Received: 15 March 2026 | Revised: 3 August 2026 | Accepted: 20 August 2026 | Published online: 16 September 2026

RESEARCH ARTICLE

Medinformatics
2026, Vol. 00(00) 1–9
DOI: 10.47852/bonviewMEDIN62029651

BON VIEW PUBLISHING

Analysis of Interpretable Machine Learning Classifiers for Myocardial Infarction Detection Using 12-Lead ECG Features

Chutian Chen[1], Valentina L. Kouznetsova[1,2,3] and Igor F. Tsigelny[1,2,3,4,*]

[1] *BiAna Institute, USA*

[2] *CureScience Institute, USA*

[3] *San Diego Supercomputer Center, University of California San Diego, USA*

[4] *Department of Neurosciences, University of California San Diego, USA*

Abstract: Swift identification of cardiac events is crucial for enhancing patient outcomes. Electrocardiograms (ECGs) are noninvasive and can be quickly taken. Machine learning (ML) can then be used to assist in recognizing possible myocardial infarction. ECG recordings were extracted from a public database, and features related to heart rate variability and morphology were selected as descriptors for ML. The extracted data were preprocessed and classified using Waikato Environment for Knowledge Analysis. The Random Forest and Logistic Model Tree ML models were the best-performing models, achieving an accuracy of 87.40%. Other models, NaïveBayes, J48, and Reduced Error Pruning Tree, achieved accuracies of 86.86%, 81.23%, and 80.16%, respectively. Additionally, BayesNet and Logistic achieved accuracies of 86.33% and 84.99%, respectively. To validate the models internally, new data were taken from a completely new group of ECG recordings, and then the models were evaluated on each new internal independent dataset with unseen ECG data. BayesNet achieved an accuracy of 84.76%, and Random Forest managed to achieve 84.49% accuracy on the first internal independent dataset. These results show the power of ML in the rapid diagnosis of myocardial infarction and how it may aid in an emergency where quick diagnosis and identification are crucial.

Keywords: myocardial infarction, electrocardiography, interpretable machine learning, feature selection

1. Introduction

Myocardial infarction (MI) is one of the leading causes of death globally and requires immediate medical attention to be properly treated [1]. MI occurs when blood flow to part of the heart is blocked, which severely limits the heart's ability to operate [2]. Without speedy restoration of blood flow to the heart, major damage to the organ and cell death can occur within just a couple of minutes [2].

MI diagnosis usually involves symptom evaluation, blood tests, and an electrocardiogram (ECG) recording [3]. Between these three options, an ECG recording is typically the first test performed during diagnosis. The recording can then be analyzed for patterns that could indicate a heart attack [4]. Commonly, MI ECGs contain ST-segment elevation, abnormal Q waves, or T-wave changes [5]. Despite this, it is often difficult for humans and computers to agree on a diagnosis, with agreements only occurring around 54.1% of the time [6].

Machine learning (ML) has already shown promise in analyzing an ECG recording to make a heartbeat classification [7]. The model has demonstrated accuracy ranging from 68% to 99.85% for arrhythmia [7]. Another study used ECG graph images that achieved a test accuracy rate of 88.79% using the Gradient Boosting Classifier for detecting MI [8].

In recent years, ML and deep learning have also emerged as formidable tools for analyzing and interpreting ECGs, particularly demonstrating strong potential for distinguishing MI from normal heart rhythms. A 2022 review of 59 different deep learning studies revealed that convolutional neural networks (CNNs) are the most popular and also the most accurate, with accuracies reaching beyond 97% [9]. Another study that took a hybrid approach found that a deep learning approach with artificial neural network models also provides accuracy around 93.6% [10].

Artificial intelligence (AI) also has the ability to come to conclusions with better accuracy than traditional clinical methods. A 2023 study showed that ML models used for the ECG diagnosis of occlusion MI could outperform practicing clinicians and other

***Corresponding author:** Igor F. Tsigelny, BiAna Institute, CureScience Institute, San Diego Supercomputer Center, University of California San Diego and Department of Neurosciences, University of California San Diego, USA. Email: itsigeln@ucsd.edu

commercial interpretation methods with better precision and sensitivity [11]. Another study from 2020 that used ML to predict acute coronary syndrome found that the final model was 52% more sensitive than commercial interpretation software and 37% more sensitive than experienced clinicians [12]. Yet another study in 2025 also found that AI-enhanced ECGs could detect acute MI with similar or better accuracy than traditional risk stratification methods and physicians [13]. However, despite these methods becoming more advanced and accurate, there are ethical concerns about implementing these models in practice, as their algorithms are not completely transparent [14].

This field continues to improve with a study from 2024 using a multitask deep learning algorithm that performed better than single-task algorithms as it reached over 90% accuracy [15]. Another 2025 study that utilized demographics such as age and sex as well as symptoms in a CNN algorithm also proved very successful [16].

Even though there has been a heavy focus on complex methods such as deep learning and AI, classical ML models remain relevant because of their transparency and efficiency. In this study, we focus specifically on features related to heart rate variability (HRV) and morphology extracted from PTB databases to train and evaluate the accuracy of classifiers such as Random Forest and other algorithms to continue to show the power of ML in the accurate diagnosis of MI. The primary purpose of this paper is to challenge the current research landscape that commonly uses uninterpretable deep learning models by using more interpretable classification models that have more transparency compared to the nontransparent nature of deep learning models. Interpretability of these models is significant because interpretability means a human can verify the model's process and conclusion since these classifiers show feature importances that are associated with existing indicators. On the other hand, deep learning models have millions of internal calculations that don't have the ability to be associated with meaningful features.

2. Materials and Methods

2.1. Database

The ECG dataset used in this study is from PTB-XL, a large publicly available electrocardiography database (version 1.0.3, 2022). ECGs are a vital diagnostic tool that assesses a patient's cardiac condition. The dataset includes a total of 21,799 12-lead ECGs, where 9514 records are labeled as normal and 5469 records are labeled as MI, indicating a large number of normal controls as well as MI records. The data were gathered using devices from Schiller AG from October 1989 to June 1996 [17–19]. Metadata, including age, sex, and weight, for each patient was also collected.

2.2. Waikato Environment for Knowledge Analysis

The Waikato Environment for Knowledge Analysis (WEKA) was developed by the University of Waikato in New Zealand. WEKA includes a variety of ML classification algorithms and data processing tools, as well as support for visualization of data [20]. In this study, WEKA's ML algorithms and attribute/descriptor filters were heavily used. The attribute filter feature was used to exclude descriptors that could possibly confuse the models and/or descriptors that aren't relevant. Finally, WEKA's ML algorithms were used for training the data and testing the models on internal independent data.

2.3. Patient selection

The PTB-XL database assigns each patient an SCP-ECG category that can be used to determine the patient's diagnosis. Only patients categorized with one of the nine categories related to infarction ("IMI," "ASMI," "ILMI," "AMI," "ALMI," "LMI," "IPLMI," "IPMI," "PMI") were extracted alongside normal subjects (Table 1).

Table 1. SCP statements and their corresponding descriptions

SCP statement	Description
IMI	Inferior myocardial infarction
ASMI	Anteroseptal myocardial infarction
ILMI	Inferolateral myocardial infarction
AMI	Anterior myocardial infarction
ALMI	Anterolateral myocardial infarction
LMI	Lateral myocardial infarction
IPLMI	Inferoposterolateral myocardial infarction
IPMI	Inferoposterior myocardial infarction
PMI	Posterior myocardial infarction

In the training set, a total of 373 patients were selected from the first 1000 patients in the database. The training set consisted of 195 normal subjects and 178 patients with infarction. The 195 normal subjects were the first 195 subjects in the 1000-patient subset, and every infarction patient in the subset was included in the training set. Only 373 patients were selected as a result of balancing the number of infarction and normal subjects to train the model on, and these came from the first block of 1000 patients, which is used as the training group in this study. The three blocks of 1000 patients were used instead of the full 21,799 ECG record database because of the intense computational load of feature extraction through NeuroKit2. The model was trained on a balanced set of classes because balancing is necessary to stop the model from ignoring the smaller group of MI cases. Without balancing, we found that the model heavily favors normal diagnosis and is more likely to miss MI cases. In other words, the model would output a high accuracy while marking a large number of MI cases as normal. Therefore, while this ratio might not reflect real-world ratios, it is necessary for model training. As for the selection of specific patients within the entire database, the selected records were compared with the unselected normal and MI records to determine if the selected set was representative of the whole database. The results are shown in the following table.

Because of the inherent nature of traditional tests such as the chi-square test that automatically amplify the test statistic as sample size increases, the standardized mean difference (SMD) and Cramér's V were used instead because of the significant difference in n. As seen in Table 2, both V values are less than 0.05, which points to a negligible association, and an SMD value of + 0.04 also means there is a negligible difference. These findings show that there is no real evidence suggesting that the selected records are unrepresentative of the database as a whole in terms of age, sex, and type of MI. Diving further into age, the normal record to MI age gap for the selected subset sits at around 22.5 years; however, this does narrow to 13.5 years in the unselected subset. After training and testing the models without age as an attribute, the original training set saw negligible change in accuracy. For example, Random Forest accuracy

Table 2. Selected subset versus remaining subset

Characteristic	Selected (n = 1119)	Unselected (n = 13,682)	Result
Age (mean ± SD)	57.7 ± 19.7	57.0 ± 16.9	SMD = + 0.04
Sex (% Female)	51.7%	47.6%	V = 0.022
Anterior localization	53.6%	45.7%	
Inferior localization	43.6%	50.9%	V = 0.048
Lateral or posterior localization	2.8%	3.4%	

*Some records carry multiple infarction codes. Compound codes were assigned to their primary territory. SMD and V are used because they are independent of sample size.

increased slightly by 1.07 percentage points, and NaiveBayes accuracy increased by 1.88 percentage points, which is a difference of 4 and 7 ECGs, respectively, out of 373. Similarly, when tested on both independent datasets, Random Forest gained 1.34 percentage points of accuracy for independent dataset 1 but lost about 1 percentage point of accuracy in independent dataset 2. This is only a difference of 5 and 4 ECGs, respectively, showing that model performance was based on ECG features instead of patient age, despite the fact that patient age differs demographically between the selected subset and the remainder of the database.

2.4. Feature extraction

The Python library Neurokit2 was used for the extraction of data [21]. The ECG recordings were first cleaned using a 0.5 Hz high-pass Butterworth filter (order = 5), followed by powerline filtering (powerline = 50) [21]. HRV metrics are then extracted from the cleaned signal as well as the location of Q, R, S, and T Peaks. The peaks are then used to calculate morphological metrics. The sex and the age of each patient were also extracted. Features were extracted from all 12 leads of each ECG. The exact same features were extracted from each lead.

2.5. Feature selection

These metrics were used as descriptors in WEKA and were filtered using WEKA's built-in AttributeSelection filter on the training set. Correlation-based feature subset selection (CfsSubsetEval) was used as the evaluator, and best-first search (BestFirst) was used as the search method [20]. CfsSubsetEval finds subsets where descriptors correlate with the class, but not with each other. Selection stops when new selected attributes stop improving the subset. The filter did not target a certain number of attributes. After running this filter, the starting 830 candidate descriptors were reduced to 44 (Table 3).

Table 3. Attribute abbreviation meanings

Attribute abbreviation	Meaning
i_HRV_HTI	HRV Triangular Index
i_mean_qrs_duration	Mean QRS complex width
i_mean_qt_interval	Mean QT interval
i_mean_st_elevation	Mean of zero-centered signal in 20–80 ms window after each S peak
i_mean_t_amplitude	Mean T-wave amplitude
ii_mean_qrs_duration	Mean QRS complex width
ii_mean_qt_interval	Mean QT interval
ii_mean_st_elevation	Mean ST-segment elevation
ii_mean_t_amplitude	Mean T-wave amplitude
iii_mean_st_elevation	Mean ST-segment elevation
avr_HRV_MCVNN	Median-based coefficient of variation of NN intervals
avr_HRV_PIP	Percentage of Inflection Points
avl_mean_qrs_duration	Mean QRS complex width
avl_mean_st_elevation	Mean ST-segment elevation
avf_HRV_MaxNN	Maximum NN interval
avf_HRV_SDNNd	SD of NN intervals on decelerations
avf_HRV_DFA_alpha1	Detrended Fluctuation Analysis (DFA) short-term scaling exponent
avf_mean_qt_interval	Mean QT interval
v1_HRV_MinNN	Minimum NN interval
v1_HRV_SD2	Long-term variability axis of the Poincaré plot
v1_HRV_ApEn	Approximate Entropy
v1_mean_qrs_duration	Mean QRS complex width

(*Continued*)

Table 3. (*Continued*)

Attribute abbreviation	Meaning
v1_mean_st_elevation	Mean ST-segment elevation
v2_HRV_SDSD	SD of successive NN-interval differences
v2_HRV_CSI_Modified	Modified Cardiac Sympathetic Index
v2_HRV_ShanEn	Shannon Entropy of the interval distribution
v2_mean_qrs_duration	Mean QRS complex width
v3_HRV_SD2a	SD2 computed on the acceleration side of the Poincaré plot
v4_HRV_CD	Correlation Dimension
v4_mean_qt_interval	Mean QT interval
v4_mean_t_amplitude	Mean T-wave amplitude
v5_mean_qt_interval	Mean QT interval
v5_mean_t_amplitude	Mean T-wave amplitude
v6_mean_qrs_duration	Mean QRS complex width
v6_mean_qt_interval	Mean QT interval
v6_mean_st_elevation	Mean ST-segment elevation
v6_mean_t_amplitude	Mean T-wave amplitude
avr_HRV_MFDFA_alpha1_Fluctuation	Multifractal DFA (alpha1)—overall fluctuation magnitude
v1_HRV_MFDFA_alpha1_Increment	Multifractal DFA (alpha1)—successive increment of the spectrum
v2_HRV_MFDFA_alpha1_Peak	Multifractal DFA (alpha1)—Hurst exponent at the spectrum's peak
v5_HRV_MFDFA_alpha1_Delta	Multifractal DFA (alpha1)—height difference across the spectrum
v6_HRV_MFDFA_alpha1_Width	Multifractal DFA (alpha1)—width of the multifractal spectrum
i_HRV_MFDFA_alpha1_Delta	Multifractal DFA (alpha1)—height difference across the spectrum
age	Patient age

When evaluating MI, cardiologists also examine key features in the ECG recording. These features include hyperacute T waves, ST-segment elevation, pathological Q waves, T-wave inversion, and reciprocal ST-segment depression. WEKA's attribute selection filter identified the majority of these features. The mean T amplitude attribute corresponds to both hyperacute T waves and T-wave inversion, and the mean ST elevation attribute corresponds to the ST-segment elevation and depression (Figure 1).

2.6. Classification

Three datasets were created, containing the extracted features: one for training and two for internal independent validation of the model trained on the original training set. The extracted data were used to create an ML model. All classifiers were used in WEKA with default parameters [20]: Random Forest [22], Logistic Model Tree (LMT) [23], NaïveBayes, J48, Reduced Error Pruning Tree (REPTree), BayesNet, and Logistic (logistic regression). Each classifier was trained only on the same training dataset that was extracted from the PTB-XL database.

2.7. Testing on internal independent data

There were two internal independent datasets on which the model trained on the training set was tested. Both of these independent datasets contain only unique recordings that were not used in the training set, as independent dataset 1 (n = 374) and independent dataset 2 (n = 372) were taken from the second and third cohort of 1000 patients in the database, respectively. WEKA's built-in test set feature was utilized for this purpose.

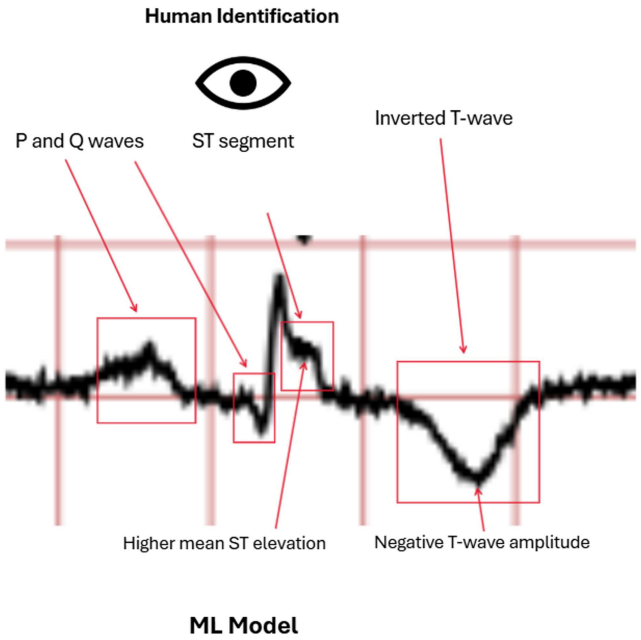

Figure 1. Important features when diagnosing myocardial infarction. This image compares the features cardiologists generally examine (eyeball) and the attributes in the ML model that correspond to them. For example, ST-segment elevation would result in a greater value of mean ST elevation, and an inverted T-wave would result in a smaller or negative value of T-wave amplitude. In this particular ECG, the ST segment is heavily elevated, which suggests an acute myocardial infarction

Each of the independent datasets was set as the supplied test set in WEKA. All seven models that were built on the original training data were then re-evaluated on both the new independent datasets within WEKA, and the accuracy of each was noted.

3. Results and Discussion

3.1. Results

The extracted data from the PTB-XL database were used to create seven different models using these seven classifiers: Random Forest, LMT, NaïveBayes, BayesNet, Logistic, J48, and REPTree. 10-fold cross-validation was used for each one as well. The Random Forest and LMT models performed the best, with an accuracy of 87.40%, followed by NaïveBayes, BayesNet, Logistic, J48, and REPTree, each with an accuracy of 86.86%, 86.33%, 84.99%, 81.23%, and 80.16%, respectively (Figure 2).

Furthermore, the highest area under the ROC curve (AUC-ROC) was 0.9445 for the LMT classifier. Other classifiers had similar AUC-ROC (Figure 3).

Each of these models was then tested on completely new data from a different subset of recordings in the same database to internally validate the data. Two internal independent datasets were extracted. Independent dataset 1 consisted of 374 patients and was derived from a subset of normal subjects and infarction patients within the second group of 1000 patients from the PTB-XL database. Independent dataset 2 consisted of 372 patients and was derived from a subset of normal subjects and infarction patients within the third group of 1000 patients from the PTB-XL database. Independent dataset 1 exhibits extremely similar average and standard deviation in the extracted lead II morphological features compared to those of the original training dataset. Only lead II comparisons are used here for simplicity. The QRS duration, QT interval, ST elevation, and T amplitude in the training data were 0.126 ± 0.042, 0.298 ± 0.054, -0.057 ± 0.043, and 0.144 ± 0.107 (mean $\pm$ standard deviation), respectively. The QRS duration, QT interval, ST elevation, and T amplitude in independent dataset 1 were 0.127 ± 0.046, 0.301 ± 0.05, -0.059 ± 0.043, and 0.149 ± 0.104 (mean $\pm$ standard deviation), respectively. Independent dataset 2 also had similar numbers (Table 4). The training set, independent dataset 1, and independent dataset 2 contained 373, 374, and 372 patients, respectively. Each set consisted of 195 normal subjects, which were the first 195 subjects in the second group of 1000 patients in the PTB-XL database for independent dataset 1 and the first 195 subjects in the third group of 1000 patients for independent dataset 2. There were different numbers of infarction patients in each set. Independent dataset 1 and independent dataset 2 had 179 and 177 infarction patients, respectively.

For independent dataset 1, the Random Forest classifier achieved an accuracy of 84.49%, LMT achieved an accuracy of 81.02%, NaïveBayes had an accuracy of 82.62%, BayesNet had the highest accuracy at 84.76%, Logistic had an accuracy of 81.82%, J48 had an accuracy of 77.54%, and REPTree had an accuracy of 79.41% (Figure 4).

Validating using an independent dataset 2 achieved similar accuracies: Random Forest achieved an accuracy of 83.60%, LMT had an accuracy of 81.18%, and NaïveBayes, BayesNet, Logistic, J48, and REPTree achieved accuracies of 83.33%, 83.33%, 79.30%, 73.66%, and 72.04%, respectively (Figure 5).

There are certain features of the classifier parameters that indicate MI. Here, we will focus specifically on the NaïveBayes model due to its easily readable parameters. Unsurprisingly, T-wave amplitude was the biggest sign of MI. In lead II, the mean T-wave amplitude fell from 0.199 in normal subjects to 0.084 in MI, with similar drops in leads I (0.167 → 0.049), V5 (0.263 → 0.107), and V6 (0.203 → 0.075), showing T-wave inversion and flattening across the board. Additionally, the QRS duration consistently widens across every lead as well, with examples in lead II where the duration increased from 0.113 to 0.141 and in aVL where it increased from 0.129 to 0.183. ST elevation is lower in nearly every lead with MI, with the exception of lead II, where it was unchanged, and the QT interval is also slightly shorter in MI, but these are much weaker indications. Moving onto multifractal features, the parameters show the avf_HRV_DFA_alpha1 attribute decreases, and therefore, MI shifts the short-term fractal scaling exponent toward randomness. The parameters also show that v6_HRV_MFDFA_alpha1_Width, avr_HRV_MFDFA_alpha1_Fluctuation, and v1_HRV_MFDFA_alpha1_Increment increase, which indicates a wider, more fluctuating multifractal spectrum. Additionally, the v2_HRV_MFDFA_alpha1_Peak and v5_HRV_MFDFA_alpha1_Delta also decrease, which means the scaling exponent decreases. MI broadens the temporal spectrum because the rhythm's regularity decreases; however, these trends are weaker but still directionally consistent.

3.2. Discussion

Currently, most published work using PTB-XL involves deep learning with interpretability being addressed afterward through different post hoc methods. This study differs as it not only

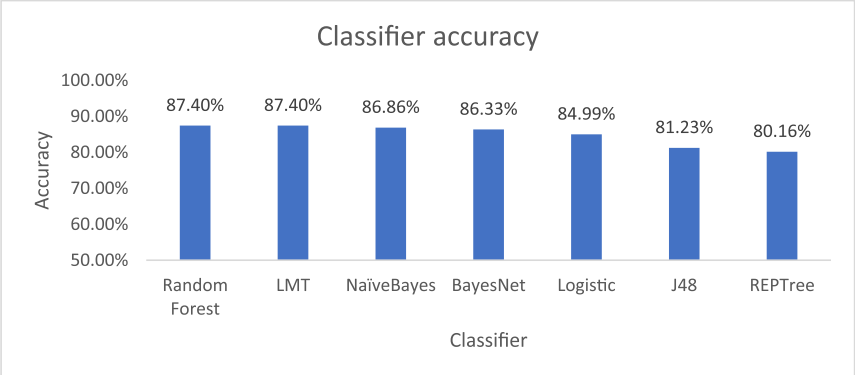

Figure 2. Accuracy of each model trained on the PTB-XL database

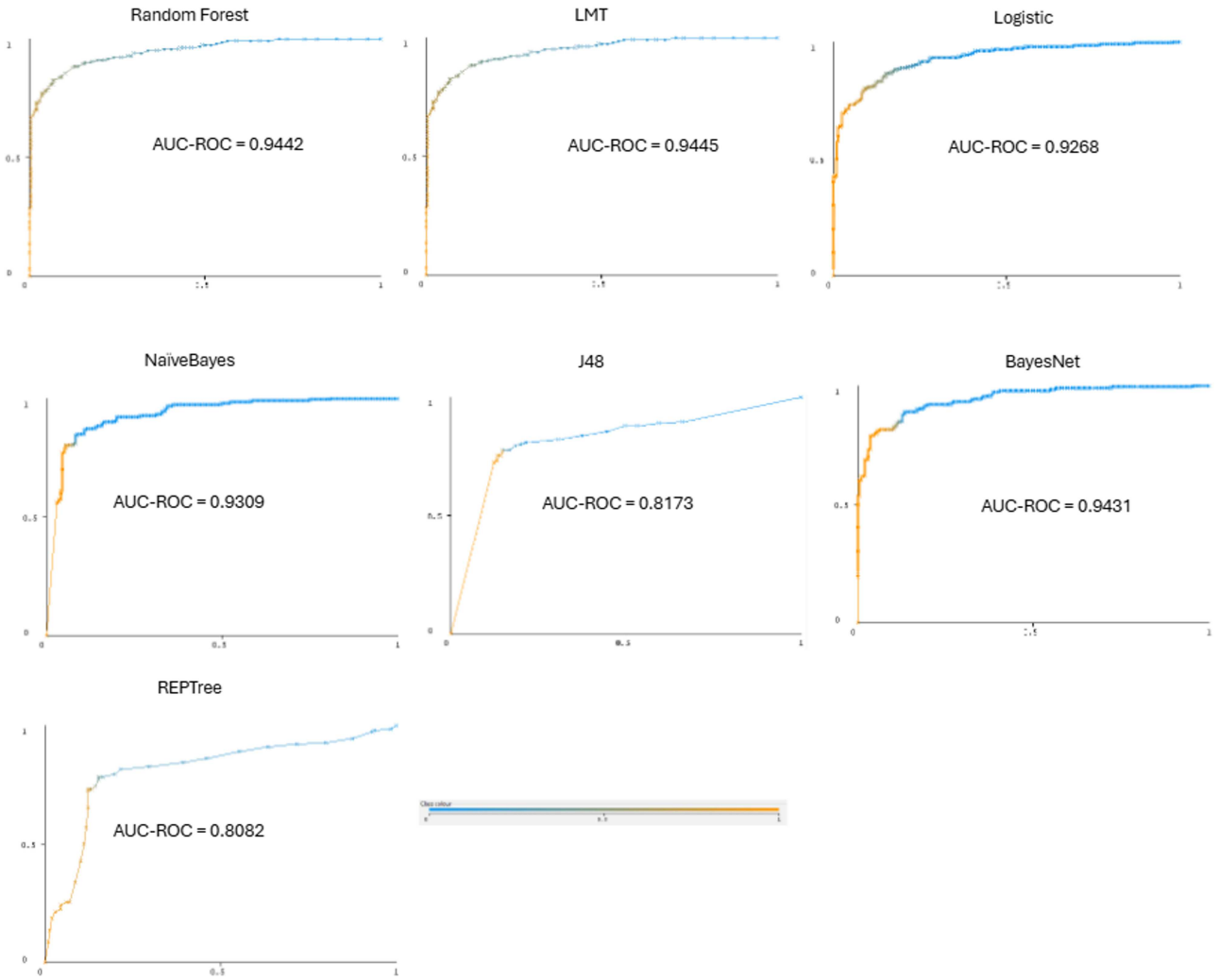

Figure 3. ROC curve for each model with its area under the curve (AUC-ROC). The Random Forest classifier has an AUC-ROC of 0.9442. The LMT classifier has an area under the curve (AUC) of 0.9445. The NaïveBayes classifier has an area of 0.9309. The J48 classifier has an area of 0.8173. The REPTree classifier has an area of 0.8082. The Logistic classifier has an area of 0.9268. The BayesNet classifier has an area of 0.9431. The color represents the threshold, and the color scale is shown

Table 4. Key attributes of each dataset include morphological features and patient numbers

	Training set	Independent dataset 1	Independent dataset 2
Normal subjects	195	195	195
Infarction patients	178	179	177
QRS duration	0.126 ± 0.042	0.127 ± 0.046	0.128 ± 0.044
QT interval	0.298 ± 0.054	0.301 ± 0.05	0.299 ± 0.058
ST elevation	–0.057 ± 0.043	–0.059 ± 0.043	–0.058 ± 0.043
T amplitude	0.144 ± 0.107	0.149 ± 0.104	0.135 ± 0.101

uses ECG features but also has intrinsic interpretability, meaning that models are trained on named descriptors with the model parameters providing justification (Table 5).

In terms of interpretability, the difference between post hoc and intrinsic interpretability is that a clinician can read our models and directly see differences in certain descriptors between MI and normal subjects, while a saliency map, for example, can only show locations considered by a neural network, not what was measured there.

One of the key advantages of the Random Forest classifier is the mitigated risk of overfitting because of the use of multiple trees. Because ECGs often contain a lot of noise and exhibit very slight differences, this is a significant benefit that contributes to their high performance. Other models, such as REPTree, are single decision tree algorithms that are prone to overfitting and result in lower accuracy. This is the same case for J48, which is the reason for its lower accuracy in the internal independent datasets. Specifically in independent dataset 2, J48 and REPTree

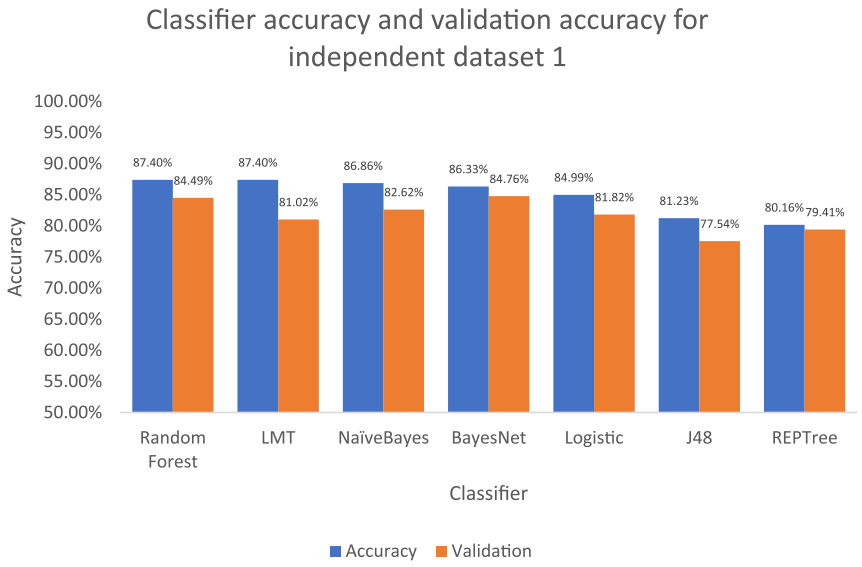

Figure 4. Accuracy of each model trained on the data from the PTB-XL database, along with the internal validation accuracy for the first independent dataset

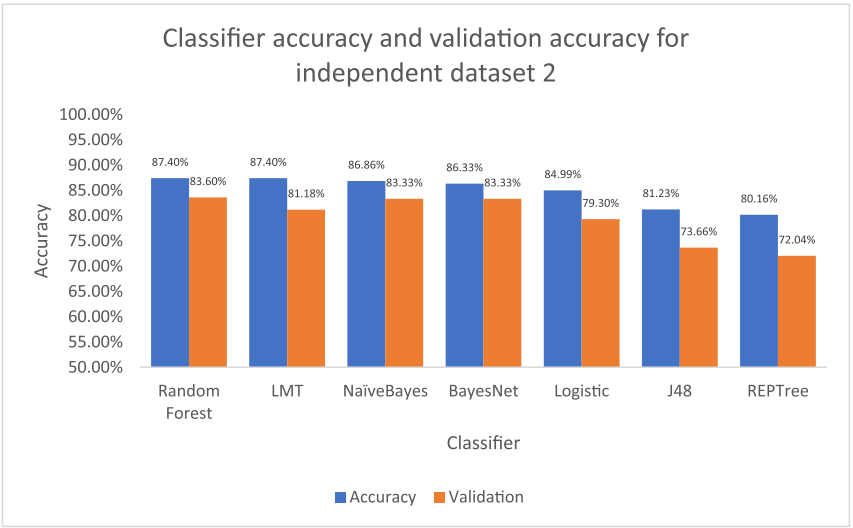

Figure 5. Accuracy of each model trained on the data from the PTB-XL database, along with the internal validation accuracy for the second independent dataset

Table 5. Comparison to other studies

Study	Input	Model	Interpretability	Reported performance
Strodthoff et al. [24]	Raw waveform	CNN/ResNet	Post hoc through input attribution	AUC = ~0.93
Xiong et al. [9]	Raw waveform	CNN review	None	Up to > 97% accuracy
Zeng et al. [25]	Raw waveform	ECA-Net, SENet, ResNet	Post hoc SHAP and Grad-CAM	88.27% accuracy
Tanyel et al. [26]	ECG features	XGBoost	Post hoc counterfactual reasoning	Not directly comparable
This work	44 named HRV, morphological and MFDFA descriptors	Random Forest, LMT, BayesNet, NaiveBayes, etc.	Intrinsic through the model parameters	87.40% accuracy, AUC 0.9445

*Performance is not directly comparable as class balance and metrics differ.

have potential overfitting problems that were predicted before, while the other models maintain stable accuracy between independent dataset 1 and independent dataset 2 with much more minor drops in performance. Another notable model is BayesNet, which achieved the most stability across the training and internal validation sets. It even edges out NaïveBayes in the first validation set. Because BayesNet doesn't assume that each feature is independent, it makes sense that it has more stability in terms of accuracy compared to NaïveBayes.

Our best-performing model, which utilizes the Random Forest classifier, achieves an accuracy of 87.40%, demonstrating the robustness of the classifier even when the data are very noisy. This demonstrates the potential for utilizing ECGs in the diagnosis of other diseases beyond MI.

Pruning attributes revealed the correlation of Multifractal Detrended Fluctuation Analysis (MFDFA) features to MI. This matches findings in another paper that used MFDFA to analyze ECG signals in the elderly [27]. In addition to the already known signs of MI, such as ST elevation, these attributes were major contributors to the accuracy of the models.

Since this study only utilized the PTB-XL ECG database, each recording was only 10 s long, which limited the number of attributes that could be calculated due to the shorter length. Another key limitation is the source of the validation datasets. While they are still unique to the training set and completely isolated, they still come from the same database, which falls short of true external validation. Additionally, patients were selected in order from the database. It was shown that they are demographically very similar, but the hardware used to take the recordings in the selected subset didn't include the CS100 machine, which makes up about 28.7% of records in the remainder of the database. However, signal-quality annotations were reviewed, and the selected records had 22.1% of records labeled with either baseline drift, static or burst noise, or electrode problems, which is very similar to the 23.2% of the remaining records that carry such annotations. This indicates that the difference in devices doesn't relate to a difference in recording quality. Nevertheless, performance on hardware not represented in the selected subset remains untested.

4. Conclusion

In this study, seven interpretable classifiers were run on 44 HRV, morphological, and multifractal features that were extracted from 12-lead PTB-XL ECG recordings. Random Forest and LMT achieved accuracies of 87.40% in 10-fold cross-validation. On the two internal validation sets, Random Forest maintained 84.49% and 83.60% accuracy, and BayesNet had similar results at 84.76% and 83.33% accuracy on validation sets 1 and 2, respectively. The attributes that were selected correspond to signs that cardiologists already use without being directed to. Limitations include the absence of external validation, as only a single database is used, and the 10-second recording lengths. The interpretability of classifiers is important when reasoning must be verified by a human. External validation on a different database is the next step.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors. No new patient data were collected, and all electrocardiographic recordings were obtained from PTB-XL version 1.0.3. The data in PTB-XL were approved for publication by the institutional ethics committee (reference PTB-2020-1) [17]. The records were also fully anonymized.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in PhysioNet at https://doi.org/10.13026/kfzx-aw45 (PTB-XL v1.0.3).

Author Contribution Statement

Chutian Chen: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Valentina L. Kouznetsova:** Conceptualization, Writing – review & editing, Visualization, Supervision. **Igor F. Tsigelny:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration.

References

[1] Thygesen, K., Alpert, J. S., Jaffe, A. S., Chaitman, B. R., Bax, J. J., Morrow, D. A., ..., & Executive Group on behalf of the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC)/American Heart Association (AHA)/World Heart Federation (WHF) Task Force for the Universal Definition of Myocardial Infarction. (2018). Fourth universal definition of myocardial infarction (2018). *Circulation*, *138*(20), e618–e651. https://doi.org/10.1161/CIR.0000000000000617

[2] Ojha, N., & Dhamoon, A. S. (2023). Myocardial infarction. In *StatPearls [Internet]*. StatPearls Publishing.

[3] Byrne, R., Rossello, X., Coughlan, J. J., Barbato, E., Berry, C., Chieffo, A., ..., & Varvarovsky, I. (2024). 2023 ESC Guidelines for the management of acute coronary syndromes. *European Heart Journal*, *44*(38), 3720–3826. https://doi.org/10.1093/eurheartj/ehad191

[4] Sattar, Y., & Chhabra, L. (2023). Electrocardiogram. In *StatPearls [Internet]*. StatPearls Publishing.

[5] Morris, F., & Brady, W. J. (2002). ABC of clinical electrocardiography: Acute myocardial infarction—Part I. *BMJ*, *324*(7341), 831–834. https://doi.org/10.1136/bmj.324.7341.831

[6] Iftikhar, A., Bond, R., Mcgilligan, V., Leslie, S. J., Knoery, C., Shand, J., ..., & Peace, A. (2021). Human–computer agreement of electrocardiogram interpretation for patients referred to and declined for primary percutaneous coronary intervention: Retrospective data analysis study. *JMIR Medical Informatics*, *9*(3), e24188. https://doi.org/10.2196/24188

[7] Aziz, S., Ahmed, S., & Alouini, M. S. (2021). ECG-based machine-learning algorithms for heartbeat classification. *Scientific Reports*, *11*(1), 18738. https://doi.org/10.1038/s41598-021-97118-5

[8] Raghukumar, B. S., & Naveen, B. (2024). Enhancing myocardial infarction diagnosis: Insights from ECG image analysis and machine learning. *SN Computer Science*, *5*(5), 448. https://doi.org/10.1007/s42979-024-02827-z

[9] Xiong, P., Lee, S. M. Y., & Chan, G. (2022). Deep learning for detecting and locating myocardial infarction

by electrocardiogram: A literature review. *Frontiers in Cardiovascular Medicine*, *9*, 860032. https://doi.org/10.3389/fcvm.2022.860032

[10] Dhandapani, S., Somasundaram, H., & Angamuthu, T. (2025). Hybrid deep learning framework for heart disease prediction using ECG signal images. *Scientific Reports*, *15*(1), 33922. https://doi.org/10.1038/s41598-025-10062-6

[11] Al-Zaiti, S. S., Martin-Gill, C., Zègre-Hemsey, J. K., Bouzid, Z., Faramand, Z., Alrawashdeh, M. O., ..., & Callaway, C. W. (2023). Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. *Nature Medicine*, *29*(7), 1804–1813. https://doi.org/10.1038/s41591-023-02396-3

[12] Al-Zaiti, S., Besomi, L., Bouzid, Z., Faramand, Z., Frisch, S., Martin-Gill, C., ..., & Sejdić, E. (2020). Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nature Communications*, *11*(1), 3966. https://doi.org/10.1038/s41467-020-17804-2

[13] Lee, M. S., Shin, T. G., Lee, Y., Kim, D. H., Choi, S. H., Cho, H., ..., & Kim, K. (2025). Artificial intelligence applied to electrocardiogram to rule out acute myocardial infarction: The ROMIAE multicentre study. *European Heart Journal*, *46*(20), 1917–1929. https://doi.org/10.1093/eurheartj/ehaf004

[14] Panjiyar, B. K., Davydov, G., Nashat, H., Ghali, S., Afifi, S., Suryadevara, V., . . . , & Arcia Franchini, A. P. (2023). Correction: A systematic review: Do the use of machine learning, deep learning, and artificial intelligence improve patient outcomes in acute myocardial ischemia compared to clinician-only approaches? *Cureus*, *17*(9), c307. https://doi.org/10.7759/cureus.c307

[15] Park, B. E., Shon, B., Cho, J., Jung, M. S., Park, J. S., Kim, M. S., . . . , & Jang, S. Y. (2025). Signal-guided multitask learning for myocardial infarction classification using images of electrocardiogram. *Cardiology*, *150*(4), 347–356. https://doi.org/10.1159/000542399

[16] Hilgendorf, L., Petursson, P., Andersson, E., Rawshani, A., Bhatt, D. L., Råmunddal, T., . . . , & Rawshani, A. (2025). Fully automated diagnosis of acute myocardial infarction using electrocardiograms and multimodal deep learning. *JACC: Advances*, *4*(8), 102011. https://doi.org/10.1016/j.jacadv.2025.102011

[17] Wagner, P., Strodthoff, N., Bousseljot, R. D., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, *7*(1), 154. https://doi.org/10.1038/s41597-020-0495-6

[18] Wagner, P., Strodthoff, N., Bousseljot, R. D., Samek, W., & Schaeffter, T. (2022). PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). *PhysioNet*. https://doi.org/10.13026/kfzx-aw45

[19] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., . . . , & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, *101*(23), e215–e220. https://doi.org/10.1161/01.cir.101.23.e215

[20] Frank, E., Hall, M. A., & Witten, H. I. (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques." The University of Waikato*. https://hdl.handle.net/10289/17096

[21] Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., ..., & Chen, S. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

[22] Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

[23] Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, *59*(1), 161–205. https://doi.org/10.1007/s10994-005-0466-3

[24] Strodthoff, N., Wagner, P., Schaeffter, T., & Samek, W. (2021). Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE Journal of Biomedical and Health Informatics*, *25*(5), 1519–1528. https://doi.org/10.1109/JBHI.2020.3022989

[25] Zeng, W., Shan, L., Chen, Y., Ma, Y., & Du, S. (2026). Interpretable deep learning framework for multi-label ECG classification: Enhancing cardiac diagnostics through feature attention and explainability. *Biomedical Signal Processing and Control*, *112*, 108447. https://doi.org/10.1016/j.bspc.2025.108447

[26] Tanyel, T., Atmaca, S., Gökçe, K., Balık, M. Y., Güler, A., Aslanger, E., & Öksüz, İ. (2025). Interpretable ECG analysis for myocardial infarction detection through counterfactuals. *Biomedical Signal Processing and Control*, *102*, 107227. https://doi.org/10.1016/j.bspc.2024.107227

[27] Jiao, D., Wang, Z., Li, J., Feng, F., & Hou, F. (2020). The chaotic characteristics detection based on multifractal detrended fluctuation analysis of the elderly 12-lead ECG signals. *Physica A: Statistical Mechanics and its Applications*, *540*, 123234. https://doi.org/10.1016/j.physa.2019.123234

How to Cite: Chen, C., Kouznetsova, V. L., & Tsigelny, I. F. (2026). Analysis of Interpretable Machine Learning Classifiers for Myocardial Infarction Detection Using 12-Lead ECG Features. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN62029651