**Bulletin of Electrical Engineering and Informatics**
Vol. 15, No. 3, June 2026, pp. 2690~2700
ISSN: 2302-9285, DOI: 10.11591/eei.v15i3.10317 

# Analysis of arrhythmia detection and classification using electrocardiogram signals with decision tree algorithm

**Putri Juniarti Marpaung[1], Nia Anggreni Matondang[1], Rince Margaretta Siregar[1], Angelica Ester Novelia Siahaan[1], Abdi Dharma[1], Arjon Turnip[2], Mardi Turnip[1]**
[1]Information Systems Study Program, Faculty of Science and Technology, Universitas Prima Indonesia, Medan, Indonesia
[2]Department of Electrical Engineering, Universitas Padjadjaran, Bandung, Indonesia

**Article Info**

***Article history:***

Received Mar 20, 2025
Revised Mar 11, 2026
Accepted Mar 31, 2026

***Keywords:***

Arrhythmia
Classification
Decision tree
Electrocardiograms
Real-time monitoring

**ABSTRACT**

Heart disease remains the primary cause of death globally, with arrhythmia diagnosis often limited by restricted access to medical personnel and the complexity of electrocardiogram (ECG) interpretation. Accurate arrhythmia classification is essential to prevent cardiovascular complications. The proposed method successfully categorized classify ECG signals into five categories: normal, abnormal, potentially arrhythmia, moderate arrhythmia risk, and highly potentially arrhythmia. Data were collected from 30 subjects under three activity scenarios: sitting, walking, and running. The proposed model achieved an accuracy of 99.4%, demonstrating strong potential for real-time monitoring applications. Performance evaluation was conducted using accuracy, precision, recall, and F1-score for each class. Although the dataset size remains relatively small, the findings highlight the effectiveness of decision tree as an efficient and interpretable classification method. Future research will involve validation using large-scale public databases like the arrhythmia database at MIT-BIH and comparisons with advanced methods including convolutional neural network (CNN), transformer-based models, and explainable artificial intelligent (XAI) frameworks.



***Corresponding Author:***

Mardi Turnip
Information Systems Study Program, Faculty of Science and Technology, Universitas Prima Indonesia
Sampul Road No. 4, Sei Putih Barat, Medan Petisah, Medan, North Sumatra 20118, Indonesia
Email: marditurnip@unprimdn.ac.id

## 1. INTRODUCTION

After stroke, heart disease continues to be the second most common cause of mortality in Indonesia and throughout the world [1]. The identification of cardiac disorders like arrhythmia is one area where technology's use in medicine is still developing. Arrhythmia is a cardiac rhythm disorder characterized by complicated waveforms that make it difficult to accurately recognize waves and irregular heartbeats of various types [2]. Arrhythmia, a primary kind of cardiovascular illness, is characterized by erratic electrical activity and often has substantial consequences [3]. Electrocardiograms (ECGs), which are very useful and practical medical instruments, have become more important in the diagnosis and management of arrhythmias due to technical advancements [4]. Early arrhythmia diagnosis outside of clinical settings is made easier by the deployment of wearable technology and telemonitoring systems that can collect ECG data in real time [5]. The accuracy of ECG interpretation has increased thanks to developments in artificial intelligence and machine learning, making it possible to quickly and precisely identify complicated arrhythmia patterns [6].

As technology advances, the computer science community is urged to contribute to the medical field. In order to support the medical community in clinical decision-making, the community creates and uses software that aids in the analysis of bioelectrical signals. An electrical impulse detector is used in an ECG, a

medical diagnostic that tracks and captures the heart's electrical activity. By transforming electrical impulses into graphs that are shown on a monitoring screen, this test helps lessen or avoid issues brought on by heart disease [7]. Especially in clinics, traditional ECGs capture the electrical activity of the heart from 12 distinct perspectives [8]. Wearable ECG, 3-lead ECG, artificial intelligence-based algorithms, wireless ECG, mobile ECG, and 3D ECG devices are only a few examples of the fast-evolving ECG technologies.

Advanced computational techniques for arrhythmia identification have been investigated in recent publications. Better patient privacy is ensured via federated learning using decision trees, which allows medical institutions to collaborate without exchanging raw data [9]. Decision trees may be used on low-power wearable devices for real-time arrhythmia detection without relying on the cloud thanks to on-device TinyML [10]. Graph neural networks (GNNs) are more accurate than traditional decision trees in capturing associations between cardiac segments by modeling ECG data as spatiotemporal graphs [11]. ECG waveform representations and decision trees are combined in time-series-based explainable artificial intelligence (XAI) modules, enabling cardiologists to verify [12]. Strong feature representation and clinical interpretability are balanced by hybrid techniques that combine transformer-based autoencoders with decision trees [13]. To increase sensitivity in identifying uncommon arrhythmias, experimental quantum-inspired feature engineering has also been investigated [14].

Current arrhythmia diagnosis methods heavily utilize machine learning and deep learning algorithms. Artificial neural networks (ANNs) are effective in recognizing complex patterns but are prone to overfitting [15]. Support vector machines (SVMs) perform well in binary classification but are less optimal for multi-class problems [16]. Convolutional neural networks (CNNs) achieve high accuracy by automatically extracting spatio-temporal features, although their black-box nature limits clinical interpretability [17]. Deep neural networks (DNNs) and transformer architectures demonstrate superior performance on large datasets but require high computational resources and offer limited transparency [18]. Compared to these approaches, decision trees provide interpretability and computational efficiency, although accuracy may be lower on large and heterogeneous datasets [19], [20]. This opens opportunities for more reliable arrhythmia diagnosis and applications across various dataset scales.

Recent developments in arrhythmia diagnosis focus on machine learning-based innovations that are still in the research stage. Approaches such as federated learning, TinyML on-device, GNN, and XAI emphasize data security and clinical interpretability. Previous studies focusing on ANNs, DNNs, and SVMs generally use the MIT-BIH arrhythmia database [21], prioritizing model performance and empirical validation but requiring complex data preprocessing [22]. These approaches complement each other in improving the effectiveness and efficiency of clinical arrhythmia diagnosis [23]. Studies also show that DNN produces the best validation accuracy, while SVM provides competitive classification results [24], [25]. Thus, current innovations focus on new ideas, while previous research relied on conventional methods.

Although deep learning has advanced arrhythmia detection, most studies focus mainly on improving classification accuracy while overlooking model interpretability and computational efficiency. Decision trees, despite being interpretable and efficient, are often used only as baseline models and rarely optimized for multi-class arrhythmia classification with structured validation and feature selection. Therefore, the limited exploration of decision trees as a primary method for arrhythmia detection represents the research gap addressed in this study. The novelty of this work lies in the systematic use of decision trees with k-fold cross-validation and optimized ECG feature preprocessing to achieve a balance between interpretability, efficiency, and competitive accuracy for wearable and clinical applications.

## 2. METHOD

This research was conducted at Universitas Prima Indonesia in collaboration with Universitas Padjadjaran in Bandung. The equipment and supplies utilized to collect the data were electrodes, data cables, AD8232 sensor boards, NodeMCU ESP32, SD cards, Raspberry Pi, Bluetooth, and Wi-Fi. The electrodes were connected to the subjects' chests, with the right chest colored red, the left chest colored yellow, and the lower right chest colored green, parallel to the red. We involved a number of vocational school students as subjects in this study, namely 30 male subjects aged between 15 and 17 years. Data was collected on each subject for 9 minutes. Three categories were created: sitting (±3) minutes, walking (±3) minutes, and running (±3) minutes. It is crucial to keep track of and document patients' physiological status in order to guarantee their quality of life, the appropriateness of their surroundings, and current health information [26]. Together, these elements guarantee that the resulting ECG data may be effectively gathered, processed, and examined. This set of resources is intended to assist the goals of the research by ensuring that data collection and analysis are optimal, effective, and organized.

The ECG components are installed on the ECG machine before it is turned on. The chest, ankles, and hands of the research subjects are cleaned with alcohol swabs before the study is conducted. Three ECG electrodes are then coated with gel and attached to the research subjects. The upper right side of the body is

connected to the red electrode, and the lower left side to the green electrode. During the activity, the ECG records and logs their heart activity. The electrical activity of the heart is captured via the AD8232 sensor and the NodeMCU ESP32 sensor. Because ECG signals might contain a lot of noise, the AD8232 sensor is helpful for increasing the accuracy of the PR and QT interval values [27]. During the activity performed by the subjects, they are instructed to breathe properly, and the ECG machine records the data.

The decision tree approach is used in this work to categorize ECG signal data by constructing a tree-based prediction model. The dataset consists of five wave intervals and three main ECG waves: P wave, QRS complex, and T wave. The P wave represents atrial depolarization with a normal duration of less than 0.12 seconds. The Q wave, characterized by small amplitude and short duration (less than 0.04 seconds), helps distinguish normal and abnormal ECG signals. The R wave, with varying amplitude depending on electrode placement, is an important indicator in heart rhythm analysis.

The decision tree method classifies data based on specific conditions until each branch reaches a single class, insufficient data, or no further feature separation is possible. Heart rate signals are recorded using chest electrodes while subjects sit, walk, and run for 3 minutes each. The data is transmitted via Bluetooth to a Raspberry Pi for processing and ECG signal analysis. The processed data is then classified into five categories: normal, abnormal, potentially arrhythmia, moderate arrhythmia risk, and highly potential arrhythmia. Finally, the results are sent to a cloud system or web server via Wi-Fi for user access, as illustrated in Figure 1.

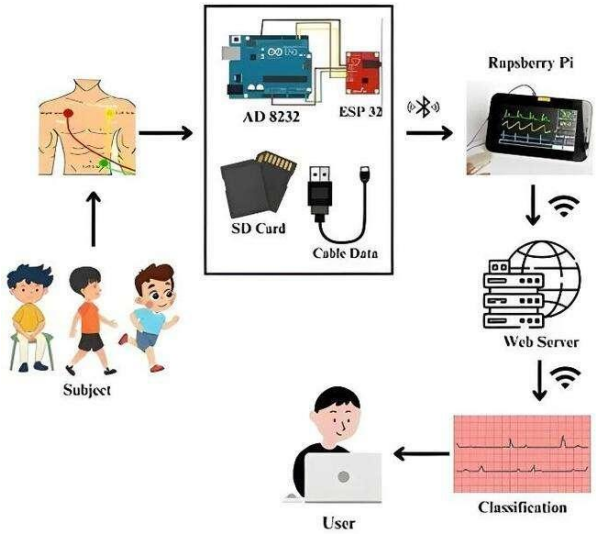



Figure 1. Block diagram

## 3. RESULTS AND DISCUSSION

Arrhythmia, heart attack (myocardial infarction), and heart failure are among the cardiac disorders that can be detected by a medical test known as an ECG. This gadget helps physicians assess a patient's heart rate and rhythm by recording the electrical activity produced by the heart. This method of data processing can increase the effectiveness and precision of diagnosing cardiac conditions. The technique improves diagnostic accuracy and patient health outcomes by reducing the possibility of mistakes that might occur during human interpretation of ECG data. Every participant in the research has a distinct wave pattern with traits that set them apart from other participants. To examine heart rhythm, spot potential arrhythmias, and evaluate general heart health, medical professionals use the ECG data's wave height or amplitude.

The extracted ECG signals are averaged, grouped, and visualized based on predetermined experimental scenarios. The activities processed are sitting, walking, and running. Figure 2 shows that the horizontal axis (X) represents the sequence of data or recorded heartbeats, while the vertical axis (Y) represents the duration of the ECG wave interval in milliseconds (ms). Figure 2(a) displays the RR interval, which is the distance between R peaks related to heartbeats. Figure 2(b) shows the PR interval, which is the time from the start of the P wave to the start of the QRS complex. Figure 2(c) shows the ST interval, which is the time from the end of the QRS complex to the start of the T wave. Figure 2(d) shows the QT interval, which is the time from the start of the QRS complex to the end of the T wave. Meanwhile, Figure 2(e) shows the QS interval, which is the time from point Q to point S on the QRS complex.

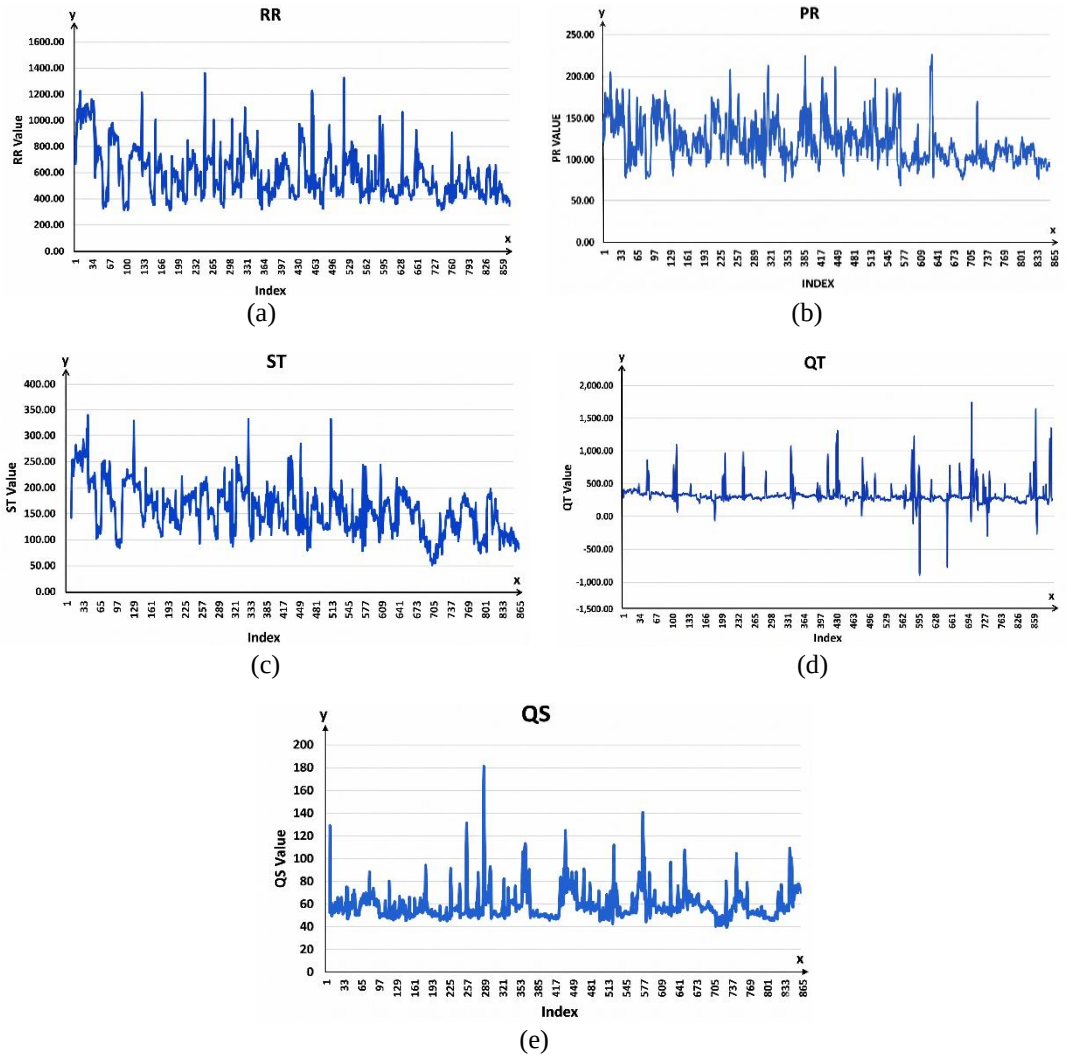



Figure 2. Classification input class; (a) RR, (b) PR, (c) ST, (d) QT, and (e) QS

The classifier uses the visualization as a foundation for identifying the input class. A structure that is used to predict the class or target value of the data during the classification process is created by applying the verified decision rules to the tree. Five classes—normal, aberrant, potentially arrhythmia, moderate arrhythmia risk, and very potentially arrhythmia—are included in this study. The test data's density is taken into consideration while determining the class value limitations. The predictions made by this strategy are based on the majority of the classes present in the leaf nodes. As a result, the prediction outcomes might differ based on how well the tree's separation rules work and how evenly the classes are distributed in the training set. To achieve the best outcomes, building necessitates careful evaluation of pertinent separation qualities and data processing. A structured dataset with seven predictor variables was created by compiling all of the retrieved features. Table 1 displays the interval measurements and derived parameters from the ECG signals, together with the specific results of the feature extraction technique for each individual.

In statistics and data analysis, a scatter plot is a typical graph style used to characterize and assess the connection between two variables with numbers. This graph displays a number of points mapped on a cartesian coordinate system, where each point represents one observation or piece of data. Whereas the position of the points on the vertical axis represents the value of the dependent variable (often the Y-axis), the position of the points on the horizontal axis displays the value of the independent variable (typically the X-axis). Potential correlations between two variables, such as positive, negative, or no discernible correlation (unrelated), can be found using scatter plots. There are five categories of data, namely “Abnormal” (blue), “Normal” (green), “Potential Arrhythmia” (turquoise), “Moderate Arrhythmia Risk” (purple), and “Highly Potential Arrhythmia” (orange) in Figure 3. The ST variable rises as the HR variable falls, as seen in Figure 3(a). The data in Figure 3(b) was dispersed randomly and showed no correlation between the PR and QS variables. The data points in Figure 3(c) cluster around a straight line that rises from left to right,

suggesting a positive correlation between the QT wave and ST variables. On the other hand, the random distribution of data in Figure 3(d) indicates that there is no link between the QS and QT variables.

Table 1. Feature extraction

| Subject | RR | PR | QS | QT | ST | HR | QTC |
|---|---|---|---|---|---|---|---|
| S1 | 769.16 | 177.36 | 127.45 | 270.94 | 155.08 | 78.01 | 202.2326394 |
| S2 | 713.56 | 162.70 | 61.33 | 266.23 | 169.19 | 84.08 | 192.6046577 |
| S3 | 1008.70 | 137.22 | 61.58 | 347.39 | 234.94 | 59.48 | 136.6224082 |
| S4 | 1212.69 | 139.30 | 68.41 | 394.28 | 261.19 | 49.48 | 126.4991742 |
| S5 | 842.86 | 125.93 | 58.73 | 325.93 | 234.39 | 71.19 | 137.1633764 |
| S6 | 1000.75 | 137.05 | 63.25 | 338.10 | 239.46 | 59.95 | 136.9966224 |
| S7 | 978.16 | 128.01 | 67.02 | 328.31 | 228.92 | 61.34 | 129.4330867 |
| S8 | 1033.95 | 120.37 | 64.04 | 385.03 | 248.46 | 58.03 | 118.3776457 |
| S9 | 1033.95 | 130.40 | 62.50 | 371.14 | 257.72 | 58.03 | 128.2424495 |
| S10 | 1087.30 | 189.59 | 60.85 | 373.02 | 266.31 | 55.18 | 181.8236595 |
| S11 | 985.43 | 127.30 | 65.18 | 385.74 | 287.58 | 60.89 | 128.238293 |
| S12 | 990.51 | 170.73 | 69.11 | 345.53 | 250.68 | 60.57 | 171.5472165 |
| S13 | 1098.64 | 176.20 | 62.50 | 383.28 | 274.10 | 54.61 | 168.1083068 |
| S14 | 1087.19 | 128.86 | 64.04 | 375.77 | 246.91 | 55.19 | 123.5829261 |
| S15 | 1119.05 | 134.04 | 69.66 | 390.65 | 255.73 | 53.62 | 126.7086459 |
| S16 | 997.69 | 181.33 | 73.30 | 390.43 | 267.75 | 60.14 | 181.537395 |
| S17 | 1126.10 | 136.68 | 62.61 | 388.01 | 264.55 | 53.28 | 128.8041104 |
| S18 | 1126.98 | 175.49 | 60.85 | 398.59 | 278.66 | 53.24 | 165.3031472 |
| S19 | 1064.86 | 136.72 | 66.61 | 389.13 | 272.57 | 56.35 | 132.4932145 |
| S20 | 1013.98 | 208.85 | 62.89 | 323.67 | 244.31 | 59.17 | 207.4066897 |
| S21 | 1074.70 | 188.26 | 72.41 | 381.10 | 249.24 | 55.83 | 181.6019382 |
| S22 | 1043.45 | 143.29 | 64.02 | 387.96 | 262.20 | 57.50 | 140.2778844 |
| S23 | 1059.96 | 178.13 | 73.19 | 360.67 | 233.69 | 56.61 | 173.0185262 |
| S24 | 1083.33 | 147.38 | 63.27 | 378.09 | 266.20 | 55.38 | 141.5948027 |
| S25 | 1009.26 | 130.40 | 64.04 | 407.41 | 297.07 | 59.45 | 129.8016855 |
| S26 | 1060.19 | 13.35 | 60.19 | 387.35 | 272.38 | 56.59 | 133.3902649 |
| S27 | 1162.26 | 128.75 | 65.26 | 392.42 | 272.49 | 51.62 | 119.4231778 |
| S28 | 1073.30 | 152.78 | 64.81 | 385.80 | 275.46 | 55.90 | 147.4684531 |
| S29 | 1084.88 | 148.15 | 64.04 | 392.75 | 260.03 | 55.31 | 142.2348658 |
| S30 | 1106.48 | 132.72 | 62.50 | 395.83 | 265.43 | 54.23 | 126.1686251 |

Model performance is evaluated by comparing actual and predicted values using a confusion matrix, a tool for two-class or multi-class classification in machine learning. The matrix displays four outcomes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). In the matrix, the vertical axis represents the actual class, while the horizontal axis represents the predicted class. Higher TP values are indicated by darker blue colors, while higher FN values appear as darker red. An ideal classifier produces values only on the diagonal, indicating that all samples are correctly classified.

The approach properly categorized 285 Abnormal test samples since the first row of the first column in Figure 4(a) (in Appendix) has a value of 285 whereas the other three columns have no values. In the second row, the number in the second column is 151, while the value in the third column is 1. As a result, not all Normal test samples were successfully classified by the approach. One test sample was mistakenly identified as a possible arrhythmia, whereas 151 test samples in the Normal class were accurately predicted. The true positive rates (TPR) and false negative rates (FNR) are displayed in Figure 4(b) (in Appendix). TPR displays the proportion of observations that are correctly assigned to their actual class. The percentage of examples that are incorrectly classified based on their actual class is represented by FNR. The false discovery rate (FDR) and positive predictive value (PPV) are displayed in Figure 4(c) (in Appendix).

Based on the metric table per class above, it can be seen that almost all categories have very high precision, recall, and F1-score values, close to 100%. However, there is a striking difference in the Abnormal class, which has a recall of only 66.7%, much lower than other classes. This shows that even though the model is able to recognize most classes very well, it still has difficulty detecting certain cases. Nearly perfect metric values in most classes in a relatively small dataset may indicate overfitting, where the model learns too specifically from the training and validation data, making it appear superior internally but at risk of failing when tested on more varied new data in Table 2.

Receiver operating characteristic (ROC) is a useful tool for assessing a classification model's performance. This curve illustrates the relationship between TPR and FPR at different categorization score thresholds. Each point on the ROC curve represents the correlation between TPR and FPR that corresponds to a certain threshold in the classification score. An indicator of overall classification performance that accounts for all potential threshold values is the area under the ROC curve, or area under the curve (AUC). Better classification ability is shown by higher AUC values, which range from 0 to 1. For the highly potential

arrhythmia class, Figure 5(a) (in Appendix) displays an AUC value of 0.83, whereas Figure 5(b) (in Appendix) displays an AUC value of 0.98. AUC=1 is displayed for the perhaps arrhythmia, moderately arrhythmia, and highly potentially arrhythmia classes, respectively, in Figures 5(c)-(e) (in Appendix). This indicates that all observations were accurately divided into three classes by the categorization model. This suggests that although the categorization model has not yet been able to correctly categorize every observation, it is getting nearby.

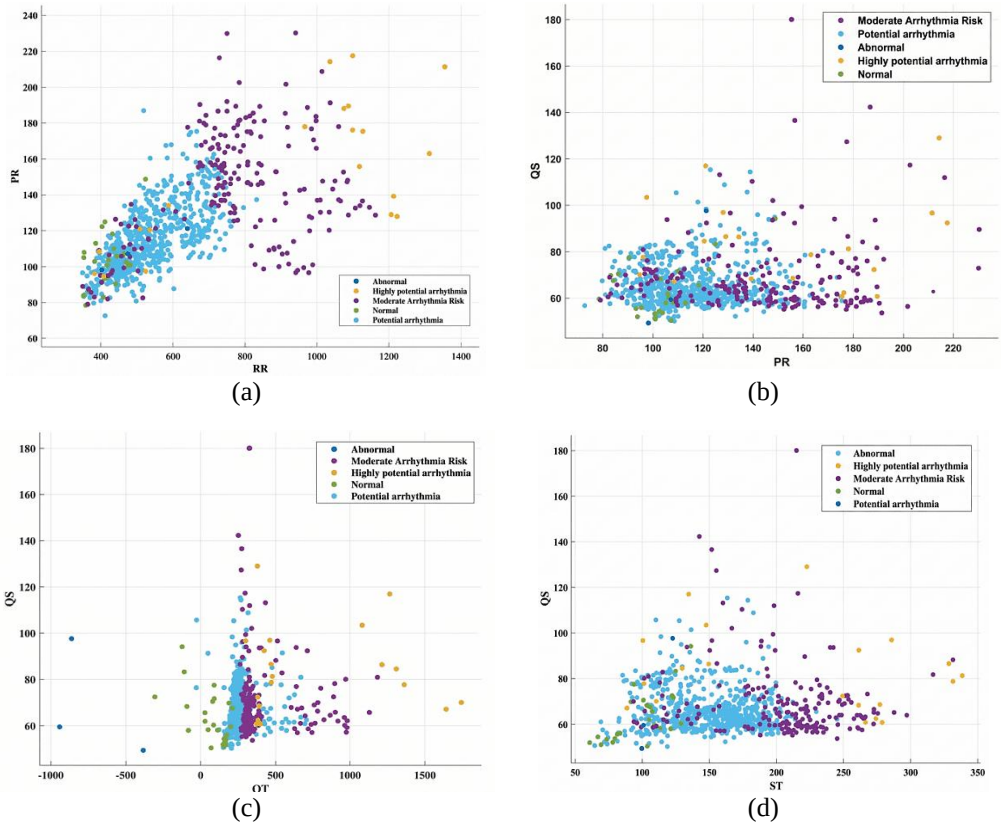



Figure 3. Scatter plot; (a) RR, (b) PR, (c) QT, and (d) ST

Table 2. Per-class metrics

| Class | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|
| Abnormal | 100 | 66.70 | 80.00 |
| Highly potential arrhythmia | 95.20 | 95.20 | 97.60 |
| Moderate arrhythmia risk | 99.50 | 99.50 | 99.50 |
| Normal | 96.20 | 100 | 98.00 |
| Potential arrhythmia | 100 | 99.80 | 99.90 |

## 4. CONCLUSION

This study demonstrates that the decision tree method is an effective tool for early detection of heart abnormalities, achieving 99.4% accuracy, precision, recall, and F1-score, with the advantage of high interpretability that supports practical clinical screening. Despite its strong performance, the model is susceptible to overfitting and may experience decreased effectiveness on complex or imbalanced datasets; therefore, future research should validate it using larger and more heterogeneous data and explore real-time clinical implementation. Scientifically, this research proves that an interpretable decision tree model can achieve near state-of-the-art performance while maintaining transparency, providing a practical, and explainable alternative to complex black-box approaches for early heart abnormality detection.

## ACKNOWLEDGMENTS

This research was primarily funded and supported by Faculty of Science and Technology, Universitas Prima Indonesia in collaboration with Department of Electrical Engineering, Universitas Padjadjaran, Indonesia.

**FUNDING INFORMATION**

Authors state no funding involved.

**AUTHOR CONTRIBUTIONS STATEMENT**

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Putri Juniarti Marpaung | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | |
| Nia Anggreni Matondang | | ✓ | | ✓ | | | | | ✓ | ✓ | | | | |
| Rince Margaretta Siregar | | ✓ | | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Angelica Ester Novelia Siahaan | | ✓ | | ✓ | | | | | ✓ | ✓ | | | | |
| Abdi Dharma | ✓ | | | | ✓ | | | ✓ | | ✓ | | ✓ | | |
| Arjon Turnip | ✓ | | ✓ | ✓ | | | ✓ | | | ✓ | ✓ | | ✓ | ✓ |
| Mardi Turnip | ✓ | | | ✓ | ✓ | ✓ | | ✓ | | ✓ | | ✓ | | |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

**CONFLICT OF INTEREST STATEMENT**

Authors state no conflict of interest.

**INFORMED CONSENT**

We have obtained informed consent from all individuals included in this study.

**ETHICAL APPROVAL**

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

**DATA AVAILABILITY**

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

**REFERENCES**

[1] H. L. Fitriani and K. A. DA Ridwan, "Knowledge on The Risk of The Ischemic Heart Disease Estimation by WHO Charts in Cempaka Baru, Indonesia," *Folia Medica Indonesiana*, vol. 58, no. 3, pp. 215–221, Sep. 2022, doi: 10.20473/fmi.v58i3.33683.

[2] Q. Ayuni, R. C. Wihandika, and N. Yudistira, "Classification of Arrhythmias Based on Electrocardiogram Results Using the Support Vector Machine Method," *Journal of Information Technology and Computer Science Development*, vol. 5, no. 6, pp. 2163–2170, 2021.

[3] Y. Tian *et al.*, "Foundation model of ECG diagnosis: Diagnostics and explanations of any form and rhythm on ECG," *Cell Reports Medicine*, vol. 5, no. 12, p. 101875, Dec. 2024, doi: 10.1016/j.xcrm.2024.101875.

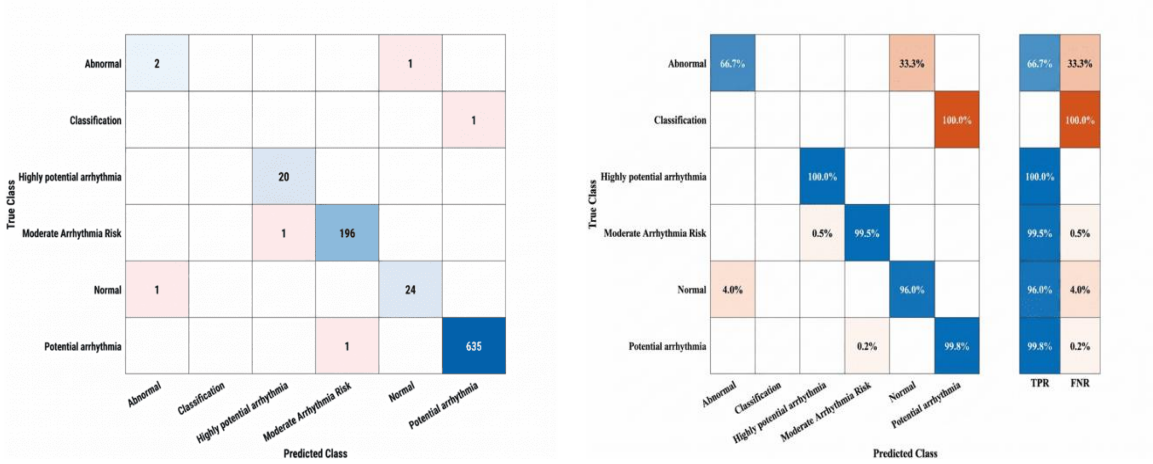
[4] Z. Li *et al.*, "An early warning method for arrhythmias in long-term ECGs based on self-supervised learning and LSTM," *Knowledge-Based Systems*, vol. 327, p. 114137, Oct. 2025, doi: 10.1016/j.knosys.2025.114137.

[5] A. Y. Hannun *et al.*, "Cardiologist-Level Arrhythmia Detection With Convolutional Neural Networks," *Nature Medicine*, vol. 28, no. 1, pp. 65–69, 2022, doi: 10.1038/s41591-018-0268-3.

[6] Z. I. Attia *et al.*, "Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram," *Nature Medicine*, vol. 25, no. 1, pp. 70–74, Jan. 2019, doi: 10.1038/s41591-018-0240-2.

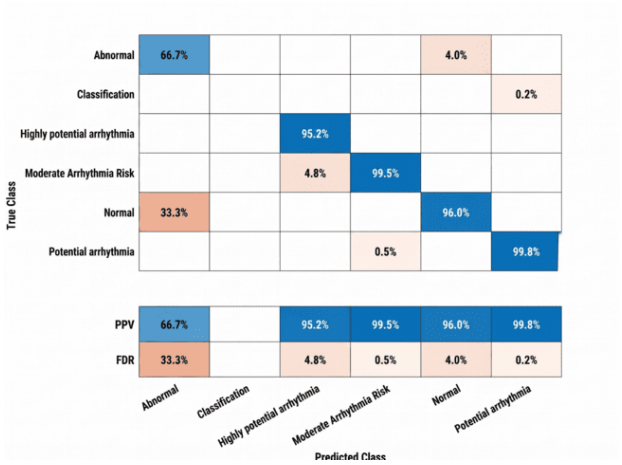
[7] F. Azhari, "Electrocardiogram (ECG)," Halodoc, 2022. https://www.halodoc.com/healt/electrocardiogram-ecg.
[8] A. F. Santamónica, R. Carratalá-Sáez, Y. Larriba, A. Pérez-Castellanos, and C. Rueda, "ECGMiner: A flexible software for accurately digitizing ECG," *Computer Methods and Programs in Biomedicine*, vol. 246, p. 108053, Apr. 2024, doi: 10.1016/j.cmpb.2024.108053.
[9] Y. Wu, S. Cai, X. Xiao, G. Chen, and B. C. Ooi, "Privacy preserving vertical federated learning for tree-based models," in *Proceedings of the VLDB Endowment*, 2020, vol. 13, no. 11, pp. 2090–2103, doi: 10.14778/3407790.3407811.
[10] Z. Jia *et al.*, "TinyML Design Contest for Life-Threatening Ventricular Arrhythmia Detection," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 43, no. 1, pp. 127–140, Jan. 2024, doi: 10.1109/TCAD.2023.3309744.
[11] Y. Qiang, X. Dong, X. Liu, Y. Yang, Y. Fang, and J. Dou, "Conv-RGNN: An efficient Convolutional Residual Graph Neural Network for ECG classification," *Computer Methods and Programs in Biomedicine*, vol. 257, p. 108406, Dec. 2024, doi: 10.1016/j.cmpb.2024.108406.
[12] C. Ding, S. Wang, X. Jin, Z. Wang, and J. Wang, "A novel transformer-based ECG dimensionality reduction stacked auto-encoders for arrhythmia beat detection," *Medical Physics*, vol. 50, no. 9, pp. 5897–5912, Sep. 2023, doi: 10.1002/mp.16534.
[13] M. A. Gelen *et al.*, "TQCPat: Tree Quantum Circuit Pattern-based Feature Engineering Model for Automated Arrhythmia Detection using PPG Signals," *Journal of Medical Systems*, vol. 49, no. 1, p. 38, Mar. 2025, doi: 10.1007/s10916-025-02169-0.
[14] Y. Ansari, O. Mourad, K. Qaraqe, and E. Serpedin, "Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023," *Frontiers in Physiology*, vol. 14, Sep. 2023, doi: 10.3389/fphys.2023.1246746.
[15] N. Alamatsaz, L. S. Tabatabaei, M. Yazdchi, H. Payan, N. Alamatsaz, and F. Nasimi, "A lightweight hybrid CNN-LSTM model for ECG-based arrhythmia detection," *Computers in Biology and Medicine*, vol. 90, 2022, doi: 10.1016/j.bspc.2023.105884.
[16] D. Kim *et al.*, "A novel hybrid CNN-transformer model for arrhythmia detection without R-peak identification using stockwell transform," *Scientific Reports*, vol. 15, no. 1, p. 7817, Mar. 2025, doi: 10.1038/s41598-025-92582-9.
[17] Z. Wu and C. Guo, "Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management," *BioMedical Engineering Online*, vol. 24, no. 1, p. 23, Feb. 2025, doi: 10.1186/s12938-025-01349-w.
[18] A. Søgaard, "On the Opacity of Deep Neural Networks," *Canadian Journal of Philosophy*, vol. 53, no. 3, pp. 224–239, 2023, doi: 10.1017/can.2024.1.
[19] A. Turnip, M. I. Rizqywan, D. E. Kusumandari, M. Turnip, and P. Sihombing, "Classification of ECG signal with Support Vector Machine Method for Arrhythmia Detection," *Journal of Physics: Conference Series*, vol. 970, no. 1, p. 012012, Mar. 2018, doi: 10.1088/1742-6596/970/1/012012.
[20] S. Wang, W. Dai, and G. Y. Li, "Distributionally Robust Receive Combining," *IEEE Transactions on Signal Processing*, vol. 73, pp. 2736–2752, 2025, doi: 10.1109/TSP.2025.3582082.
[21] G. Xu, Y. Shi, X. Sun, and W. Shen, "Internet of things in marine environment monitoring: A review," *Sensors (Switzerland)*, vol. 19, no. 7, p. 1711, Apr. 2019, doi: 10.3390/s19071711.
[22] M. Aqib, R. Mehmood, A. Alzahrani, I. Katib, A. Albeshri, and S. M. Altowaijri, "Smarter traffic prediction using big data, in-memory computing, deep learning and gpus," *Sensors (Switzerland)*, vol. 19, no. 9, p. 2206, May 2019, doi: 10.3390/s19092206.
[23] S. Leonelli and N. Tempini, Eds., *Data journeys in the sciences*. Cambridge, MA: MIT Press, 2020.
[24] N. Stylos and J. Zwiegelaar, "Big data as a game changer: How does it shape business intelligence within a tourism and hospitality industry context?," in *Big Data and Innovation in Tourism, Travel, and Hospitality: Managerial Approaches, Techniques, and Applications*, Singapore: Springer Singapore, 2019, pp. 163–181, doi: 10.1007/978-981-13-6339-9_11.
[25] Q. Song, H. Ge, J. Caverlee, and X. Hu, "Tensor completion algorithms in big data analytics," *ACM Transactions on Knowledge Discovery from Data*, vol. 13, no. 1, pp. 1–48, Feb. 2019, doi: 10.1145/3278607.
[26] R. Hariri, L. Hakim, and R. F. Lestari, "Heart Rate Monitoring System Using Sensors AD8232," *Journal Zetroem*, vol. 2, no. 2, Oct. 2020, doi: 10.36526/ztr.v2i2.1017.
[27] J. J. A. M. Junior *et al.*, "AD8232 to biopotentials sensors: Open source project and benchmark," *Electronics*, vol. 12, no. 4, p. 833, 2023, doi: 10.3390/electronics12040833.

## APPENDIX




(a) (b)

Figure 4. Confusion matrix; (a) number of observations and (b) TPR and FNR




(c)

Figure 4. Confusion matrix; (c) PPV and FDR *(continued)*

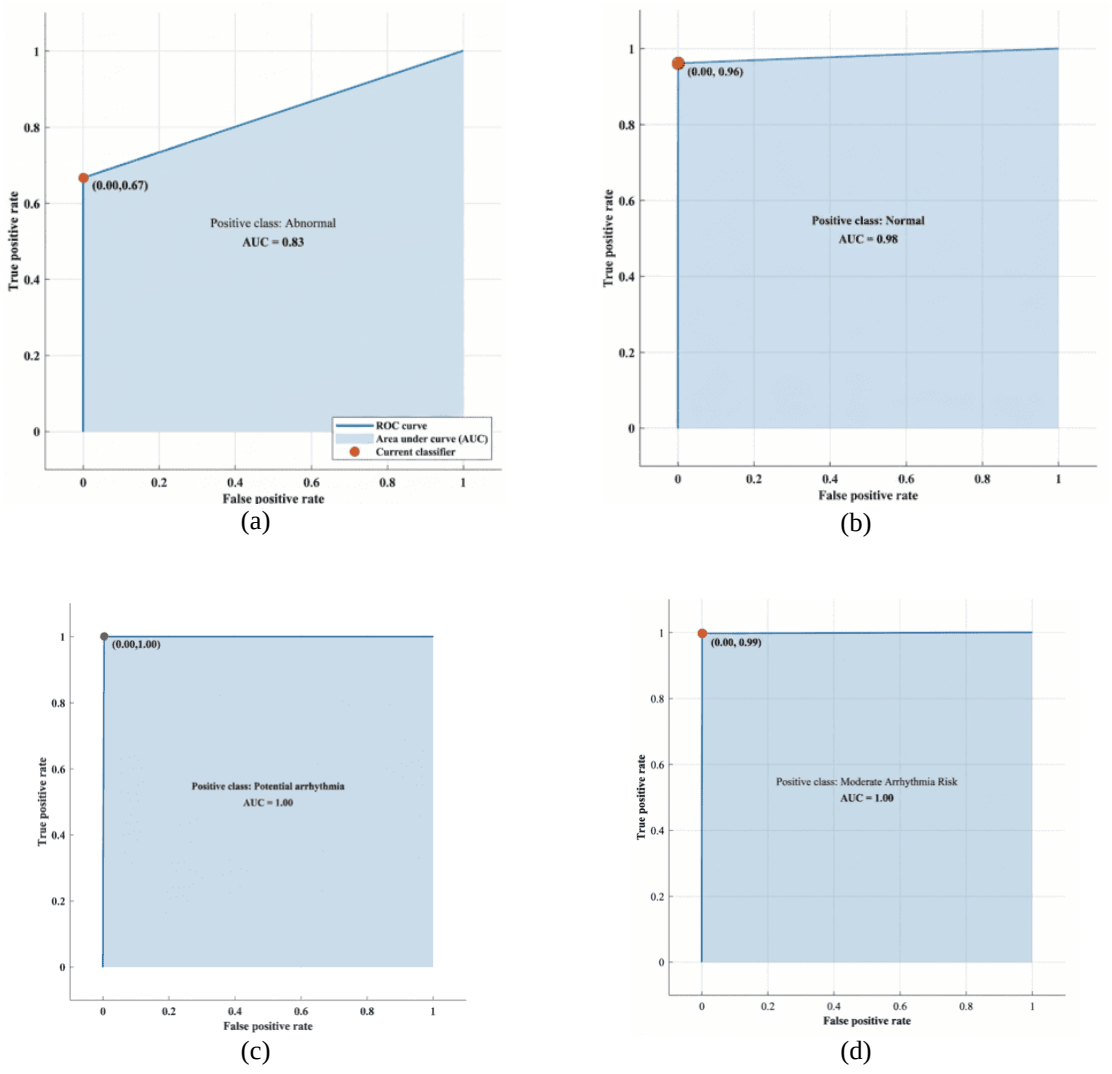
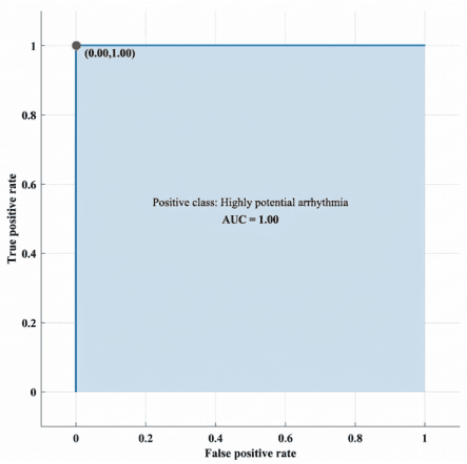



Figure 5. ROC curves representing the average positive data for; (a) abnormal, (b) normal, (c) potential arrhythmia, and (d) moderate potential arrhythmia




(e)

Figure 5. ROC curves representing the average positive data for; (e) highly potential arrhythmia" *(continued)*

## BIOGRAPHIES OF AUTHORS

**Putri Juniarti Marpaung** is an undergraduate student in the Information Systems program at Universitas Prima Indonesia (UNPRI). She is currently beginning her academic journey in the field of research, despite having no prior experience in formal research projects. She can be contacted at email: putrimarpaung28@gmail.com.

**Nia Anggreni Matondang** is an undergraduate student in the Information Systems program at Universitas Prima Indonesia (UNPRI). She is currently beginning her academic journey in the field of research, despite having no prior experience in formal research projects. She can be contacted at email: niaanggrainy3@gmail.com.

**Rince Margaretta Siregar** is an undergraduate student in the Information Systems program at Universitas Prima Indonesia (UNPRI). She is currently beginning her academic journey in the field of research, despite having no prior experience in formal research projects. She can be contacted at email: rincemargaretta@gmail.com.

**Angelica Ester Novelia Siahaan** is an undergraduate student in the Information Systems program at Universitas Prima Indonesia (UNPRI). She is currently beginning her academic journey in the field of research, despite having no prior experience in formal research projects. She can be contacted at email: angelesther873@gmail.com.

**Abdi Dharma** received the Bachelor's degree in Informatics Engineering from STMIK Mikroskil from 1998 to 2003. He obtained a Master's degree in Informatics Engineering from North Sumatra University (USU) from 2012 to 2014. He obtained his doctorate in Computer Science from North Sumatra University (USU) from 2024 to the present. He can be contacted at email: abdidharma@unprimdn.ac.id.

**Arjon Turnip** received the B.Eng. and M.Eng. degrees in Engineering Physics from the Institute of Technology Bandung (ITB), Indonesia, in 1998 and 2003, respectively, and the Ph.D. degree in Mechanical Engineering from Pusan National University, Busan, Korea, under the World Class University program in 2012. He is currently working in the Technical Implementation Unit for Instrumentation Development, Indonesian Institute of Sciences, Indonesia, as a research coordinator. He can be contacted at email: turnip@unpad.ac.id.

**Mardi Turnip** is a lecturer at the Faculty of Science and Technology, Universitas Prima Indonesia, received a Bachelor's degree in Informatics Engineering from the National Institute of Technology (ITENAS) from 2005 to 2010. He obtained a Master's degree in Information Systems from STMIK LIKMI Bandung from 2011 to 2013. He can be contacted at email: marditurnip@unprimdn.ac.id.