COMMUNICATIONS IN
SCIENCE AND TECHNOLOGY
Homepage: cst.kipmi.or.id

Interpretability evaluation of rule-based classifier in myocardial infarction classification based on syntactical features of ECG signal

Farhatul Fityah[a], Noor Akhmad Setiawan[b,*], Dyah Wulan Anggrahini[c]

[a]*Master Program of Biomedical Engineering, Universitas Gadjah Mada, Yogyakarta 55284, Indonesia*
[b]*Department of Electrical and Information Engineering, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia*
[c]*Department of Cardiology and Vascular Medicine, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia*

Article history:
Received: 30 November 2025 / Received in revised form: 20 December 2025 / Accepted: 21 December 2025

Abstract

Cardiovascular diseases remain the leading cause of mortality on a global scale, with myocardial infarction (MI) representing a critical and life-threatening condition. Electrocardiography (ECG) is a widely utilized method for the detection of myocardial infarction (MI), and artificial intelligence (AI) has demonstrated a promising performance in the automated ECG-based diagnosis. However, most existing studies emphasize predictive accuracy while failing to provide substantial evidence that model decision logic aligns with clinical reasoning, thereby limiting clinical adoption. This present study evaluates the interpretability of three rule-based machine learning classifiers—Decision Tree, RIPPER, and Rough Set—for MI detection from ECG signals, including a comparison between models with and without feature selection. Interpretability of the system is assessed through rule complexity analysis and a standardized qualitative clinical validation protocol involving three cardiologists, based on contemporary AHA/ESC ECG diagnostic guidelines. The findings indicate that the Rough Set classifier attains the optimal overall performance, with 80% of its generated rules demonstrating clinically aligned, thereby outperforming the other models regarding interpretability. The findings demonstrate the benefit of guideline-based clinical validation for advancing trustworthy ECG-based MI diagnostic systems.

Keywords: Electrocardiography; intrinsic interpretability; myocardial infarction; rule-based classifier

1. Introduction

Cardiovascular disease, as released by the World Heart Federation in 2023, remains one of the leading causes of mortality on a global scale [1]. In Indonesia, these diseases are the leading cause of annual deaths, with a total of 651,481 fatalities each year. While, coronary heart disease is the second most prevalent cause of death after stroke [2]. Among its most critical manifestations is myocardial infarction (MI), which results from partial or complete occlusion of the coronary arteries due to atherosclerotic plaque formation, leading to myocardial ischemia and necrosis [3]. The early and accurate detection of MI is essential to prevent severe complications, including sudden cardiac death, and to enable timely clinical intervention [4].

Electrocardiography (ECG) is a non-invasive, cost-effective, and widely available diagnostic modality routinely used for the detection of MI through the analysis of cardiac electrical activity [5]. In clinical practice, ECG interpretation relies on cardiologists' expertise in identifying characteristic waveform abnormalities, such as pathological Q waves, ST-segment deviations, and T-wave changes [6].

* Corresponding author. Tel.: +62-816-4265-148
Email: noorwewe@ugm.ac.id
https://doi.org/10.21924/cst.10.2.2025.1851

In recent years, advances in artificial intelligence (AI), particularly in the realm of machine learning and deep learning, have demonstrated substantial potential for automating the detection and classification of MI from ECG signals, Numerous studies have reported high predictive performance, indicating the promising capabilities of AI-based systems in this field.

Deep learning-based approaches have achieved particularly notable results in automated disease detection tasks. For instance, convolutional neural network–based models have reported very high classification accuracy in several diagnostic applications [7], while hybrid architectures with the combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have also demonstrated robust predictive performance [8]. Conventional machine learning methods, including Support Vector Machines (SVM) and Multilayer Perceptron (MLP), have similarly been extensively applied, frequently achieving competitive accuracy [9]. Parallel to this, intelligent systems based on fuzzy logic have been developed for the diagnosis of cardiovascular disease, including coronary artery disease. These system encode expert knowledge to support decision-making under uncertainty [10]. Despite these promising results across different diagnostic settings, most existing studies predominantly emphasize predictive performance, while issues related to interpretability, transparency, and clinical trustworthiness remain relatively underexplored.

In healthcare applications, interpretability is a critical requirement for ensuring clinical trust, ethical accountability, and practical usability of AI-based decision-support systems [11,12]. It refers to the degree to which human users—particularly clinicians—can understand and rationalize a model's predictions in terms of established medical knowledge and diagnostic reasoning. High predictive accuracy alone is inadequate for clinical adoption as long as model decisions are not adequately explained, scrutinized, or justified within a medical context.

Rule-based classifiers, characterized by explicit if–then decision structures, are widely regarded as intrinsically interpretable in view of their transparent reasoning mechanisms [13,14]. Consequently, such models are frequently considered suitable for clinical applications where explainability is essential. Interpretability evaluation methods are generally categorized into two distinct approaches: quantitative (machine-level) assessments, which include metrics such as rule count, rule length, or statistical support and qualitative (human-level) validation. While most of prior research on interpretable machine learning models has largely focused on quantitative indicators, qualitative validation involving domain experts, particularly cardiologists, remains limited [15 – 17]. In the context of medical practice, such qualitative validation is crucial to ensure that model-generated rules are not only syntactically interpretable but also clinically meaningful and aligned with established diagnostic standards.

Based on this gap, this present study evaluates the interpretability of three rule-based classifiers—Decision Tree, RIPPER, and Rough Set—for ECG-based myocardial infarction detection. It emphasizes formal qualitative clinical validation by multiple cardiologists rather than solely rule structure or quantitative metrics. The present work is grounded in expert judgment and contemporary ECG guidelines, and offers a clinically oriented perspective on interpretability assessment.

2. Materials and Methods

This present study applied systematic workflow for ECG-based myocardial infarction classification using the PTB-XL dataset. The workflow included signal preprocessing, feature extraction and discretization, performance evaluation, and clinical validation to evaluate rule interpretability and clinical relevance.

2.1. Dataset

This present study utilized the PTB-XL ECG dataset [17] from PhysioNet, which consists of 12-lead ECG recordings, each with a duration of 10 seconds, validated according to the SCP-ECG standard. The dataset comprises a total of 5,310 records, including 1,388 Anterior Myocardial Infarction (AMI), 1,267 Inferior Myocardial Infarction (IMI), and 2,655 normal recordings.

2.2. Signal filtering

To enhance the clarity and reliability of ECG signal, preprocessing was applied purposely to reduce noise from powerline interference, electrode artifacts, and baseline wander, all of which have the potential to obscure diagnostic information. In this study, an Adaptive Mean Filter was applied to effectively remove baseline wander [18]. Subsequently, Discrete Wavelet Transform (DWT) was implemented by means of the Daubechies6 (db6) mother wavelet at level 8 as the final filtering step [19].

2.3. Feature extraction

In this study, ECG feature extraction targets MI-related abnormalities, including pathological Q waves, ST-segment elevation, T wave inversion, and elevated T wave. Accordingly, Q wave, T wave, and J-point amplitude were extracted as key features due to their direct correlation with ST-segment changes and MI detection.

The feature extraction process involves the segmentation of each ECG lead into individual P–QRS–T complexes through the detection of R peaks and definition of RR intervals. The analysis of heartbeat is conducted by delineating each heartbeat from the midpoint of one RR interval to the next. The identification of R peaks is facilitated by the Two-Average algorithm due to its reliable detection accuracy [20]. After segmentation, features extraction is achieved through the implementation of a windowing technique [21], incorporating the following point detection adjustment algorithm:

A. R-Peak Detection: The Two-Average algorithm is employed to detect the R-peak. Once identified, a window of 2% of the sampling frequency is applied to examine any higher amplitude within the window. This adjustment has been demonstrated to enhance the accuracy of R-peak detection, particularly in cases where the Two-Average algorithm detects only the QRS complex.
B. Q-Peak Detection: The Q-peak is identified as the first negative deflection or zero crossing within the window preceding the R-peak.
C. S-Peak Detection: The S-peak is determined as the first negative deflection or zero crossing following the R-peak.
D. J-Point Detection: The J-point is detected as the subsequent deflection following the S-peak, marking the transition between the QRS complex and the ST segment.
E. T-Peak Detection: The T-peak is identified as the absolute maximum amplitude (positive or negative) within the window extending from the J-point to the end of the segment.
F. T-Onset Detection: Similar to Q-peak detection, the T-onset is determined by locating the first negative deflection or zero crossing before the T-peak, referenced relative to the R-peak.

This feature extraction approach enables accurate identification of MI-related ECG key points and provides relevant features as inputs for effective classification of myocardial infarction abnormalities.

2.4. Data discretization

Data discretization is implemented through the categorization of feature values into predefined ranges in accordance with clinical guidelines and medical relevance [22 – 24]. In this study, each feature was discretized according to the threshold values as specified in Table 1.

Table 1. Discretization category

Feature	Lead	Threshold Condition	Category
Q Peak Amplitude	All	≤ -0,2 mV	Pathologic
		Else	Normal
S Peak Amplitude	V1 – V6	> 0,2 mV	High
	I, II, III, aVL, aVR, aVF	> 0,1 mV	
	V1 – V6	< -0,2 mV	Low
	I, II, III, aVL, aVR, aVF	< -0,1 mV	
	V1 – V6	-0,2 < S peak < 0,2 mV	Normal
	I, II, III, aVL, aVR, aVF	-0,1< S peak < 0,1 mV	
T Onset Amplitude	All	> 0,2 mV	High
	All	< -0,2 mV	Low
	All	-0,2 < S < 0,2 mV	Normal
	Except aVR, V1	≤ -0,1 mV	Inverted
T Peak Amplitude	All	-0,05 < T < 0,05	Undetected
	All	0,1 ≤ T ≤ 1,0 mV	Normal
	All	-1,0 > T > 1,0 mV	Hyperacute

2.5. Classification

This present study compared three intrinsically interpretable classifiers: Decision Tree, RIPPER, and Rough Set. The Decision Tree classifier constructs rules based on a hierarchical branching structure, generated through recursive partitioning of data attributes [25]. Each path from the root to a leaf node represents a distinct rule, encapsulating specific patterns within the data [26]. RIPPER (Repeated Incremental Pruning to Produce Error Reduction) is a rule-based algorithm that iteratively refines its rules to enhance classification accuracy. Initially, rules are generated from the training data, followed by a pruning process to mitigate overfitting [27]. Rough Set Theory (RST), introduced by Pawlak in 1982, provides a mathematical framework for managing uncertainty and incomplete information by defining upper and lower approximations of data classes [28]. These approximations serve as the basis for rule generation within the classifier.

The Decision Tree and RIPPER classifiers were implemented in Python using Visual Studio Code, while the Rough Set classifier was executed in RSES2. The evaluation of all models was conducted with and without feature selection, employing stratified 10-fold cross-validation to ensure balanced class representation.

2.6. Evaluation

The model evaluation process involved a comparison of classification performance with and without feature selection. This was conducted by means of confusion matrix–based metrics (i.e. accuracy, precision, recall, and F1-score) and the Area Under the ROC Curve (AUC) to assess overall ranking performance [29].

For experiments involving feature selection, different strategies were applied with the selection of strategy dependent upon the classifier. The Recursive Feature Elimination (RFE) [30] was employed for the Decision Tree and RIPPER classifiers, where features were iteratively ranked and removed based on importance scores derived from estimator, purposely to retain the most informative subset. In contrast, the feature selection process for the Rough Set classifier was performed through the utilization of the reduct mechanism implemented in the Rough Set Exploratory System 2 (RSES2). This mechanism is designed to identify minimal attribute subsets that preserve decision class discernibility in accordance with the principles of rough set theory.

Subsequently, interpretability evaluation was performed based on two complementary criteria: model complexity and clinical qualitative validation. The complexity of the model was quantified by the analysis of the total number of rules generated by each classifier and the average number of features involved per rule.

The objective of this present study was to undertake clinical qualitative validation to assess the extent to which model-generated decision rules align with established clinical diagnostic reasoning for myocardial infarction. The validation protocol was grounded in the Fourth Universal Definition of Myocardial Infarction, as endorsed by the European Society of Cardiology (ESC) [31], American College of Cardiology (ACC), and American Heart Association (AHA) [32]. In accordance with these guidelines, electrocardiographic evidence of MI is characterized by pathological Q-waves, ST-segment elevation at the J-point, or T-wave abnormalities. These abnormalities must occur in a minimum of two anatomically contiguous ECG leads corresponding to the same myocardial territory (e.g., anterior or inferior regions).

Three independent cardiologists served as clinical validators. For each classifier, five representative rules per class (NORMAL, AMI, and IMI) were selected based on the highest support values and compiled into structured validation sheets for independent review.

A rule was considered clinically aligned if it satisfied the following operational criteria: (1) involvement of ECG features corresponding to recognized ischemic abnormalities (i.e. pathological Q-wave, ST-segment deviation at the J-point, or T-wave inversion/hyperacute T-wave), and (2) manifestation of these abnormalities across two or more contiguous leads consistent with guideline-based myocardial territory definitions.

The aggregation of clinical judgments was achieved through a consensus-based approach, whereby a rule was deemed aligned if it received concurrence from at least two of the three cardiologists. This aimed to minimize subjectivity and enhance validation robustness.

3. Results and Discussion

3.1. Classification result

The classification was performed hierarchically, encompassing binary NORMAL–MI classification followed by AMI–IMI subclass classification.

3.1.1. NORMAL – MI classification

Table 2 depicts the summary of the classification performance for the NORMAL–MI task. The Rough Set classifier demonstrated highest overall performance across most metrics, including accuracy, F1-score, and AUC, both with and without feature selection. This is likely attributable to its capacity o derive decision rules from multiple feature combinations while managing data uncertainty.

Based on a paired t-test, the performance of models with and without feature selection was found only marginally different and not statistically significant. This indicates that discriminative information was relatively evenly distributed across ECG features. Clinically, this suggests that global ischemic patterns that distinguish normal and MI ECGs can be reliably captured without reliance on a narrowly selected feature subset.

Table 2. Class classification (Normal – MI) result

Evaluation Metrics	Decision Tree		RIPPER		Rough Set	
	Feature Selection					
	x	√	x	√	x	√
Accuracy	0.8064	0.8122	0.7851	0.7836	0.8363	0.8312
Precision	0.7835	0.7995	0.8242	0.8154	0.8377	0.8348
Recall	0.8471	0.8342	0.7262	0.7382	0.8363	0.8312
F1-Score	0.8138	0.8162	0.7712	0.7718	0.8366	0.8315
AUC	0.8707	0.8757	0.7949	0.7969	0.832	0.8335

x = without feature selection; √ = with feature selection

3.1.2. AMI – IMI subclass classification

Table 3 presents the performance of the AMI–IMI subclass classification. Consistent with the binary task, the Rough Set classifier attained the maximum accuracy and F1-score. Contrasting to the NORMAL–MI task, feature selection demonstrated a statistically significant effect on the Rough Set model, while no significant impact was observed for the Decision Tree and RIPPER classifiers.

This effect can be attributed to the Rough Set rule induction mechanism. Without feature selection, LEM2 induces compact, high-precision rules with limited coverage, leaving some instances unclassified and inflating apparent accuracy. The application of feature selection via Genetic Algorithm–based reduction produces a larger rule set that enhances coverage at the cost of slightly reduced accuracy. This behavior reflects a core principle of Rough Set theory, in which uncertain instances are assigned to the boundary region rather than forcibly classified, thus favoring error avoidance over exhaustive prediction—an important consideration in clinical decision-making.

Table 3. Subclass Classification (AMI – IMI) Result

Evaluation Metric	Decision Tree		RIPPER		Rough Set	
	Feature Selection					
	x	√	x	√	x	√
Accuracy	0.8210	0.8222	0.8007	0.8000	0.8535	0.8259
Precision	0.8383	0.8395	0.8582	0.8539	0.8566	0.8269
Recall	0.7765	0.7781	0.7001	0.7024	0.8535	0.8259
F1-Score	0.8049	0.8062	0.7690	0.7690	0.8533	0.8261
AUC	0.8812	0.8821	0.8057	0.8087	0.854	0.8265

x = without feature selection; √ = with feature selection

3.2. Interpretability evaluation

Interpretability was assessed using two complementary dimensions: quantitative rule complexity and qualitative clinical validation based on established myocardial infarction diagnostic criteria.

3.2.1. Rule complexity

Fig. 1 illustrates the number of rules generated by each classifier, while Fig. 2 presents the average number of features per rule.

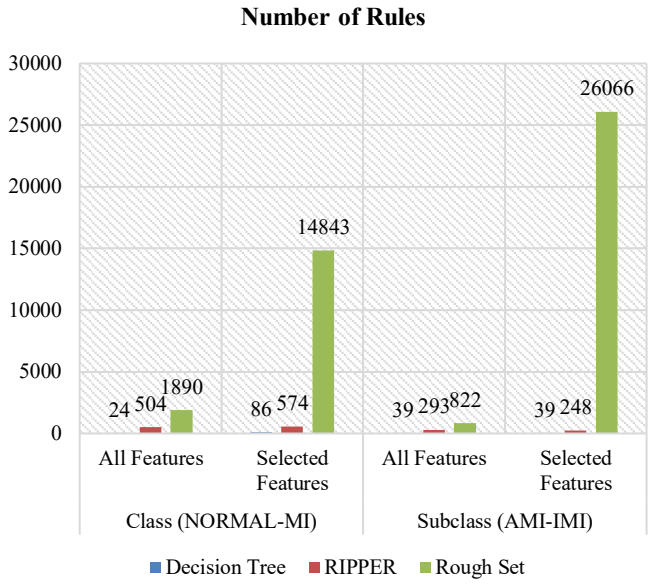

Fig. 1. Number of rules extracted

The Rough Set classifier produced the most complex models, with a higher number of rules and features per rule due to its exhaustive rule-generation strategy, which captures fine-grained ECG feature relationships but increases the interpretative burden.

RIPPER demonstrated moderate complexity, which was consistent with its incremental learning strategy. Feature selection increased rule counts in the NORMAL–MI task to compensate for reduced feature diversity, but simplified rules in the AMI–IMI task. This indicated sufficient retained features for infarction localization.

The Decision Tree generated the fewest rules owing to its hierarchical structure and pruning. The implementation of feature selection increased rule complexity in the NORMAL–MI task but it had minimal impact in the AMI–IMI task, suggesting preserved discriminative power for subclass differentiation.

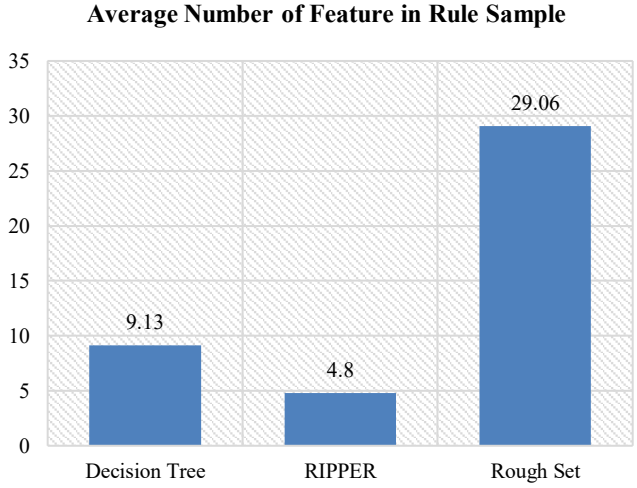

Fig. 2. Average Number of Feature in Rule Sample

Overall, these results obtained highlight fundamental structural differences among the classifiers and demonstrate the manner of feature selection interacts with each model's rule-generation process.

3.2.2. Clinical qualitative validation

Clinical qualitative validation in this present study was conducted to assess the alignment between machine-generated rules and established ECG-based diagnostic criteria for myocardial infarction (MI). In adherence to contemporary clinical practice and international guidelines, the diagnosis of MI requires the presence of ischemic ECG abnormalities such as ST-segment elevation at the J-point, pathological Q waves, or T-wave inversion, in at least two anatomically contiguous leads. Accordingly, the validation process focused on clinically relevant lead groupings, including inferior (I–aVL, II–III–aVF) and anterior (V1–V2, V3–V4, V5–V6) regions.

To ensure a focused yet clinically meaningful assessment, five representative rules per class (NORMAL, AMI, and IMI) were selected based upon the highest support values, reflecting rules that were most frequently activated during classification. These rules were then evaluated by cardiologists through a consensus-based adjudication process, where a rule was considered clinically aligned if its feature conditions reflected recognized ECG abnormalities across contiguous lead groups consistent with guideline-based MI diagnosis. The comparison of clinical validation results is presented in Fig. 3.

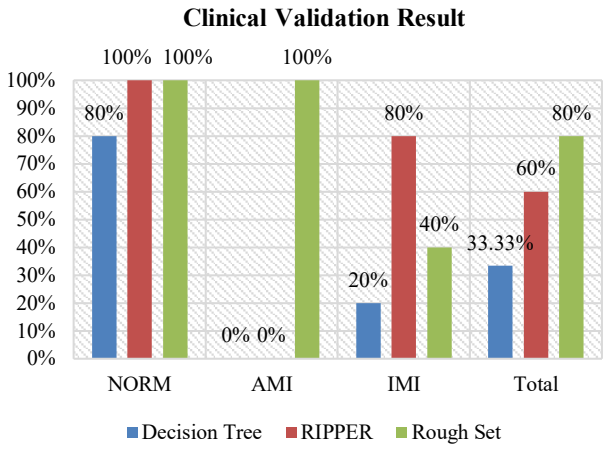

Fig. 3. Clinical validation result

3.2.2.1. Rough set model

The Rough Set model demonstrated the highest clinical agreement, with 80% of evaluated rules considered to be clinically aligned. Valid rules commonly captured pathological Q-wave or abnormal T-wave patterns across contiguous inferior or anterior leads, whereas mismatches occurred primarily in a small number of IMI rules, characterized by the absence of paired inferior-lead involvement.

This behavior is indicative of the underlying rule-generation mechanisms. LEM2 produced fewer, high-precision rules with limited coverage, while GA-based feature selection generated a larger rule set that improved coverage and consistency. Despite an increase in the volume of rule, interpretability was maintained through explicit if–then rules, thereby enabling clinicians to trace decision logic and capture multi-lead dependencies that are central to the interpretation of MI.

3.2.2.2. RIPPER model

The RIPPER classifier achieved moderate clinical validation accuracy (60%), performing well for NORMAL cases but exhibiting limitations in the discrimination of AMI and IMI. Misaligned rules frequently relied on abnormalities in a single inferior lead (e.g., pathological Q-waves in lead II) without corroboration from contiguous leads, and some AMI rules incorrectly emphasized inferior-lead patterns characteristic of IMI.

This behavior is indicative of RIPPER's incremental, parsimonious rule induction strategy, which yields concise and easily interpretable rules but does not enforce clinical constraints on lead contiguity. Consequently, although RIPPER rules were considered to be the easiest to understand, their consistency with formal diagnostic criteria remained limited.

3.2.2.3. Decision tree model

The Decision Tree classifier demonstrated the lowest clinical validation accuracy of 33.33%. Despite its satisfactory overall classification performance, its rules frequently failed to capture clinically meaningful MI patterns. This occurred when abnormalities were observed in a single lead with normal findings in others without enforcing contiguous lead involvement, particularly in cases of AMI and IMI.

This limitation reflects the hierarchical, locally optimized split strategy of Decision Trees and the effects of pruning, which simplify models but may remove subtle inter-lead relationships that are critical for the interpretation of MI. Consequently, despite reasonable statistical accuracy, clinical alignment was limited.

Overall, these findings indicate that clinical interpretability is dependent not only on rule simplicity but on the capacity to encode multi-lead ECG patterns consistent with diagnostic guidelines, a property more effectively captured by the Rough Set classifier.

3.3. Comparative analysis

Several studies have explored the use of interpretable and rule-based machine learning models for ECG-based cardiovascular disease analysis, primarily focusing on classification performance rather than clinical interpretability. These studies are relevant to the present work in view of their use of similar classification models or comparable ECG datasets, allowing methodological comparison.

A number of comparative studies have evaluated traditional machine learning classifiers, including Decision Tree, Naïve Bayes, Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest, and Convolutional Neural Networks, for ECG-based heart disease classification tasks. These studies typically employed publicly available ECG datasets with the purpose of evaluating model performance using accuracy-based metrics. Within this setting, Decision Tree classifiers were reported to achieve moderate classification accuracy of approximately 62.86% [33]. Despite the fact that Decision Trees provide explicit and transparent decision structures, these studies did not include clinical validation of the extracted rules, and interpretability was implicitly assumed based on model form rather than assessed in relation to myocardial infarction diagnostic reasoning.

Other studies have focused on arrhythmia classification using the MIT-BIH Arrhythmia dataset, a benchmark dataset widely utilized for the evaluation of ECG classification algorithms. In these works, rule-based classifiers such as Decision Tree C4.5, PART, and RIPPER were compared, achieving reported accuracies of up to 92.48% [34]. While these studies demonstrate the effectiveness of rule-based models on standardized ECG data, interpretability assessment remained limited to the presentation of induced rules. No formal qualitative evaluation by clinical experts has been conducted to determine whether the learned rules reflected clinically meaningful ECG patterns or diagnostic logic.

Further investigations employing the UCI Arrhythmia dataset integrated statistical feature selection techniques with rule-based classifiers. In this context, of several competing models, RIPPER was reported to achieve the highest classification accuracy [35]. These studies highlight the impact of feature selection on rule induction and classification performance. However, the interpretability evaluation remained confined to rule visualization and quantitative metrics, without examining whether the selected features and resulting rules aligned with real-world clinical reasoning or diagnostic standards.

Rough Set Theory has been applied more directly to myocardial infarction classification using ECG-derived features such as ST-segment elevation, pathological Q waves, and T-wave abnormalities. The classification accuracies reported in these studies range up to 99.8% [36], demonstrating the potential of Rough Set–based rule induction for capturing diagnostically relevant ECG characteristics. Nevertheless, the evaluation in these works primarily focused on rule strength, coverage, and predictive accuracy. The generated rules were not subjected to explicit qualitative validation against contemporary ECG interpretation guidelines or assessed by cardiologists. Similar evaluation strategies have been observed in other Rough Set–based approaches for heart disease diagnosis, where interpretability is inferred from quantitative measures with minimal involvement of medical experts [37].

These studies, taken together, demonstrate that rule-based and interpretable machine learning models can achieve competitive performance in a range of ECG-based classification tasks and datasets. However, across different problem formulations—ranging from general heart disease classification to arrhythmia detection and myocardial infarction diagnosis—interpretability is consistently treated as an implicit model property. The absence of formal qualitative clinical validation limits the extent to which the reported rules can be considered clinically interpretable or aligned with real-world diagnostic reasoning.

In contrast, the present study addresses this limitation by explicitly incorporating cardiologist-driven qualitative validation of model-generated rules in the context of myocardial infarction classification. Rather than inferring interpretability from model structure or quantitative indicators alone, this study evaluates whether the induced decision rules reflect established ECG interpretation principles. By grounding interpretability assessment in expert clinical judgment, this present study proposes a more rigorous and clinically meaningful framework for evaluating rule-based machine learning models for ECG-based myocardial infarction detection.

4. Conclusion

This present study evaluated the interpretability of three intrinsically interpretable rule-based classifiers—Decision Tree, RIPPER, and Rough Set—for ECG-based myocardial infarction detection, with primary emphasis on clinical alignment rather than predictive performance alone. Despite comparable classification accuracy across models, qualitative validation by cardiologists revealed substantial differences in interpretability and clinical relevance. The Rough Set classifier demonstrated the highest clinical alignment, with numerous rules conforming to guideline-based ECG patterns, particularly those involving anatomically contiguous lead involvement. Despite generating a larger rule set, its explicit if–then structure facilitated transparent clinical reasoning. RIPPER produced fewer, more concise rules that were easier to read, though less consistently aligned with diagnostic criteria, reflecting differences in rule structure rather than interpretability per se. In contrast, the Decision Tree exhibited minimal responsiveness to lead-level diagnostic patterns, thereby constraining its clinical applicability despite consistent performance. These findings highlight the potential of rule-based models—particularly Rough Set approaches—for interpretable ECG-based MI detection, while also acknowledging limitations related to amplitude-based features, the emphasis on Q-wave MI patterns, and the scalability of extensive rule sets. Future research should incorporate temporal ECG features, expand evaluation to broader MI subtypes, and address rule redundancy. Overall, this present study demonstrates that qualitative clinical validation is essential for assessing the real-world applicability of interpretable machine learning in cardiology beyond conventional performance metrics.

References

1. World Heart Federation, *World Heart Report 2023: Confronting the World's Number One Killer*, World Heart Federation, 2023, pp. 1–52.
2. N. Azizah, *Kemenkes: Penyakit kardiovaskular jadi penyebab kematian terbanyak di Indonesia*, Republika News, 2023.
3. N. Ojha, A.S. Dhamoon, *Myocardial infarction*, StatPearls Publishing, 2023.
4. N. Strodthoff, C. Strodthoff, *Detecting and interpreting myocardial infarction using fully convolutional neural networks*, Physiol. Meas. 40 (2019) 1–11.
5. S. Śmigiel, *ECG classification using orthogonal matching pursuit and machine learning*, Sensors 22 (2022).
6. S.M. Salerno, P.C. Alguire, H.S. Waxman, *Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence*, Ann. Intern. Med. (2003) 43–45.
7. P. Lu et al., *KecNet: a light neural network for arrhythmia classification based on knowledge reinforcement*, J. Healthc. Eng. (2021).
8. S. Hasbullah, M.S. Mohd Zahid, S. Mandala, *Detection of myocardial infarction using hybrid models of convolutional neural network and recurrent neural network*, BioMedInformatics 3 (2023) 478–492.
9. S. Aziz, S. Ahmed, M.S. Alouini, *ECG-based machine-learning algorithms for heartbeat classification*, Sci. Rep. 11 (2021) 1–14.
10. A.M. Sajiah, N.A. Setiawan, O. Wahyunggoro, *Interval type-2 fuzzy logic system for diagnosis coronary artery disease*, J. Telecommun. Electron. Comput. Eng. 1 (2016) 51–55.
11. J. Amann, A. Blasimme, E. Vayena, D. Frey, V.I. Madai, *Explainability*

for artificial intelligence in healthcare: a multidisciplinary perspective, BMC Med. Inform. Decis. Mak. 20 (2020) 1–9.
12. A.J. London, *Artificial intelligence and black-box medical decisions: accuracy versus explainability*, Hastings Cent. Rep. 49 (2019) 15–21.
13. A. Vellido, *The importance of interpretability and visualization in machine learning for applications in medicine and health care*, Neural Comput. Appl. 32 (2020) 18069–18083.
14. C. Molnar, *Interpretable machine learning: a guide for making black box models explainable*, Creative Commons License, 2021.
15. H. Turbé, M. Bjelogrlic, C. Lovis, G. Mengaldo, *Evaluation of post-hoc interpretability methods in time-series classification*, Nat. Mach. Intell. 5 (2023) 250–260.
16. N. Alangari, M.E.B. Menai, H. Mathkour, I. Almosallam, *Exploring evaluation methods for interpretable machine learning: a survey*, Information 14 (2023) 1–29.
17. P. Wagner et al., *PTB-XL, a large publicly available electrocardiography dataset*, Sci. Data 7 (2020) 1–15.
18. W. Jenkal et al., *An efficient algorithm of ECG signal denoising using the adaptive dual threshold filter and the discrete wavelet transform*, Biocybern. Biomed. Eng. 36 (2016) 499–508.
19. M. AlMahamdy, H.B. Riley, *Performance study of different denoising methods for ECG signals*, Procedia Comput. Sci. 37 (2014) 325–332.
20. S. Śmigiel, K. Pałczyński, D. Ledziński, *Deep learning techniques in the classification of ECG signals using R-peak detection based on the PTB-XL dataset*, Sensors 21 (2021) 1–18.
21. L. De Palma et al., *ECG wave segmentation algorithm for complete P-QRS-T detection*, Proc. IEEE Int. Symp. Med. Meas. Appl. (2023) 1–6.
22. Y. Sattar, L. Chhabra, *Electrocardiogram*, StatPearls Publishing, 2023.
23. C. Hodnesdal et al., *Rapidly upsloping ST-segment on exercise ECG: a marker of reduced coronary heart disease mortality risk*, Eur. J. Prev. Cardiol. 20 (2013) 541–548.
24. W.J. Brady, *ST segment and T wave abnormalities not caused by acute coronary syndromes*, Emerg. Med. Clin. North Am. 24 (2006) 91–111.
25. J. Han, M. Kamber, J. Pei, *Data mining: concepts and techniques*, Elsevier, 2011.
26. S. Fletcher, M.Z. Islam, *Decision tree classification with differential privacy: a survey*, ACM Comput. Surv. 52 (2019) 1–35.
27. W.W. Cohen, *Fast effective rule induction*, Morgan Kaufmann, 1995.
28. Z. Pawlak, *Rough sets: theoretical aspects of reasoning about data*, Kluwer Academic Publishers, 1991.
29. A. Rakotomamonjy, *Optimizing area under ROC curve with SVMs*, Proc. ICML (2004) 71–80.
30. X. Chen, J.C. Jeong, *Enhanced recursive feature elimination*, Proc. ICMLA (2007) 429–435.
31. A. Task et al., *2023 ESC guidelines for the management of acute coronary syndromes*, Eur. Heart J. (2023) 3720–3826.
32. M.H. Crawford et al., *ACC/AHA guidelines for ambulatory electrocardiography*, J. Am. Coll. Cardiol. 34 (1999).
33. M. Ciocirlan, A. Udrea, *Techniques of biological signals classification and comparisons using machine learning techniques*, Proc. CSCS (2023) 467–472.
34. N.A. Setiawan et al., *Classification of arrhythmia ECG signal using cascade transparent classifier*, J. Intell. Fuzzy Syst. 42 (2022) 1015–1025.
35. S. Singh, *Comparative study ID3, CART and C4.5 decision tree algorithm: a survey*, Int. J. Adv. Inf. Sci. Technol. 27 (2014) 97–103.
36. B. Halder, S. Mitra, M. Mitra, *Classification of complete myocardial infarction using rule-based rough set method and rough set explorer system*, IETE J. Res. 68 (2022) 85–95.
37. K.P. Ahmed, D.P. Acharjya, *A hybrid scheme for heart disease diagnosis using rough set and cuckoo search technique*, J. Med. Syst. 44 (2020).