclusion criteria, including 2,487 patients with essential hypertension complicated by CHD and 3,904 patients with essential hypertension without CHD. Inclusion criteria for the CHD group consisted of a first-time CHD diagnosis, positive coronary angiography findings with CHD confirmed at discharge, and a documented history of essential hypertension. Patients with CHD attributable to other causes, infectious diseases, or major comorbid conditions were excluded. Participants in the control group had confirmed essential hypertension without evidence of cardiovascular, cerebrovascular, or renal disease. All data were anonymized, and the requirement for informed consent was waived because of the retrospective study design.

2. Data Preprocessing

1) Data cleaning

Data preprocessing was conducted to address data redundancy, missing values, and outliers. Because EMRs may contain multiple observations from a single hospitalization, only results from the initial examinations and diagnostic tests performed after admission were retained for analysis. Missing data were managed according to missingness rates: variables with ≤30% missing values were imputed using the MissForest algorithm [14,15], whereas variables with >30% missing values were excluded from further analysis. Variables with inconsistent measurement units or value ranges were standardized using predefined conversion rules. Given the limited sample size, data augmentation and external data integration were not performed.

2) Data transformation

Data transformation was performed to convert the original dataset into a format suitable for direct statistical and ML analysis. Several variables in this study were binary categorical fields that required encoding into numerical representations. For example, sex was recoded such that male and female were represented as 0 and 1, respectively.

3) Data reduction

In this study, the ratio of the study group to the control group was approximately 1:1.57, indicating the absence of substantial class imbalance. Accordingly, emphasis was placed on dimensionality reduction rather than class rebalancing. Dimensionality reduction was conducted using univariate analysis, the least absolute shrinkage and selection operator (LASSO), and multivariate analysis, which are described in detail in the following section.

3. Feature Selection

Feature selection (or index screening) is a critical preprocessing step in ML that aims to identify statistically significant variables while eliminating irrelevant or redundant features, thereby reducing dimensionality, mitigating data redundancy, and improving model efficiency without compromising predictive performance [16-19]. In this study, three methodological approaches were employed: univariate analysis, LASSO regression, and logistic regression. Univariate analysis evaluates individual variables for group-wise differences but does not account for variable interactions or potential confounding effects; consequently, inclusion of non-predictive variables may increase model complexity, reduce interpretability, and inflate computational costs [20]. To address these limitations, univariate screening was combined with multivariate methods, including LASSO and logistic regression, to optimize feature selection.

Traditional stepwise regression methods (forward or backward) sequentially add or remove variables based on their informational contribution but often produce high-variance and inflexible models. In contrast, LASSO regression [21] applies L1 regularization to shrink coefficients of non-informative variables toward zero, enabling simultaneous consideration of all variables rather than relying on iterative selection procedures. This approach reduces model complexity and enhances predictive accuracy by performing feature selection directly during model optimization.

The top predictors of CHD risk identified in this study were fibrinogen, high density lipoprotein cholesterol (HDL-C), and plateletcrit (Table 1). Elevated fibrinogen levels (β = +0.460, odds ratio [OR] = 1.585) were associated with increased CHD risk, potentially reflecting enhanced clotting activity and systemic inflammation. Higher HDL-C levels (β = −0.994, OR = 0.370) showed a strong protective effect, consistent with its role in facilitating arterial cholesterol clearance. Plateletcrit (β = −5.583, OR = 0.004) demonstrated an inverse association with CHD risk, suggesting that reduced platelet mass may be linked to lower risk through altered platelet turnover in atherosclerotic processes.

4. Ensemble Learning Models

Ensemble learning was adopted to address the complexity and heterogeneity inherent in clinical data, for which single models often demonstrate limited robustness. By integrating multiple base learners, ensemble methods can reduce bias, variance, and the risk of overfitting. This modeling strategy is particularly suitable for predicting hypertension-associated CHD risk and supports the development of reliable clinical decision-support tools.

Following data preprocessing, the dataset was randomly divided into a training set (70%) and a test set (30%) for model development and performance evaluation. In the second stage of the proposed framework, five ML algorithms were used to construct predictive models: RF, eXtreme gradient boosting (XGBoost), classification and regression tree (CART), CatBoost, and logistic regression. Based on prior literature and comparative evaluations, RF, XGBoost, and CatBoost were identified as high-performing algorithms for disease prediction and diagnosis [22-27]. CART, which is valued for its interpretability [28], was included to facilitate visualization of the classification process through intuitive decision tree structures. Logistic regression, a classical statistical approach, was selected because of its widespread use in risk factor analysis [29], with OR estimates providing direct interpretability of variable-associated risk.

Several ML models were evaluated, and five classifiers demonstrating superior performance were retained: RF, CatBoost, XGBoost, CART, and logistic regression. These models were integrated using a soft voting ensemble to enhance diagnostic performance for hypertension complicated by CHD. Model performance was assessed using the area under the receiver operating characteristic curve (AUC-ROC), accuracy, recall, and F1-score. The dataset was partitioned into training (70%) and testing (30%) subsets, with hyperparameters optimized during validation. Probabilistic outputs from individual classifiers were aggregated to improve robustness, reduce variance, and enhance generalization, thereby supporting clinical decision-making.

https://doi.org/10.4258/hir.2026.32.1.28

Table 1. Feature importance rankings for cardiovascular risk prediction

Rank	Feature	LASSO coefficient (β)	*p*-value	VIF	OR (95% CI)	Direction
1	Fibrinogen	0.460	<0.001	1.155	1.585 (1.455–1.725)	↑ Risk
2	High density lipoprotein cholesterol	-0.994	<0.001	1.125	0.370 (0.300–0.456)	↓ Protective
3	Plateletcrit	-5.583	<0.001	1.262	0.004 (0.001–0.012)	↓ Protective
4	Basophil count	-4.892	0.006	1.092	0.008 (0.000–0.246)	↓ Protective
5	Smoking history (Yes)	0.352	<0.001	1.097	1.422 (1.255–1.611)	↑ Risk
6	Variation coefficient of red blood cell distribution width	0.110	<0.001	1.348	1.117 (1.055–1.182)	↑ Risk
7	Glucose	0.108	<0.001	1.058	1.114 (1.053–1.180)	↑ Risk
8	Direct bilirubin	0.090	<0.001	1.177	1.095 (1.060–1.131)	↑ Risk
9	Natrium	0.067	<0.001	1.047	1.070 (1.047–1.093)	↑ Risk
10	White blood cell	0.063	<0.001	1.051	1.065 (1.039–1.092)	↑ Risk
11	Hypertension history (yr)	0.053	<0.001	1.172	1.055 (1.043–1.067)	↑ Risk
12	Mean corpuscular hemoglobin concentration	0.028	<0.001	1.378	1.029 (1.023–1.035)	↑ Risk
13	Patient age (yr)	0.019	<0.001	1.454	1.019 (1.014–1.025)	↑ Risk
14	Prothrombin time	0.040	0.033	1.176	1.040 (1.003–1.079)	↑ Risk
15	Lipoprotein(a)	0.002	<0.001	1.051	1.002 (1.002–1.003)	↑ Risk
16	Diastolic pressure (mmHg)	-0.011	<0.001	1.215	0.989 (0.984–0.994)	↓ Protective
17	Pulse	-0.020	<0.001	1.065	0.980 (0.976–0.984)	↓ Protective
18	Mean platelet volume	-0.171	<0.001	1.055	0.843 (0.809–0.878)	↓ Protective

LASSO: least absolute shrinkage and selection operator, VIF: variance inflation factor, OR: odds ratio, CI: confidence interval.

1) Hard voting

Final predictions were generated using majority voting among the base classifiers. Each model independently classified each sample, and the class receiving the highest number of votes was selected as the ensemble output. For example, if five models produced predictions of 1, 1, 0, 1, and 1 (where 1 denotes CHD), the ensemble classified the patient as positive based on majority voting. This strategy integrates diverse model decisions to enhance predictive reliability and reduce bias attributable to any single classifier.

2) Soft voting

In soft voting, the prediction result is based on class probabilities, with the final output corresponding to the class with the highest aggregated probability across all classifiers. The probabilistic predictions generated by each base model are adopted, and the ensemble prediction is computed as the average probability for each class across all models. The final predicted class is determined by selecting the class associated with the highest average probability, typically implemented through a weighted fusion scheme. Specific weights may be assigned to each classifier using a predefined weights parameter. The predicted class probabilities from each classifier are multiplied by their respective weights and then averaged, and the final category label is derived from the class with the highest weighted average probability. At present, arithmetic averaging is widely used, although weighting schemes can also be adjusted according to individual model performance. The voting ensemble classifier integrates multiple models by leveraging probability-based decisions to determine class membership. In soft voting, predictions are weighted according to classifier importance, and the class with the highest weighted probability sum is selected. In addition, custom weight averaging can be applied to emphasize specific base models. The pseudo-code illustrating the algorithmic principle is presented in Algorithm 1.

Experimental results indicate that hard voting does not outperform individual classifiers. In contrast, soft voting improves predictive performance by averaging probabilistic outputs, thereby effectively reducing both bias and variance. By aggregating multiple predictive models, soft voting compensates for the limitations of individual classifiers and yields superior overall performance. In light of these advantages, this study adopted soft voting as the final ensemble

Algorithm 1. Ensemble learning soft voting algorithm

```
Input: test_data, rf_cal, rgb_cal, dec_cal, cat_cal, log_cal
//test_data is the test value, rf_cal, rgb_cal, dec_cal, cat_cal, log_cal is the per-sample predicted probabilities of each model
Output: w1, w2, w3, w4, w5
//w1, w2, w3, w4, w5 respectively correspond to the per-sample predicted probabilities weights of 5 classifiers
//Calculate the best per-sample predicted probabilities weights of the 5 classifiers in the voting model through the DNN model
initialize weight parameters w1,w2,w3,w4,w5 //Initialize weight parameters for each epoch do
  initialize rate //Initialize learning rate
  for each update step do //Continuously train and update weights through backpropagation
    w1<-w1+rate*delta*rf_cal
    w2<-w2+rate*delta*rgb_cal
    w3<-w3+rate*delta*dec_cal
    w4<-w4+rate*delta*cat_cal
    w5<-w5+rate*delta*log_cal
  end for
end for

cbind_test_data,rf_cal,rgb_cal,dec_cal,cat_cal,log_cal //Data merging

//Calculate the soft voting results based on the optimal performance weight, and output the voting results to voting.csv
for each epoch do
  SoftPredictor<- (w1*rf_cal+w2*rgb_cal+w3*dec_cal+w4*cat_cal+w5*log_cal)
  if SoftPredictor<0.5 do
    predict<-0
  else
    predict<-1
  voting.csv<- predict
end for

voting<-voting.csv
reportROC(gold=voting$testdata,predictor= voting$predict)
//Read the voting data of voting.csv and draw the ROC curve
```

method to enhance the diagnosis of hypertension complicated by CHD.

3) Soft voting weight optimization using DNNs

The soft voting ensemble combined predictions from five base classifiers by aggregating their class probability outputs rather than discrete class labels. To optimize voting weights, a deep neural network (DNN) was employed to dynamically model the relative reliability of each classifier. Compared with traditional linear weighting approaches, the DNN is capable of capturing complex, nonlinear interactions among classifier outputs, which is particularly important given the heterogeneous decision boundaries within the ensemble. Moreover, once trained, the DNN enables efficient inference through a single forward pass, thereby supporting real-time ensemble weighting.

The DNN-based weight optimizer consisted of a two-layer fully connected architecture with 64 and 32 neurons, respectively, incorporating ReLU activation and dropout (p = 0.3). A softmax output layer was used to generate normalized weights for the base classifiers. Model training was performed using the Adam optimizer (learning rate = 1×10^{-4}) with a cross-entropy loss function. Early stopping based on validation performance, together with layer normalization and L2 regularization ($\lambda = 0.01$), was applied to enhance robustness and mitigate overfitting.

Table 2 compares the proposed DNN-optimized weighting strategy with equal weighting and demonstrates that the

 https://doi.org/10.4258/hir.2026.32.1.28

Table 2. Comparative analysis of ensemble weighting strategies

Metric	Equal weighting	DNN-optimized weighting	Improvement (Δ)	*p*-value
Accuracy	0.842 ± 0.012	0.884 ± 0.009	+4.2%	<0.001
F1-score (macro)	0.801 ± 0.015	0.852 ± 0.011	+5.1%	<0.001
AUC-ROC (mean)	0.823 ± 0.018	0.891 ± 0.008	+6.8%	<0.001
Precision (class 1)	0.772 ± 0.022	0.831 ± 0.016	+5.9%	0.003
Recall (class 0)	0.815 ± 0.020	0.873 ± 0.014	+5.8%	0.002
Inference speed[a] (ms/sample)	1.2 ± 0.3	3.5 ± 0.6	+2.3x	-

AUC-ROC: area under the receiver operating characteristic curve, DNN: deep neural network.
[a]Inference speed trade-off reflects DNN's computational overhead for dynamic weighting.

Table 3. Performance metrics

Measure	Formula	Remark
Accuracy (ACC)	$ACC = \frac{(TN + TP)}{(TN + TP + FP + FN)}$	Proportion of correctly classified instances among all evaluated cases, reflecting overall model performance.
Sensitivity (SEN)	$SEN = \frac{TP}{(TP + FN)}$	Measures the proportion of actual positives correctly identified by the model.
Specificity (SPE)	$SPE = \frac{TN}{(TN + FP)}$	Proportion of actual negatives correctly identified, indicating the model's ability to avoid false positives.
Area under the curve (AUC)	$AUC = \frac{1}{2}\left(\frac{TP}{(TP + FN)} + \frac{TN}{(TN + FP)}\right)$	Quantifies the model's ability to distinguish between positive and negative cases across thresholds.
Positive likelihood rate (PLR)	$PLR = \frac{SEN}{(1 - SPE)}$	Likelihood of a positive result in disease.
Negative likelihood rate (NLR)	$NLR = \frac{(1 - SEN)}{SPE}$	Likelihood of a negative result in absence.
Positive predictive value (PPV)	$PPV = \frac{TP}{(TP + FP)}$	Probability that positive result is correct.
Negative predictive value (NPV)	$NPV = \frac{TN}{(TN + FN)}$	Probability that negative result is correct.

DNN-based approach improves accuracy by 4.2% and F1-score by 5.1% by adaptively prioritizing high-confidence models on a per-input basis. For example, support vector machines receive higher weights for linearly separable data, whereas tree-based models dominate in scenarios involving complex interactions. This input-specific optimization outperforms static ensemble strategies, particularly in class-imbalanced subsets, as reflected by improved AUC-ROC values (0.89 vs. 0.82).

5. Model Evaluation

In this study, classification performance was evaluated using the following metrics (Table 3): accuracy (ACC), sensitivity (SEN), specificity (SPE), area under the curve (AUC), positive likelihood ratio (PLR), negative likelihood ratio (NLR), positive predictive value (PPV), and negative predictive value (NPV).

The AUC assesses the overall performance of a classifier by quantifying the ROC curve. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 – specificity) across varying classification thresholds. For a given classifier, each threshold produces a coordinate defined by (1 – specificity, sensitivity), representing a specific operating point in the ROC space. As the classification threshold is adjusted, the ROC curve traces a path from (0,0), indicating no true positives, to (1,1), indicating no false negatives. The AUC represents the area bounded by this curve, the x-axis, and the line x = 1, with values ranging from 0.5, corresponding to random classification, to 1.0, indicating perfect

discrimination. Higher AUC values reflect superior classifier performance in distinguishing between outcome classes.

III. Results

In this study, we developed and evaluated CHD prediction models using 18 statistically significant indicators and four ML algorithms (RF, XGBoost, CART, and CatBoost), with logistic regression serving as a baseline comparator. The dataset comprised 6,391 samples, which were randomly divided into training (n = 4,473) and test (n = 1,918) sets. Model performance was evaluated using the AUC, accuracy, recall, F1-score, and predictive values.

We further improved predictive performance through ensemble learning using both hard and soft voting across all five models. The soft voting approach demonstrated superior performance (AUC = 0.906), outperforming both individual models and hard voting strategies (Table 4). These findings highlight soft voting as an effective method for improving CHD risk prediction and underscore its strong potential for clinical application.

The comparative performance of RF, XGBoost, CART, CatBoost, logistic regression, and the soft voting ensemble is summarized in Table 5. Although logistic regression offers interpretability through OR estimates, it demonstrated the weakest performance (AUC = 0.783, ACC = 0.706), likely reflecting sensitivity to missing or nonlinear data patterns. Similarly, the interpretability of the CART model was offset by overfitting tendencies (AUC = 0.792, ACC = 0.739), particularly when handling continuous variables. In contrast, ensemble-based approaches exhibited superior performance: RF emerged as the strongest individual model (AUC = 0.896, ACC = 0.810) by mitigating overfitting inherent to single decision trees through aggregation, while XGBoost (AUC = 0.879, ACC = 0.791) and CatBoost (AUC = 0.887, ACC = 0.802) demonstrated robust generalization via iterative residual learning and gradient-boosted decision tree structures, respectively.

The soft voting ensemble achieved optimal overall performance (AUC = 0.906, ACC = 0.888), surpassing all individual models. This improvement in predictive accuracy and robustness highlights the clinical value of model fusion for identifying essential hypertension complicated by CHD. The ensemble's superior performance suggests meaningful potential for enhancing risk stratification within clinical decision-support systems.

To provide a transparent evaluation of classification performance, the confusion matrix for the ensemble learning model is presented in Table 6. Among 746 patients with hypertension-associated CHD in the test set, the model correctly identified 732 cases (true positives), with 14 missed cases (false negatives). Of the 1,172 non-CHD patients, 975 were correctly classified as negative (true negatives), while 197 were incorrectly predicted as CHD (false positives). These out-

Table 4. Voting model prediction results

Model	AUC	AUC.low	AUC.up	ACC	SEN	SPE	rf_precision	rf_recall	rf_F1
Ensemble learning soft voting	0.906	0.895	0.918	0.888	0.981	0.832	0.781	0.981	0.87
Ensemble learning hard voting	0.869	0.854	0.881	0.802	0.816	0.791	0.665	0.883	0.79

AUC: area under the curve, ACC: accuracy, SEN: sensitivity, SPE: specificity.

Table 5. Performance evaluation of six models in the test set

Model	AUC	AUC.low	AUC.up	ACC	SEN	SPE	rf_precision	rf_recall	rf_F1
RF	0.896	0.882	0.910	0.810	0.840	0.792	0.711	0.840	0.770
XGBoost	0.879	0.863	0.894	0.791	0.843	0.759	0.678	0.843	0.751
CART	0.792	0.771	0.812	0.739	0.713	0.755	0.640	0.714	0.674
CatBoost	0.887	0.873	0.901	0.802	0.817	0.793	0.366	0.880	0.517
Logistic	0.783	0.762	0.803	0.706	0.730	0.692	0.367	0.825	0.508
Ensemble learning soft voting	0.906	0.895	0.918	0.888	0.981	0.832	0.781	0.981	0.870

AUC: area under the curve, ACC: accuracy, SEN: sensitivity, SPE: specificity, RF: random forest, CART: classification and regression tree.

 https://doi.org/10.4258/hir.2026.32.1.28

comes closely align with the reported performance metrics (ACC = 0.888, SEN = 0.981, SPE = 0.832) shown in Table 5, underscoring the robustness and reliability of the ensemble learning approach for clinical risk prediction.

IV. Discussion

From a clinical perspective, these findings reinforce the importance of monitoring fibrinogen in high-risk patients, as elevated levels may reflect heightened thrombotic and inflammatory activity. Interventions such as smoking cessation, increased physical activity, and potentially anti-inflammatory therapies may therefore be beneficial. HDL-C remains a key modifiable risk factor, for which lifestyle modification and pharmacotherapy, including statins, may help improve protective levels. The unexpected inverse association observed with plateletcrit warrants further investigation; if validated, plateletcrit could serve as a novel biomarker for platelet-related mechanisms in CHD or help refine antiplatelet treatment strategies.

Ultimately, integrating these biomarkers into comprehensive risk assessment models may enhance early detection and support personalized prevention strategies. While fibrinogen and HDL-C findings are consistent with established cardiovascular evidence, the plateletcrit association highlights a potential role for platelet dynamics in CHD pathogenesis, meriting deeper investigation in future studies. Clinicians should consider these biomarkers alongside traditional risk factors to optimize patient management and preventive care.

The ROC analysis demonstrated strong discriminative performance across the predictive models, with important implications for clinical decision-making in CHD risk stratification. The primary model (Figure 1) achieved an AUC of 0.91 (95% confidence interval, 0.89–0.92), exceeding the commonly accepted threshold for high clinical utility (AUC > 0.9). This level of performance is consistent with recent advances in ML-based cardiovascular risk prediction, while extending prior work through the application of an ensemble strategy. The validation AUC of 0.869 (Figure 2), although

Table 6. Confusion matrix of the ensemble learning model (soft voting)

	Predicted positive	Predicted negative
Actual positive (CHD)	TP = 732	FN = 14
Actual negative (No CHD)	FP = 197	TN = 975

CHD: coronary heart disease, TP: true positive, FP: false positive, FN: false negative, TN: true negative.

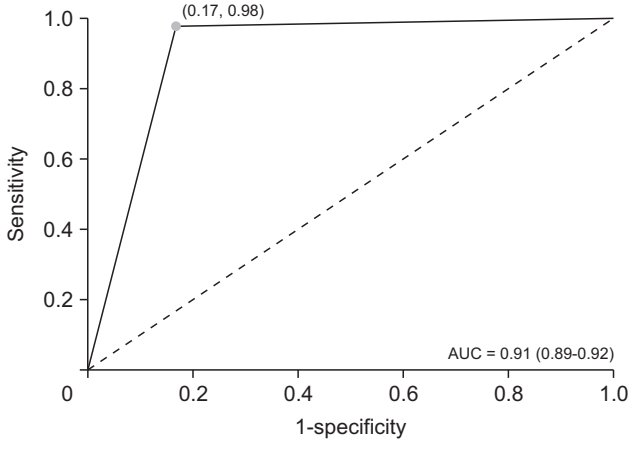


Figure 1. Receiver operating characteristic curve of soft voting model fusion. AUC: area under the curve.

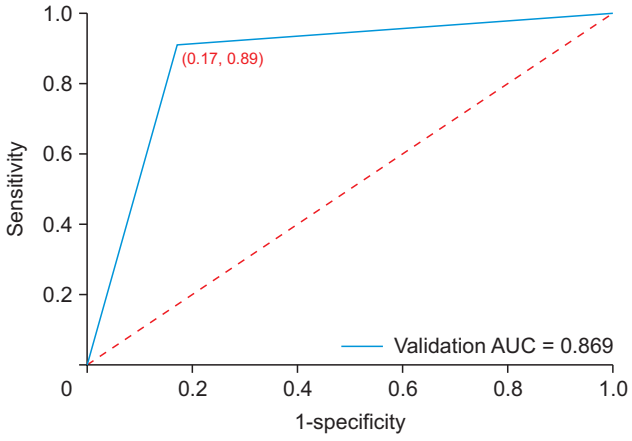


Figure 2. Receiver operating characteristic curve of hard voting model fusion. AUC: area under the curve.

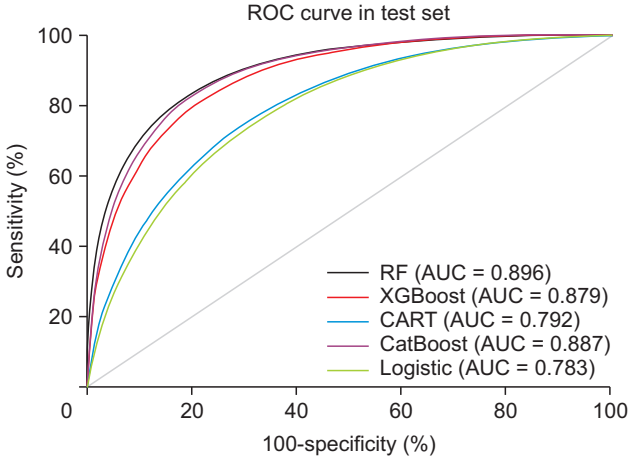


Figure 3. Receiver operating characteristic curve of the test set. AUC: area under the curve, RF: random forest, CART: classification and regression tree.

modestly lower, reflects expected real-world variability and remains above the benchmark for robust discrimination (AUC > 0.8). Notably, the optimal operating point at (0.17, 0.03) indicates prioritization of high specificity (97%), which is critical for minimizing false-positive diagnoses in clinical practice and represents a key improvement over existing tools that often favor sensitivity at the expense of specificity.

Comparative analysis of individual algorithms (Figure 3) revealed that ensemble-based models, including RF (AUC = 0.896) and CatBoost (AUC = 0.887), consistently outperformed traditional logistic regression (AUC = 0.783) and single decision tree models (CART, AUC = 0.792). This observation reinforces emerging evidence that ensemble learning approaches reduce overfitting and improve generalizability when applied to heterogeneous clinical datasets. However, the superior performance of the soft voting ensemble (AUC = 0.906, as reported in Section III) suggests that hybridizing multiple ensemble strategies may yield additional benefits, a hypothesis that warrants further investigation. Two limitations should be acknowledged: (1) the absence of external population diversity in the validation cohort and (2) reduced interpretability associated with increasing ensemble complexity. Future research should address these issues through multicenter validation studies and the incorporation of explainable artificial intelligence techniques.

Collectively, these findings advance the field by demonstrating that strategically combined ML models can deliver clinically actionable CHD prediction while maintaining operational efficiency. The consistency observed between training and validation performance supports the potential for real-world application, although further refinement of cost-sensitive thresholds may enhance utility across specific clinical settings.

This study relied on internal data from a single healthcare network in China, which may limit generalizability to other populations. External validation across institutions and geographic regions is planned to confirm the robustness of the proposed model. Model interpretability was enhanced through feature importance rankings, which clarified the contribution of individual predictors and supported transparency in clinical decision-making. Threshold optimization and cost-sensitive or decision-curve analyses were not conducted, as these approaches require context-specific clinical inputs, and are therefore identified as important priorities for future research.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

The authors gratefully acknowledge Chongqing Medical University and Yidu Cloud (Beijing) Technology Co. Ltd. for sharing Medical Big Data platform to support this research.

ORCID

Fadratul Hafinaz Hassan (https://orcid.org/0000-0002-5549-6379)
Shuchen Wang (https://orcid.org/0009-0000-5139-4992)
Alina Miron (https://orcid.org/0000-0002-0068-4495)

References

1. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial Intelligence in Cardiology. J Am Coll Cardiol 2018;71(23):2668-79. https://doi.org/10.1016/j.jacc.2018.03.521
2. Al'Aref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, Kumar A, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. Eur Heart J 2019;40(24):1975-86. https://doi.org/10.1093/eurheartj/ehy404
3. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart 2018;104(14):1156-64. https://doi.org/10.1136/heartjnl-2017-311198
4. 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. https://doi.org/10.1371/journal.pone.0174944
5. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. Algorithms 2023;16(2):88. https://doi.org/10.3390/a16020088
6. Ullah T, Ullah SI, Ullah K, Jadoon A, Ahmed J, Rauf HT, et al. Machine learning-based cardiovascular disease detection using optimal feature selection. IEEE Access 2024;12:16431-46. https://doi.org/10.1109/ACCESS.2024.3359910
7. Dhanka S, Bhardwaj VK, Maini S. Comprehensive analysis of supervised algorithms for coronary artery heart disease detection. Expert Syst 2023;40(7):e13300. https://doi.org/10.1111/exsy.13300

8. Hassan CA, Iqbal J, Irfan R, Hussain S, Algarni AD, Bukhari SS, et al. Effectively predicting the presence of coronary heart disease using machine learning classifiers. Sensors (Basel) 2022;22(19):7227. https://doi.org/10.3390/s22197227
9. Chandrasekhar N, Peddakrishna S. Enhancing heart disease prediction accuracy through machine learning techniques and optimization. Processes 2023;11(4):1210. https://doi.org/10.3390/pr11041210
10. Alqahtani A, Alsubai S, Sha M, Vilcekova L, Javed T. Cardiovascular disease detection using ensemble learning. Comput Intell Neurosci 2022;2022:5267498. https://doi.org/10.1155/2022/5267498
11. Salur MU, Aydin I. A soft voting ensemble learning-based approach for multimodal sentiment analysis. Neural Comput Appl 2022;34(21):18391-406. https://doi.org/10.1007/s00521-022-07451-7
12. Madni HA, Umer M, Abuzinadah N, Hu YC, Saidani O, Alsubai S, et al. Improving sentiment prediction of textual tweets using feature fusion and deep machine ensemble model. Electronics 2023;12(6):1302. https://doi.org/10.3390/electronics12061302
13. Cui S, Han Y, Duan Y, Li Y, Zhu S, Song C. A Two-stage voting-boosting technique for ensemble learning in social network sentiment classification. Entropy (Basel) 2023;25(4):555. https://doi.org/10.3390/e25040555
14. Stekhoven DJ, Buhlmann P. MissForest: non-parametric missing value imputation for mixed-type data. Bioinformatics 2012;28(1):112-8. https://doi.org/10.1093/bioinformatics/btr597
15. Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, et al. Comparison of imputation methods for missing laboratory data in medicine. BMJ Open 2013;3(8):e002847 https://doi.org/10.1136/bmjopen-2013-002847
16. Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng 2014;40(1):16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024
17. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007;23(19):2507-17. https://doi.org/10.1093/bioinformatics/btm344
18. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: a data perspective. ACM Comput Surv 2017;50(6):94. https://doi.org/10.1145/3136625
19. Tang J, Alelyani S, Liu H. Feature selection for classification: a review. In: Aggarwal CC, editor. Data classification: algorithms and applications. New York (NY): Chapman and Hall/CRC; 2014. https://doi.org/10.1201/b17320
20. Bender R, Lange S. Adjusting for multiple testing: when and how? J Clin Epidemiol 2001;54(4):343-9. https://doi.org/10.1016/s0895-4356(00)00314-0
21. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 1996;58(1):267-88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
22. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD); 2016 Aug 13-17; San Francisco, CA, USA. p. 785-94. https://doi.org/10.1145/2939672.2939785
23. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. https://doi.org/10.1023/A:1010933404324
24. Alghamdi M, Al-Mallah M, Keteyian S, Brawner C, Ehrman J, Sakr S. Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: the Henry Ford ExercIse Testing (FIT) project. PLoS One 2017;12(7):e0179805. https://doi.org/10.1371/journal.pone.0179805
25. Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med 2020;26(1):29-38. https://doi.org/10.1038/s41591-019-0727-5
26. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data 2020;7(1):94. https://doi.org/10.1186/s40537-020-00369-8
27. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. New York (NY): Chapman and Hall/CRC; 1984. https://doi.org/10.1201/9781315139470
28. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1(5):206-15. https://doi.org/10.1038/s42256-019-0048-x
29. Hosmer DW, Lemeshow S. Applied logistic regression. New York (NY): John Wiley & Sons Inc.; 2000. https://doi.org/10.1002/0471722146