*JAMIA Open*, 2025, **8(5)**, ooaf114
https://doi.org/10.1093/jamiaopen/ooaf114
**Research and Applications**

Research and Applications

# Early auxiliary diagnosis model for chest pain triad based on artificial intelligence multimodal fusion

**Jun Tang, MEng[1], Fang Chen, MD*[,2], Dongdong Wu, PhD*[,1]**

[1]Department of Information, Daping Hospital, Army Medical University, Chongqing 400042, China, [2]Department of Intensive Care Unit, Daping Hospital, Army Medical University, Chongqing 400042, China

*Corresponding authors: Fang Chen, MD, Department of Intensive Care Unit, Daping Hospital, Army Medical University, Chongqing 400042, China (cfang4@tmmu.edu.cn) and Dongdong Wu, PhD, Department of Information, Daping Hospital, Army Medical University, Chongqing 400042, China (604269346@qq.com)

## Abstract

**Objectives:** Acute chest pain is a common presentation in the emergency department, characterized by sudden onset with high morbidity and mortality. Traditional diagnostic methods, such as computed tomography (CT) and CT angiography (CTA), are often time-consuming and fail to meet the urgent need for rapid triage in emergency settings.

**Materials and Methods:** We developed a multimodal model that integrates Bio-ClinicalBERT and ensemble learning (AdaBoost, Gradient boosting, and XGBoost) based on 41 382 patient data from April 1, 2013 to April 1, 2025 at Chongqing Daping Hospital. By integrating clinical texts and laboratory indicators, the model aims to classify the 3 major causes of fatal chest pain (acute coronary syndrome, pulmonary embolism, and aortic dissection), as well as other causes of chest pain, aiding rapid triage. We adopt strict data preprocessing and rank importance feature selection.

**Results:** The multimodal fusion model based on Gradient boosting exhibits the best performance: accuracy of 88.40%, area under the curve of 0.951, F1-score of 74.56%, precision of 77.50%, and recall of 72.52%. SHapley Additive exPlanations (SHAP) analysis confirmed the clinical relevance of key features such as d-dimer and high-sensitivity troponin. When reducing the number of numerical features to 30 key indicators, the model enhanced robustness without compromising performance.

**Discussion and Conclusion:** We developed an artificial intelligence model for chest pain classification that effectively addresses the problem of overlapping clinical symptoms through multimodal fusion, and the model has high accuracy. However, future work needs to better integrate model development with clinical workflows and practical constraints.

## Lay Summary

Acute chest pain is a common emergency in the emergency department and can be caused by several very dangerous cardiac or vascular conditions. Rapid identification of the specific etiology is crucial to save lives, but traditional tests (eg, CT or CTA) are often time-consuming and are not conducive to rapid triage in emergency medicine. We developed a novel artificial intelligence multimodal model that integrates 2 aspects of information: a textual description of the medical record recorded by the doctor and the patient's laboratory test results. Utilizing data from over 40 000 chest pain patients at Chongqing Daping Hospital, we completed training and testing of the model. The model performed excellently in differentiating between the 3 major fatal causes of chest pain (acute coronary syndrome, pulmonary embolism, and aortic coarctation) and other causes of chest pain, with an accuracy of 88.40%. The study also confirmed the value of certain key indicators in determining etiology. The model has the potential to become an efficient intelligent triage tool for emergency chest pain, helping physicians make faster decisions.

**Key words:** chest pain; acute coronary syndrome; pulmonary embolism; aortic dissection; ensemble learning.

## Introduction

Acute chest pain is one of the most common diseases in cardiovascular and emergency departments,[1] with a rapid onset and often the first symptom of various life-threatening conditions. Cardiologists and general practitioners often face the challenge of managing chest pain symptoms caused by complex etiologies.[2] The 2022 Heart Disease and Stroke Statistics Report released by the American Heart Association,[3] based on 2019 data, shows that the mortality rate caused by cardiovascular disease in the United States is 214.6 cases per 100 000 people. Statistics show that cardiovascular disease causes 1 death every 36.1 seconds, with a total of 2396 deaths per day. Meanwhile, in China, the 2022 China Cardiovascular Health and Disease Report shows that in 2020,[4] cardiovascular disease accounted for 48.00% and 45.86% of deaths in rural and urban areas, respectively. Although preliminary screening can be conducted among various differential diagnoses through routine methods such as medical history inquiry, physical examination, and electrocardiogram, a considerable number of patients still require further examination to confirm the diagnosis.[5] In emergencies, quickly and accurately identifying the cause of chest pain in patients, and effectively screening high-risk chest pain patients, is crucial for reducing complications, lowering mortality rates, and making reasonable treatment decisions.

Chest pain triple-rule-out (TRO) refers to 3 diseases and complications characterized by acute chest pain attacks, including acute coronary syndrome (ACS), pulmonary

**Received:** June 30, 2025; **Revised:** September 2, 2025; **Accepted:** September 4, 2025

embolism (PE), and aortic dissection (AD).[6] These diseases have a dangerous onset, high misdiagnosis rate, and mortality rate. Their clinical symptoms mainly include chest pain, difficulty breathing, and hemoptysis. Although chest pain caused by different etiologies has similarities in symptoms and different characteristics, it can manifest as pain in different parts, of different natures, and to varying degrees. Early identification of the cause and timely and effective treatment are key to the treatment of patients with acute chest pain.[7] However, traditional diagnostic methods such as computed tomography and computed tomography angiography have certain limitations. These examination methods are time-consuming, not conducive to early diagnosis, and difficult to meet the needs of rapid triage in emergency departments.

In recent years, with the rapid development of artificial intelligence (AI) technology, the application of machine learning and pre-trained language models in the medical field has gradually emerged. Machine learning can automatically learn patterns and rules from large amounts of data by constructing algorithmic models, discovering complex patterns and associated features hidden in the data, and making predictions and classifications based on them.[8,9] Pre-trained language models, with their powerful natural language processing capabilities, can understand human language and provide new avenues for processing and analyzing medical text data.[10] Although there have been numerous studies exploring the application of AI in chest pain diagnosis, most of them have focused on the diagnosis of a single chest pain disease, such as using AI to assist in the diagnosis of ACS,[11–13] PE,[14–17] or AD.[18–20] TRO has high clinical overlap in clinical manifestations and further research is needed.[21]

This study innovatively integrates pre-trained language models and machine learning techniques to construct a multimodal classification model, aiming to achieve an early auxiliary diagnosis of the 3 major emergencies of TRO and other causes of chest pain. By integrating clinical texts and laboratory indicators, this framework is expected to improve the efficiency of chest pain diagnosis, reduce misdiagnosis and missed diagnosis rates, and provide intelligent decision support for rapid triage in emergency departments.

## Methods

### Patients and dataset

We collected data on hospitalized patients with acute chest pain admitted to Chongqing Daping Hospital from April 1, 2013 to April 1, 2025. Inclusion criteria: (1) chief complaint of acute chest pain; (2) age $\geq$ 18 years old; and (3) no multiple TRO symptoms. The final dataset included a total of 41 382 patients, including ACS (n = 23 604), PE (n = 2073), AD (n = 1361), and Non-TRO chest pain (Non-TRO, such as coronary heart disease, myocarditis, heart failure, etc, n = 14 344). All clinical examinations were performed in strict accordance with the European Society of Cardiology (ESC) guidelines for ACS, PE, and AD diagnosis and management.[22–24] After the clinical diagnosis was established, an adjudication board of experienced cardiologists grouped the patients into final diagnoses.

To enable rapid classification during early diagnostic stages, we selected clinically accessible data aligned with established clinical guidelines. These data included demographics (gender, age, and BMI), clinical records (chief complaint, present medical history, and past medical history), and examination findings (physical exam, vital signs, and electrocardiographic). Laboratory data encompassed 56 indicators, including those from complete blood count (12 items), coagulation panel (6 items), cardiac injury markers (6 items), thyroid function tests (5 items), liver function tests (9 items), renal function tests (3 items), lipid profile (4 items), electrolytes (6 items), fasting blood glucose, glycated hemoglobin, procalcitonin, c-reactive protein, and carbon dioxide ($CO_2$). Detailed indicator information is provided in Supplementary File S1.

### Data preprocessing and feature engineering

To address the issue of missing values in raw data, mode imputation is used for categorical variables, and k-nearest neighbor interpolation (K = 5) is used for continuous variables. All numerical features are standardized using Z-score.

To optimize the feature space, we used the permutation_importance function[25] of scikit-learn to analyze the contribution of numerical features by selecting 30 high-impact features (eg, D-dimer, high-sensitivity troponin [hs-Tn], etc) from all numerical features, as shown in Figure 1. The core idea of the method is to measure the importance of a feature by randomly disrupting the value of that feature on the validation set and observing the impact this operation causes on the model performance. To improve the stability of the results, we repeat the reordering process n_repeats = 10 times for each feature and take the average of its performance degradation as the final importance score of that feature.

To address the issue of class imbalance, we adopt a class weighted learning strategy combined with ensemble algorithms to enhance the ability to recognize minority classes (PE or AD). In addition, clinical text data are transformed into structured features through vectorization techniques.

### Construction of multimodal models

This study proposes a multimodal architecture that integrates Bio_ClinicalBERT[26] and ensemble learning (as shown in Figure 2) to achieve early classification of TRO and other chest pain diseases by integrating clinical text and numerical data. The specific process is as follows: (1) Text feature processing: 5 types of clinical text features (chief complaint, current medical history, past medical history, physical examination, and electrocardiogram report) are input into the Bio_ClinicalBERT model (pre-trained based on MIMIC III clinical notes) to generate 512-dimensional semantic embedding vectors; (2) Numerical feature processing: The filtered 30 key numerical features are input into 3 ensemble models[27]: AdaBoost (dynamically adjusting sample weights), Gradient boosting (iterative decision tree ensemble), and XGBoost (regularized optimization tree structure); (3) Multimodal fusion: concatenate text embedding vectors with the probability outputs of 3 numerical models, train a random forest meta learner[28] through stacking strategy, and learn the optimal feature combination decision boundary.

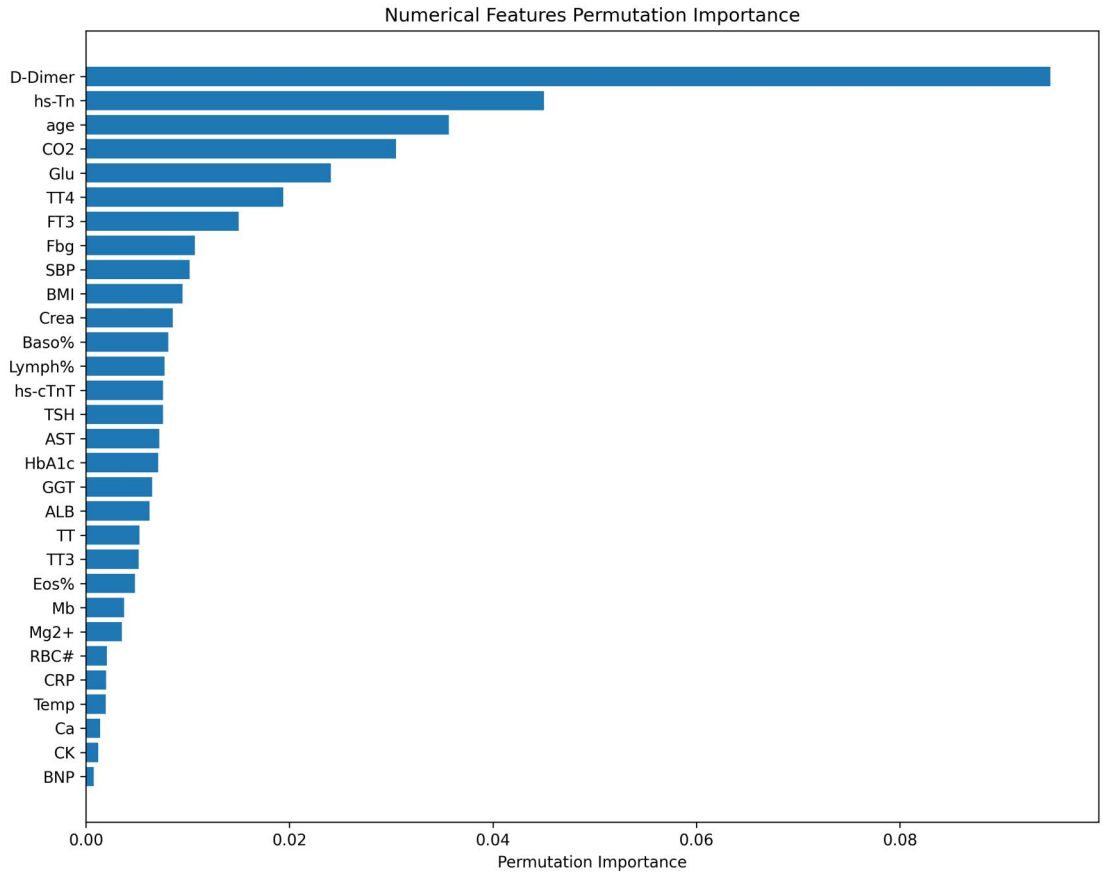


**Figure 1.** The importance ranking of the 30 key numerical features included. The vertical axis lists feature names in descending order of importance, while the horizontal axis represents the corresponding importance scores.

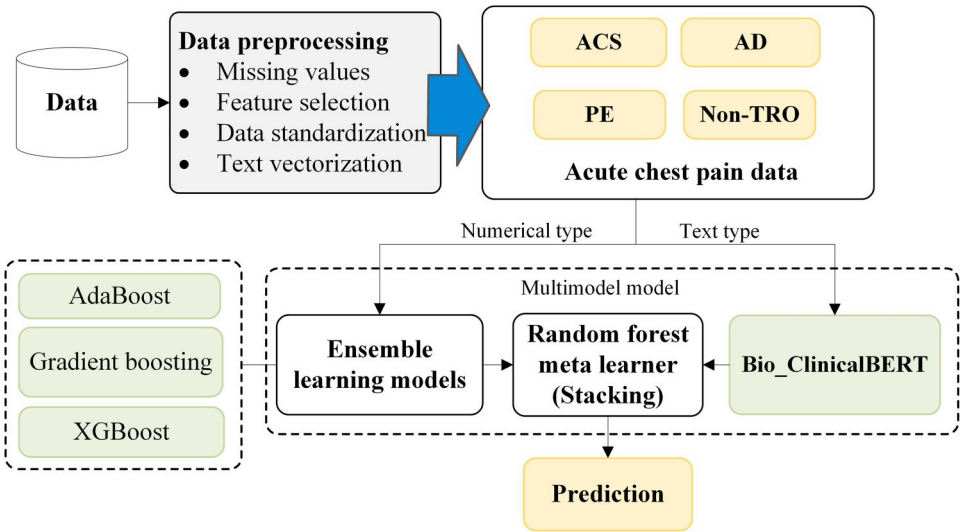


**Figure 2.** Multimodal model architecture. Displayed the process of text features, numerical features, and multimodal fusion.

## Model evaluation

In the experiment, we used 5 evaluation metrics to measure the performance of the model, including accuracy, F1-score, recall, precision, and area under the curve (AUC) value. To obtain more accurate and less biased results, we used a 5-fold cross-validation method. Finally, we took the average as the final evaluation result of the model and calculated the standard deviation (SD) to reflect the volatility of the results.

During optimization, we set the learning rate at 2e-5 and used a batch size of 32. The model was trained on a computer equipped with a 24-core CPU, 64 GB of RAM, and an NVIDIA RTX A6000 GPU with 48 GB of memory. We implemented the model via Python 3.10 and PyTorch 2.6.

## Results

### Patient characteristics

This study ultimately included 30 key numerical features (selected by feature importance ranking, as shown in Figure 1), covering clinical testing dimensions such as blood routine, biochemistry, coagulation function, and myocardial injury markers. Among them, core biomarkers such as D-dimer and hs-Tn have been confirmed by literature to have a significant correlation with TRO.[29] Table 1 presents the statistical results of the characteristics of patient groups with 4 types of chest pain causes (ACS, AD, PE, and Non-TRO), with the distribution differences of baseline data in each group presented in the form of mean ± SD.

### Overall predictive performance of multimodel models

We constructed 3 multimodal models: the Bio_ClinicalBERT model was uniformly used for text feature extraction, and the numerical models were based on AdaBoost, Gradient boosting, and XGBoost ensemble learning algorithms, respectively. The performance of these 3 models was evaluated through 5-fold cross-validation (Table 2). See Supplementary File S2 for more detailed results per fold.

From the experimental results in Table 2, it can be seen that the multimodal model Gradient boosting-based has the best comprehensive performance, with an accuracy of 88.40% ± 0.007, F1-score of 74.56% ± 0.014, precision of 77.50% ± 0.018, recall of 72.52% ± 0.012, and AUC value of 0.951 ± 0.004. Next is the Adaboost-based model, with an accuracy of 87.75% ± 0.003 and an AUC value of 0.946 ± 0.004. The XGBoost-based model is slightly inferior in accuracy and AUC, with an accuracy of 86.16% ± 0.006 and 0.917 ± 0.007, respectively. The SD of all 3 models is below 0.018, confirming the stability and generalization ability of the experiment.

Table 2 shows the average results of the 3 models separately for all classes in the 5-fold cross-validation, and as can be seen from Supplementary File S2, the performance of the results for each fold remains relatively stable. Therefore, in Table 3, we report the individual performance of the 4 classes in 1-fold of the 5-fold cross-validation for the best performing Gradient boosting-based model. Specifically, classes 1 to 4 correspond to PE, AD, ACS, and Non-TRO, respectively. The results show that the performance of classes 1 and 2 is slightly lower than that of the other 2 classes, but the difference is not significant, and the differences between the different classes across the metrics are all less than 15%. This indicates that our model can effectively identify minority classes. In addition, we further visualize the receiver operating characteristic (ROC) curves of the 3 models for each class in Figure 3. All models show high AUC values (>0.88) in all 4 classes, which indicates that our models have good discriminative ability in categorization tasks.

To improve clinical applicability, we studied the relationship between the number of features and the prediction performance using the Gradient boosting-based multimodal model with the best prediction performance (Table 4). The experimental results show that when the number of numerical features is reduced to 30, the model performance does not

**Table 1.** The statistical distribution results (mean and SD) of the 30 key numerical features included in the patient population of 4 types of chest pain causes (ACS, AD, PE, and Non-TRO).

| Characteristics | ACS (n = 23604) | AD (n = 1361) | PE (n = 2073) | Non-TRO (n = 14344) |
|---|---|---|---|---|
| Age (years) | 66.41 (11.50) | 58.69 (13.67) | 64.93 (13.51) | 57.74 (15.65) |
| BMI | 25.01 (12.99) | 24.19 (3.85) | 24.86 (20.95) | 25.09 (36.86) |
| Temperature (Temp) | 36.62 (7.06) | 36.49 (0.28) | 36.53 (0.44) | 36.66 (8.69) |
| Systolic blood pressure (SBP) | 130.69 (25.28) | 139.12 (24.66) | 127.84 (20.13) | 125.43 (23.00) |
| Red blood cell count (RBC#) | 4.43 (0.64) | 4.45 (0.77) | 4.21 (0.78) | 4.51 (0.61) |
| Lymphocyte percentage (Lymph%) | 23.53 (9.50) | 18.43 (10.62) | 19.13 (10.22) | 24.01 (9.49) |
| Basophil percentage (Baso%) | 0.33 (0.27) | 0.34 (0.29) | 0.35 (0.28) | 0.45 (0.51) |
| Eosinophil percentage (Eos%) | 2.04 (2.38) | 1.65 (2.14) | 1.93 (2.97) | 2.22 (2.67) |
| D-dimer (ug/L) | 374.03 (2323.60) | 2362.74 (9466.07) | 3060.39 (16094.29) | 408.26 (1292.42) |
| Thrombin time (TT) | 15.22 (8.43) | 15.26 (6.67) | 15.10 (7.39) | 14.65 (7.2) |
| Fibrinogen (Fbg) | 3.25 (1.01) | 3.39 (1.34) | 3.74 (1.48) | 3.58 (1.28) |
| Myoglobin (Mb) | 116.77 (231.05) | 107.41 (236.40) | 112.58 (207.83) | 52.03 (80.59) |
| Ceatine kinase (CK) | 163.41 (702.68) | 125.35 (198.21) | 377.55 (2156.19) | 155.41 (576.57) |
| hs-Tn | 0.23 (0.77) | 0.08 (0.30) | 0.06 (0.21) | 0.03 (0.17) |
| High-sensitivity cardiac troponin T (hs-cTnT) | 0.33 (1.02) | 0.02 (0.03) | 0.04 (0.07) | 0.01 (0.05) |
| Brain natriuretic peptide (BNP) | 677.97 (2202.67) | 153.85 (242.03) | 477.23 (2064.04) | 144.04 (582.12) |
| Thyroid stimulating hormone (TSH) | 2.89 (7.20) | 3.08 (9.97) | 2.96 (4.87) | 2.72 (5.07) |
| Free triiodothyronine (FT3) | 4.49 (1.14) | 4.60 (0.81) | 4.34 (1.01) | 4.86 (1.31) |
| Total triiodothyronine (TT3) | 1.35 (0.41) | 1.32 (0.36) | 1.23 (0.43) | 1.44 (0.41) |
| Total thyroxine (TT4) | 109.48 (28.74) | 114.81 (34.57) | 102.85 (49.39) | 105.08 (37.41) |
| Aspartate aminotransferase (AST) | 60.36 (219.63) | 63.90 (486.77) | 53.30 (162.84) | 29.05 (40.98) |
| γ-Glutamyl transferase (GGT) | 44.39 (71.08) | 58.18 (139.82) | 63.89 (100.67) | 43.33 (70.56) |
| Albumin (ALB) | 43.78 (9.30) | 40.91 (7.66) | 38.60 (8.97) | 42.78 (8.41) |
| Creatinine (Crea) | 83.98 (61.41) | 98.88 (116.64) | 81.95 (71.79) | 71.62 (46.98) |
| Total calcium (Ca) | 2.27 (0.13) | 2.22 (0.14) | 2.20 (0.17) | 2.27(0.14) |
| Magnesium (Mg2+) | 0.87 (0.11) | 0.86 (0.10) | 0.87 (0.11) | 0.86 (0.09) |
| Glucose (Glu) | 6.40 (2.84) | 6.19 (2.71) | 5.97 (2.19) | 5.48 (1.75) |
| Hemoglobin A1c (HbA1c) | 6.36 (1.46) | 5.98 (1.19) | 6.35 (1.43) | 6.23 (1.31) |
| C-reactive protein (CRP) | 14.61 (30.08) | 30.94 (46.95) | 42.93 (55.80) | 31.73 (48.61) |
| Carbon dioxide ($CO_2$) | 24.34 (3.00) | 24.18 (3.31) | 24.32 (3.67) | 25.27 (2.72) |

ACS, acute coronary syndrome; AD, aortic dissection; PE, pulmonary embolism; Non-TRO, Non chest pain triple-rule-out

**Table 2.** Comparison of prediction results of 3 multimodal models with 5-fold cross-validation (mean and SD).

| Multimodel model | Accuracy | F1-score | Precision | Recall | AUC |
|---|---|---|---|---|---|
| XGBoost-based | 86.16 (0.006) | 71.15 (0.007) | 73.22 (0.009) | 69.68 (0.007) | 0.917 (0.007) |
| Adaboost-based | 87.75 (0.003) | 73.13 (0.013) | 76.99 (0.008) | 70.76 (0.013) | 0.946 (0.004) |
| Gradient boosting-based | 88.40 (0.007) | 74.56 (0.014) | 77.50 (0.018) | 72.52 (0.012) | 0.951 (0.004) |

AUC, area under the curve.

**Table 3.** Individual performance of 4 classes of 1-fold in 5-fold cross-validation of the Gradient boosting-based model.

| Class | Accuracy | F1-score | Precision | Recall | AUC |
|---|---|---|---|---|---|
| Class 1 | 82.64 | 72.39 | 72.62 | 70.36 | 0.932 |
| Class 2 | 77.81 | 69.49 | 73.13 | 68.43 | 0.935 |
| Class 3 | 87.51 | 79.79 | 85.09 | 77.39 | 0.984 |
| Class 4 | 92.56 | 83.66 | 87.61 | 80.10 | 0.984 |
| Mean | 89.42 | 76.33 | 79.61 | 74.07 | 0.959 |

AUC, area under the curve.

significantly degrade, but instead the accuracy improves from 88.34% ± 0.004 to 88.40% ± 0.007. This suggests that most of the removed features are redundant or noisy features. Feature selection enhances the generalization ability and robustness of the model, which is especially important when the feature dimension is high. The slight decrease in the other metrics is as expected. A small amount of information that contributes slightly to the metrics may have been removed in the process of removing features, resulting in a small decrease in these metrics.

Next, the model framework proposed in this study utilizes SHapley Additive exPlanations (SHAP)[30] values to extract feature importance. When a model makes specific predictions based on a set of features, its contribution is determined by calculating the impact of the presence or absence of each feature on the performance of the model. If the absence of a certain feature leads to a significant decline in model performance, that feature will be assigned a higher contribution score. Figure 4 shows the SHAP value analysis results of a multimodal model based on Gradient boosting. The figure not only displays the SHAP values for each class but also shows the average SHAP values for each class. Each point corresponds to a sample in the training set and its SHAP value. The variables are arranged in descending order of SHAP values, with D-dimer having the highest average feature importance and BNP having the lowest average feature importance. The horizontal position of the point represents the impact of the feature value on the prediction result, with red indicating a higher feature value and the horizontal axis depicting the "positive" or "negative" impact of the feature on the prediction.

To more comprehensively evaluate the performance of our multimodal model and verify its advantages over unimodal models, we have added comparisons with models using only numerical data and text data in Supplementary File S3. These unimodal models utilized numerical and text features for prediction, respectively, to assess the contribution of different data sources to model performance. The experimental results show that the unimodal models do not perform as well as the multimodal model on all metrics. Specifically, the model using only text data slightly outperforms the model using only numerical data on all metrics, but neither can reach the performance level of the multimodal model (Gradient boosting-based). This indicates that the multimodal model can more comprehensively capture information from different data sources when dealing with complex clinical scenarios, thus achieving more accurate predictions.

## Discussion

This study is based on medical data from Daping Hospital in Chongqing, and it successfully developed and validated a multimodal model that integrates a pre-trained language model (Bio_ClinicalBERT) and ensemble learning (AdaBoost, Gradient boosting, and XGBoost) techniques, achieving preliminary classification of the 3 major emergencies of TRO and other chest pain related diseases. To our knowledge, this is the first study to comprehensively predict TRO using AI technology.

In terms of model construction, we used the outstanding Gradient boosting model and Bio_ClinicalBERT for stacking fusion. The model demonstrated excellent performance on the complex dataset of this study, with an accuracy of 88.40%, an F1-score of 74.56%, a precision of 77.50%, a recall of 72.52%, and an AUC value of 0.951. This experimental result fully demonstrates that our multimodal model architecture effectively integrates text and numerical data information, combines the advantages of deep learning and machine learning, and can better capture the complex correlations between features and chest pain etiology.

Given the crucial role of rapid diagnosis of acute chest pain in improving prognosis, we are committed to building a more streamlined and suitable classification model for early diagnosis. Through feature importance analysis, we identify and remove redundant or minimally contributing features. After significantly reducing the number of numerical features to 30, the model's performance was not only maintained, but its accuracy even slightly improved. This move speeds up the diagnostic decision-making process for acute chest pain by significantly reducing the cost of time spent on clinical tests by retaining only the most relevant features. This streamlined model is more focused on key clinical factors, which improves the interpretability of the results, reduces the risk of model overfitting to the training data, and enhances the model's generalization ability and robustness. In addition, this process significantly reduces the computational cost and time required for model training and prediction.

The interpretability analysis using SHAP values revealed key features of model decision-making, such as D-dimer and hs-Tn biomarkers. The results were highly consistent with the value of these biomarkers in chest pain diagnosis confirmed by existing research, significantly enhancing the credibility and clinical practicality of the model results.

Most current research focuses on using AI technology to assist in the diagnosis of a single chest pain disease. For example, the ensemble learning framework constructed by

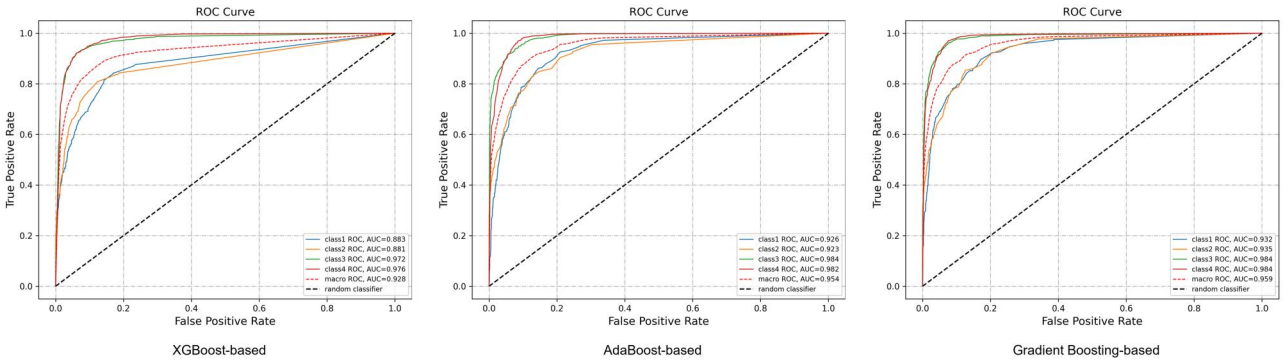


**Figure 3.** ROC curves of 3 multimodal models for each category. ROC: receiver operating characteristic.

**Table 4.** Prediction results of the Gradient boosting-based multimodal model on different number of features (mean and SD).

| Numerical features | Accuracy | F1-score | Precision | Recall | AUC |
|---|---|---|---|---|---|
| 30 features | 88.40 (0.007) | 74.56 (0.014) | 77.50 (0.018) | 72.52 (0.012) | 0.951 (0.004) |
| All features | 88.34 (0.004) | 74.83 (0.012) | 77.54 (0.009) | 72.91 (0.014) | 0.953 (0.005) |

AUC, area under the curve.

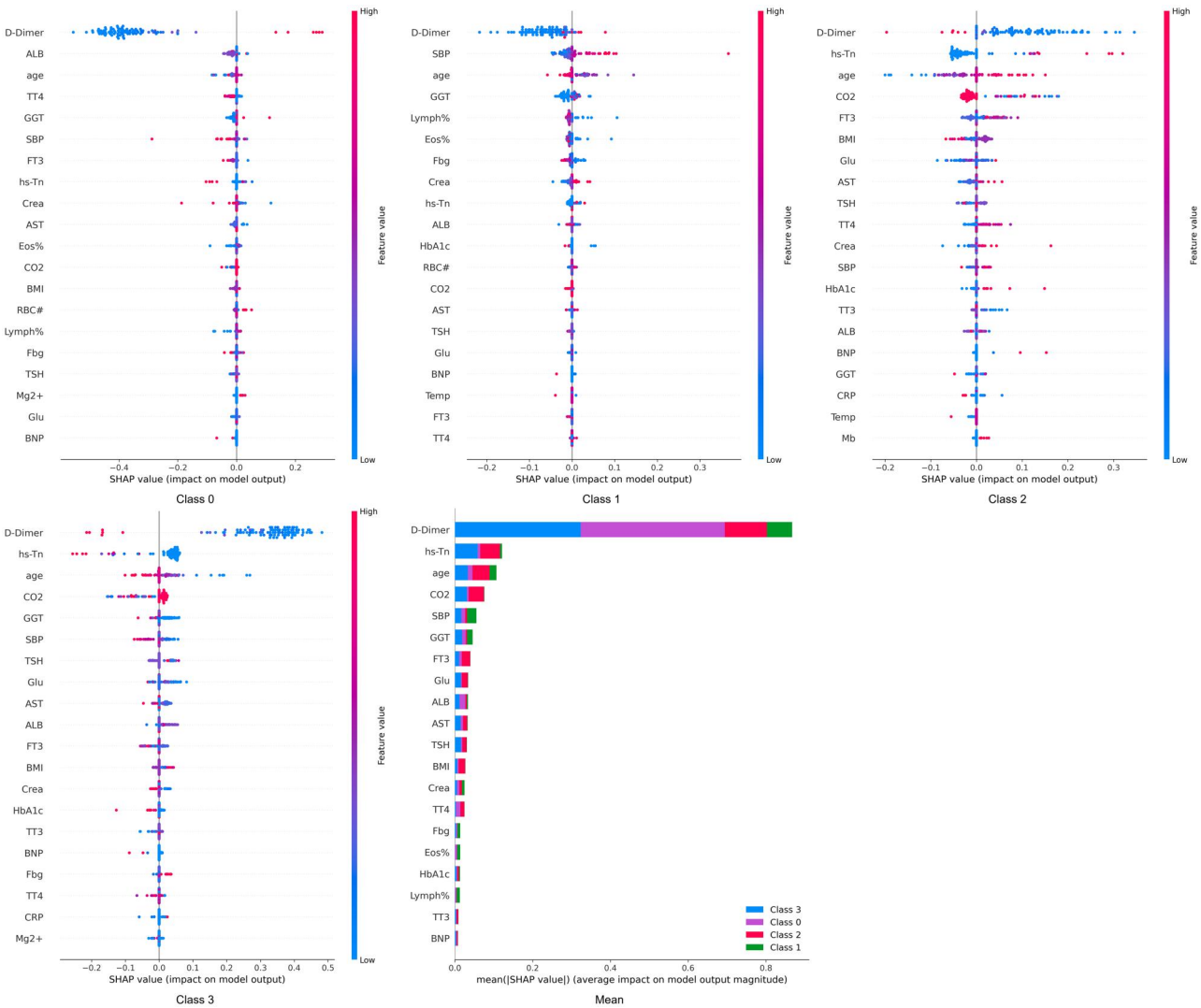


**Figure 4.** SHAP value analysis of multimodal models based on Gradient boosting. Among them, Class 0 to Class 3 correspond to the SHAP result graphs of the 4 types of chest pain causes, and Mean represents the average SHAP result graph. Each point in the figure corresponds to a sample in the training set and its SHAP value, and the variables are arranged in descending order of SHAP value. The horizontal position of the point represents the degree of influence of the feature value on the prediction result, with red indicating a higher feature value and the horizontal axis indicating the "positive" or "negative" impact of the feature on the prediction. SHAP: SHapley Additive exPlanations.

Emakhu et al[11] as a diagnostic aid for ACS can avoid misdiagnosis. It analyzed data from 31 228 patients and achieved an AUC value of 0.933. The Bayesian network model developed by Huo et al[19] was used for early diagnosis of aortic dissection, with a prediction accuracy of 84.55% and an AUC value of 0.857. Xi et al[9] developed a machine learning model based on 8 clinical features to assist in PE diagnosis, with an AUC value of 0.726, and compared it with the current clinical probability assessment model. These studies have made significant contributions to improving the accuracy of diagnosing a single cause of chest pain. However, due to the highly overlapping clinical manifestations of TRO, the limitations of single disease diagnostic research in practical applications are highlighted, making it difficult to meet the clinical demand for rapid differentiation of multiple chest pain diseases. This study innovatively focuses on acute chest pain, covering the classification of ACS, AD, PE, and Non-TRO diseases, which is more applicable in clinical practice and promises to provide comprehensive and efficient solutions for early triage of chest pain patients, with a significant value for clinical promotion and application. However, despite the better performance of the model, the possibility of misclassification still exists. Therefore, model outputs should be considered as one of the references for clinical decision-making rather than the sole basis. Clinicians should combine other clinical information and experience when using the model to ensure the accuracy and safety of decision-making.

However, this study also has certain limitations. First, although the overall performance of the model is good, especially in terms of AUC and accuracy, the clinical manifestations of chest pain triad are highly overlapping, which poses certain challenges to F1-score, precision, and recall metrics, making their performance relatively limited. Therefore, future research needs to further optimize feature engineering and model parameters, explore the integration of more potential biomarkers with clinical features, to continuously improve the diagnostic performance and generalization ability of the model.

Second, while we used routine interpolation when dealing with missing values in laboratory data, we did not adequately consider the nonrandom nature of missing data in acute care scenarios. As noted in a recent study,[31] tests are often performed selectively based on clinical suspicion, and simple interpolation may lead to bias. For this reason, in the future, we will analyze the mechanisms of missing data in depth, and further investigate the patterns of missing data through statistical analyses and recommendations from clinical experts to determine whether missing data are associated with the clinical presentation and initial diagnosis of patients. The use of multiple interpolation and machine learning-based predictive filling methods will also be explored, which can better take into account the underlying mechanisms of missing data and reduce the bias introduced by the mishandling of missing values.

Third, although a large number of features have been reduced through the feature selection work, some of the features are still difficult to obtain quickly in real clinical acute situations. Therefore, we should prioritize the acquisition of key features and gradually validate the effectiveness of the model in the clinical scenario. Currently, our feature selection work has been participated in and supported by 2 clinical experts. Going forward, we plan to further expand the involvement of clinical experts to ensure that the model design closely matches the actual clinical needs, which will cover several key aspects such as model interpretation, result validation, etc, thus introducing more clinical perspectives.

Finally, it is critical to test the performance of the model in clinical deployments, with particular attention to its false negative and false positive rates, to ensure its reliability and safety in real-world clinical settings. Therefore, external validation through larger sample sizes and multicenter studies is needed to apply the model to clinical applications.

## Conclusions

This study constructed a multimodal framework that integrates a pre-trained language model of Bio_ClinicalBERT with Gradient boosting ensemble learning, achieving efficient classification of the 3 major emergencies of TRO and Non-TRO etiology. The experiment based on 41 382 patients showed that the model maintained excellent performance even after numerical feature reduction to 30 key indicators. This framework is expected to overcome the difficult problem of highly overlapping clinical symptoms in TRO diseases and provide a fast and accurate intelligent auxiliary tool for emergency triage.

## Author contributions

Jun Tang (Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing), Fang Chen (Conceptualization, Formal analysis, Funding acquisition, Project administration, Resources, Supervision), and Dongdong Wu (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Writing—review & editing)

## Supplementary material

Supplementary material is available at *JAMIA Open* online.

## Funding

This work was supported by the National Natural Science Foundation of China (grant number: 72374206), the Artificial Intelligence Medical Research Project of Daping Hospital in Chongqing, China (grant numbers: ZXAIYB015 and ZXAIYB010), and the Basic Research Project of Daping Hospital in Chongqing, China (grant number: ZXJCYJ202414).

## Conflicts of interest

The authors have no competing interests to declare.

## Data availability

The dataset underlying this study cannot be publicly shared in a repository due to current restrictions imposed by the Chongqing Daping Hospital Institutional Review Board (IRB) and the participants did not consent to the public sharing of their data. However, the data are available upon request from the Chongqing Daping Hospital IRB for researchers who meet the criteria for accessing confidential information. Please contact the corresponding author to request the dataset.

## Ethics statement

This study complied with the Declaration of Helsinki and was approved by the Ethics Committee of Chongqing Daping Hospital (Ethics Approval No. 2025_131). Due to the retrospective nature of the research, the requirement for informed consent was waived.

## References


1. Mahmood SS, Levy D, Vasan RS, et al. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet* 2014;383:999-1008.
2. Dawson LP, Nehme E, Nehme Z, et al. Chest pain management using prehospital point-of-care troponin and paramedic risk assessment. *JAMA Intern Med*. 2023;183:203-211.
3. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. *Circulation*. 2022;145:e153-e639.
4. China Cardiovascular Health and Disease Report Writing Group. Report on cardiovascular health and diseases in China 2022: an updated summary. *Chin J Interv Cardiol*. 2023;31:485-508.
5. Pasteur-Rousseau A, Souibri K, Fouassier D, et al. Le scanner thoracique injecté « triple-rule-out » a-t-il un intérêt dans le bilan des douleurs thoraciques aiguës pour éliminer le syndrome coronaire aigu, l'embolie pulmonaire et la dissection aortique? [Benefits and drawbacks of CT scan as a triple rule-out exam in acute chest pain to exclude acute coronary syndrome, pulmonary embolism and aortic dissection]. *Ann Cardiol Angeiol (Paris)*. 2023;72:101641.
6. Halpern EJ. Triple-rule-out CT angiography for evaluation of acute chest pain and possible acute coronary syndrome. *Radiology* 2009;252:332-345.
7. Chae MK, Kim EK, Jung KY, et al. Triple rule-out computed tomography for risk stratification of patients with acute chest pain. *J Cardiovasc Comput Tomogr*. 2016;10:291-300.
8. Takeda M, Oami T, Hayashi Y, et al. Prehospital diagnostic algorithm for acute coronary syndrome using machine learning: a prospective observational study. *Sci Rep*. 2022;12:14593.
9. Xi L, Kang H, Deng M, et al. A machine learning model for diagnosing acute pulmonary embolism and comparison with wells score, revised Geneva score, and years algorithm. *Chin Med J (Engl)*. 2024;137:676-682.
10. Song M, Wang J, Yu Z, et al. PneumoLLM: harnessing the power of large language model for pneumoconiosis diagnosis. *Med Image Anal*. 2024;97:103248.
11. Emakhu J, Monplaisir L, Aguwa C, et al. Acute coronary syndrome prediction in emergency care: a machine learning approach. *Comput Methods Programs Biomed*. 2022;225:107080.
12. Kayvanpour E, Gi WT, Sedaghat-Hamedani F, et al. microRNA neural networks improve diagnosis of acute coronary syndrome (ACS). *J Mol Cell Cardiol*. 2021;151:155-162.
13. Chan PZ, Ramli MAIB, Chew HSJ. Diagnostic test accuracy of artificial intelligence-assisted detection of acute coronary syndrome: a systematic review and meta-analysis. *Comput Biol Med*. 2023;167:107636.
14. Laffafchi S, Ebrahimi A, Kafan S. Efficient management of pulmonary embolism diagnosis using a two-step interconnected machine learning model based on electronic health records data. *Health Inf Sci Syst*. 2024;12:17.
15. Ben Yehuda O, Itelman E, Vaisman A, et al. Early detection of pulmonary embolism in a general patient population immediately upon hospital admission using machine learning to identify new, unidentified risk factors: model development study. *J Med Internet Res* 2024;26:e48595.
16. Hu Z, Hu Y, Zhang S, et al. Machine-learning-based models assist the prediction of pulmonary embolism in autoimmune diseases: a retrospective, multicenter study. *Chin Med J (Engl)*. 2024;137:1811-1822.
17. Islam NU, Zhou Z, Gehlot S, et al. Seeking an optimal approach for computer-aided diagnosis of pulmonary embolism. *Med Image Anal*. 2024;91:102988.
18. Luo J, Zhang W, Tan S, et al. Aortic dissection auxiliary diagnosis model and applied research based on ensemble learning. *Front Cardiovasc Med*. 2021;8:777757.
19. Huo D, Kou B, Zhou Z, et al. A machine learning model to classify aortic dissection patients in the early diagnosis phase. *Sci Rep*. 2019;9:2701.
20. Liu L, Zhang C, Zhang G, et al. A study of aortic dissection screening method based on multiple machine learning models. *J Thorac Dis*. 2020;12:605-614.
21. Araoz PA, Gadam S, Bhanushali AK, et al. Triple rule out CT in the emergency department: clinical risk and outcomes (triple rule out in the emergency department). *Acad Radiol*. 2025;32:1297-1305.
22. Ibanez B, James S, Agewall S, et al.; ESC Scientific Document Group. 2017 ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: the task force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J*. 2018;39:119-177.
23. Konstantinides SV, Torbicki A, Agnelli G, et al. Task force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J*. 2014;35:3033-3069.
24. Erbel R, Aboyans V, Boileau C, et al.; ESC Committee for Practice Guidelines. 2014 ESC guidelines on the diagnosis and treatment of aortic diseases: document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The task force for the diagnosis and treatment of aortic diseases of the European Society of Cardiology (ESC). *Eur Heart J*. 2014;35:2873-2926.
25. Altmann A, Toloşi L, Sander O, et al. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26:1340-1347.
26. Alsentzer E, Murphy JR, Boag W, et al. Publicly available clinical BERT embeddings. 2019. Accessed September 19, 2025. https://doi.org/10.18653/v1/W19-1909
27. Mienye ID, Sun Y. A survey of ensemble learning: concepts, algorithms, applications, and prospects. *IEEE Access* 2022;10:99129-99149.
28. Hu J, Szymczak S. A review on longitudinal data analysis with Random Forest. *Brief Bioinform*. 2023;24:bbad002.
29. Asha SE, Miers JW. A systematic review and meta-analysis of d-dimer as a rule-out test for suspected acute aortic dissection. *Ann Emerg Med*. 2015;66:368-378.
30. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc.; 2017:4768-4777.
31. Overmars LM, Niemantsverdriet MSA, Groenhof TKJ, et al. A wolf in sheep's clothing: reuse of routinely obtained laboratory data in research. *J Med Internet Res*. 2022;24:e40516.



JAMIA Open, 2025, 8, 1–8
https://doi.org/10.1093/jamiaopen/ooaf114
Research and Applications