Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1277-1295 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Hybrid Time-Frequency ECG Arrhythmia Classification with Feature-Model Compatibility and Ensemble Decision Fusion

J. Prayoga[1], Melinda Melinda[2], Teuku Yuliar Arif[2] and Herlina Dimiati[3]

[1] Doctoral Program of Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia
[2] Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia
[3] Department of Pediatrics, Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh, Indonesia

Corresponding author: Melinda Melinda (e-mail: melinda@usk.ac.id), **Author(s) Email**: J. Prayoga (e-mail: prayoga@mhs.usk.ac.id), Teuku Yuliar Arif (e-mail: yuliar@usk.ac.id), Herlina Dimiati (e-mail: herlinadimiati@usk.ac.id)

Abstract: Cardiac arrhythmia is a critical cardiovascular disorder associated with high global mortality, while electrocardiogram (ECG)-based diagnosis remains time-consuming and susceptible to inter-observer variability. Although recent artificial intelligence approaches have improved automated ECG analysis, many existing studies rely on single time-frequency representations and uniform classification strategies, limiting their ability to capture complementary ECG characteristics. This study proposed a hybrid time-frequency ECG arrhythmia classification framework incorporating feature–model compatibility and ensemble decision fusion to improve classification robustness and reliability. ECG signals from the MIT-BIH Arrhythmia Database were preprocessed using a fourth-order Butterworth bandpass filter and segmented using overlapping windows. To prevent data leakage, patient-wise cross-validation was employed, ensuring that ECG segments originating from the same patient were assigned exclusively to either training or testing folds. Three complementary time-frequency representations, namely Mel-Spectrogram, Short-Time Fourier Transform (STFT), and Discrete Wavelet Transform (DWT), were paired with their most compatible classifiers: Convolutional Neural Network (CNN); Random Forest (RF); and Support Vector Machine (SVM); respectively. Decision fusion was performed using stacking, hard voting, and soft voting strategies. Experimental results showed that the DWT-SVM model with stacking achieved the best overall performance, attaining 94.82% accuracy and a 94.59% F1-score, while STFT-RF with hard voting achieved comparable performance with 94.75% accuracy and a 94.50% F1-score. In contrast, Mel-Spectrogram-CNN produced substantially lower performance with 81.45% accuracy, indicating limited suitability of Mel-scale representations for ECG morphology analysis. Statistical analysis confirmed significant performance differences among models ($p < 0.001$). The findings demonstrate that integrating hybrid time-frequency representations with feature–model compatibility and model-dependent decision fusion provides a robust framework for automated ECG arrhythmia classification with strong potential for clinical decision support and real-time cardiac monitoring applications.

Keywords **Arrhythmia classification; ECG signal processing; Time-frequency analysis; Ensemble learning; Decision fusion.**

I. Introduction

Cardiovascular diseases remain one of the leading causes of global mortality, accounting for millions of deaths annually, with cardiac arrhythmia representing one of the most critical abnormalities due to irregular electrical activity in the heart [1]. Early and accurate detection of arrhythmia is essential because delayed diagnosis may lead to severe complications, including stroke, heart failure, and sudden cardiac death. Electrocardiogram (ECG) signals are widely used as the primary diagnostic tool for identifying cardiac rhythm abnormalities because they provide non-invasive recordings of the heart's electrical activity [2], [3]. However, manual ECG interpretation is time-consuming, highly dependent on clinical expertise, and susceptible to inter-observer variability, particularly in long-term monitoring scenarios. These limitations highlight the importance of developing automated and reliable ECG arrhythmia classification systems capable of assisting clinicians in real-time cardiac diagnosis [4].

Recent advances in artificial intelligence and machine learning have significantly improved automated

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

ECG analysis. Deep learning architectures, particularly Convolutional Neural Networks (CNNs), have demonstrated strong capability in extracting discriminative patterns from ECG representations [5], [6], [7]. In parallel, conventional machine learning approaches such as Support Vector Machine (SVM) and Random Forest (RF) remain effective for structured ECG classification, particularly when combined with wavelet-derived features [8]. These methods provide robust decision boundaries and ensemble-based learning mechanisms for handling non-linear and high-dimensional ECG feature spaces. Furthermore, time-frequency analysis techniques including Short-Time Fourier Transform (STFT), Discrete Wavelet Transform (DWT), and Mel-Spectrogram have been extensively utilized to capture both spectral and temporal ECG characteristics. These approaches have shown promising results in arrhythmia detection by transforming non-stationary ECG signals into more informative representations suitable for machine learning and deep learning models [9], [10].

Despite these advances, several important limitations remain unresolved in current ECG arrhythmia classification studies. First, many existing approaches rely on a single time-frequency representation, limiting their ability to capture complementary ECG characteristics across multiple frequency resolutions. ECG signals are inherently non-stationary and contain both global rhythmic structures and localized transient morphological variations that may not be sufficiently represented using a single transformation technique [11]. Second, most previous studies applied uniform classifiers across heterogeneous feature representations without considering structural compatibility between feature characteristics and learning models [12], [13]. This may reduce the effectiveness of representation learning because different feature types possess distinct spatial, spectral, and temporal properties. Third, although ensemble learning has been widely explored, many studies utilized generalized fusion mechanisms without investigating whether fusion effectiveness depends on the characteristics of the underlying classifiers. In addition, rigorous statistical validation remains insufficiently reported in many existing studies, limiting the reliability and reproducibility of the reported performance improvements [2].

To address these limitations, this study proposed a hybrid time-frequency ECG arrhythmia classification framework incorporating feature–model compatibility and ensemble decision fusion. Three complementary time-frequency representations, namely Mel-Spectrogram, Short-Time Fourier Transform (STFT), and Discrete Wavelet Transform (DWT), were integrated to capture diverse ECG characteristics across different spectral and temporal resolutions [14]. Unlike conventional approaches that apply a single classifier uniformly across all feature representations, the proposed framework explicitly paired each representation with the classifier most compatible with its structural characteristics. Specifically, Mel-Spectrogram representations were processed using CNN due to their spatial time-frequency structure, STFT features were classified using Random Forest to exploit robust ensemble partitioning, and DWT coefficients were analyzed using Support Vector Machine to capture non-linear transient ECG morphology. Furthermore, model outputs were integrated through stacking, hard voting, and soft voting strategies to evaluate model-dependent decision fusion effectiveness [15], [16].

The main contributions of this study are summarized as follows. First, a hybrid multi-representation framework was proposed by integrating Mel-Spectrogram, STFT, and DWT to capture complementary ECG spectral-temporal information. Second, this study introduced a feature–model compatibility strategy that explicitly aligns each time-frequency representation with the most suitable classification model. Third, a model-dependent ensemble fusion strategy was investigated to evaluate the effectiveness of stacking, hard voting, and soft voting across heterogeneous classifiers. Fourth, rigorous statistical validation using Friedman and Wilcoxon signed-rank tests with Bonferroni correction and Cohen's d analysis was conducted to ensure the robustness and significance of the obtained results. The proposed framework was expected to contribute toward more reliable automated ECG arrhythmia classification systems for clinical decision support and real-time cardiac monitoring applications [17].

II. Method

This study adopted a systematic and structured methodology to develop a robust ECG arrhythmia classification framework based on hybrid time-frequency feature extraction and ensemble learning. The overall research workflow consisted of several sequential stages, including data acquisition, signal preprocessing, segmentation, hybrid feature extraction, feature representation, model training, decision fusion, and performance evaluation with statistical validation. Each stage was designed to ensure that both spectral and temporal characteristics of ECG signals were effectively captured and utilized for classification [18]. The complete pipeline of the proposed framework is illustrated in Fig. 1, which highlights the integration of multiple time-frequency representations and heterogeneous classifiers within a unified ensemble learning strategy. As shown in Fig. 1, the proposed approach emphasizes parallel feature

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1277-1295 e-ISSN: 2656-8632

processing and model training, followed by a stacking-based decision fusion mechanism to enhance classification reliability and robustness.

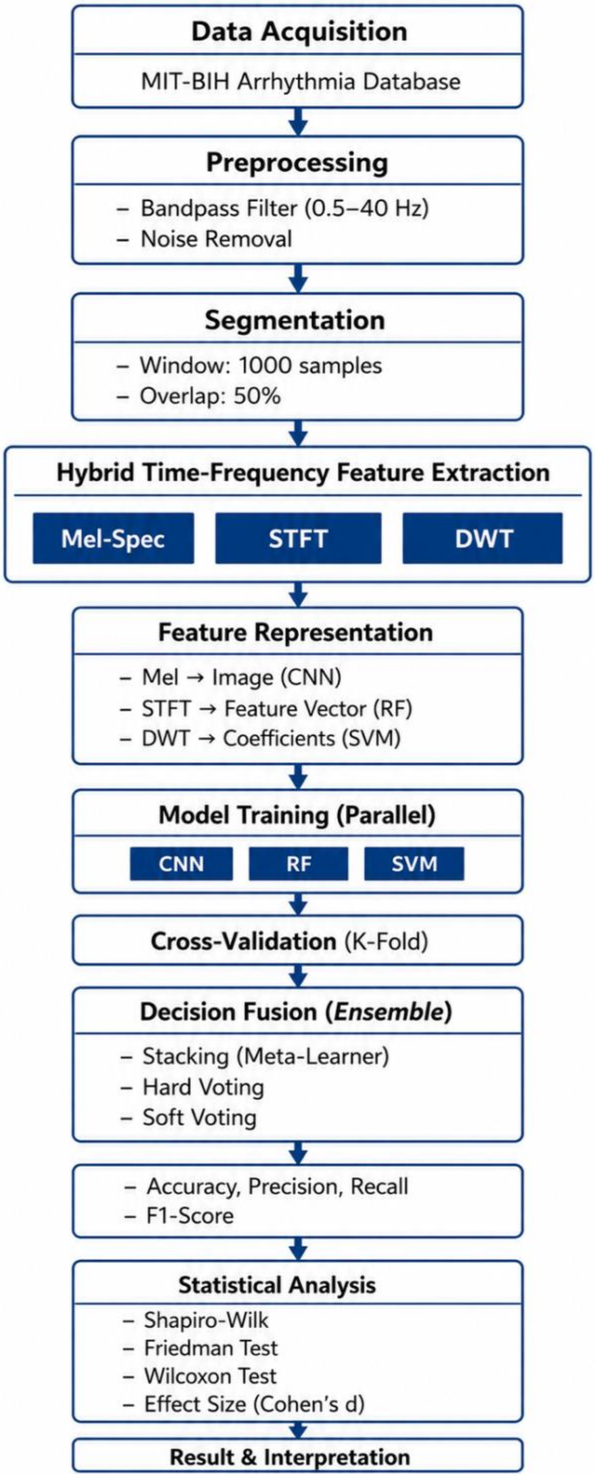


Fig. 1. Proposed research workflow for hybrid ECG arrhythmia classification

A. Data Acquisition

The ECG data used in this study were obtained from the MIT-BIH Arrhythmia Database available through PhysioNet, which is one of the most widely used benchmark datasets for automated arrhythmia classification research. The database contains 48 half-hour ambulatory ECG recordings collected from 47 subjects, sampled at 360 Hz with 11-bit resolution over a 10 mV signal range. The dataset includes both normal sinus rhythms and various clinically significant arrhythmia conditions, making it suitable for evaluating automated ECG classification systems [19]. To construct a balanced binary classification scenario, 11 recordings representing normal cardiac rhythms and 11 recordings containing arrhythmic abnormalities were selected. The selected records were chosen to provide representative ECG variability while maintaining a balanced class distribution between normal and abnormal categories. Binary classification was adopted in this study to focus on robust abnormality detection and to reduce the severe class imbalance commonly encountered in multi-class MIT-BIH classification tasks, The selected MIT-BIH record IDs and their corresponding class assignments are summarized in Table 1.

Table 1. Selected MIT-BIH records used in this study

Class	MIT-BIH Record IDs	Description
Normal	100, 101, 103, 105, 115, 117, 121, 122, 123, 124, 230	Predominantly normal sinus rhythm
Arrhythmia	200, 201, 202, 203, 205, 207, 208, 209, 210, 213, 214	Arrhythmic ECG recordings

Prior to feature extraction, the ECG recordings were segmented into overlapping windows to increase the number of training samples and improve representation learning capability. To avoid classification bias, class balancing was maintained throughout the segmentation and validation processes. Furthermore, to prevent data leakage caused by overlapping ECG segments originating from the same patient, all segments derived from a single patient record were assigned exclusively to either the training or testing fold during patient-wise cross-validation. This strategy ensures that patient-specific ECG morphology does not simultaneously appear in both training and evaluation stages, thereby providing a more realistic assessment of model generalization capability. The resulting dataset provides a reliable benchmark for evaluating the effectiveness of the proposed hybrid time-frequency representation framework and ensemble

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

decision fusion strategy for automated ECG arrhythmia classification.

B. Preprocessing

Raw ECG signals are inherently susceptible to various forms of noise and artifacts, including baseline wander, muscle noise, power-line interference, and motion artifacts, which may negatively affect the performance of automated arrhythmia classification systems. Therefore, a preprocessing stage was performed to enhance signal quality while preserving diagnostically important ECG morphology, particularly the P-wave, QRS complex, and T-wave components [20]. In this study, a fourth-order Butterworth bandpass filter with cutoff frequencies of 0.5–40 Hz was applied to suppress low-frequency baseline drift and high-frequency noise while maintaining clinically relevant cardiac waveform characteristics. The Butterworth filter was selected because of its smooth frequency response and minimal signal distortion, making it highly suitable for biomedical signal processing applications. To avoid phase distortion, zero-phase filtering was implemented during the filtering process. The filtering operation is expressed in Eq. (1) [20]:

$$y(t) = x(t) * h(t) \tag{1}$$

where $x(t)$ represents the raw ECG signal, $h(t)$ denotes the impulse response of the Butterworth filter, and $y(t)$ is the filtered ECG output. The symbol $*$ indicates the convolution operation between the ECG signal and the filter response. Following the filtering stage, z-score normalization was applied to standardize the ECG amplitude distribution across recordings. The normalization operation is defined in Eq. (2) [21]:

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

where x denotes the ECG sample value, μ represents the mean signal amplitude, and σ is the standard deviation of the ECG signal. After normalization, the filtered ECG signals were segmented into overlapping windows to increase the number of training samples and preserve localized temporal ECG patterns. A window length of 1000 samples was selected, corresponding to approximately 2.78 seconds at a sampling frequency of 360 Hz. This duration is sufficient to capture multiple cardiac cycles while maintaining important temporal morphology information. Overall, the preprocessing stage plays a critical role in improving signal reliability and ensuring that subsequent time-frequency feature extraction methods operate on standardized and noise-reduced ECG inputs.

C. Segmentation

The segmentation stage was performed to transform continuous ECG recordings into structured fixed-length signal windows suitable for automated feature extraction and classification. Since ECG signals are inherently non-stationary and contain temporal variations across cardiac cycles, segmentation enables localized analysis of ECG morphology while increasing the number of training samples available for machine learning models [22], [23]. After preprocessing, each filtered ECG recording was divided into overlapping segments with a fixed window length of 1000 samples. Considering the MIT-BIH sampling frequency of 360 Hz, each segment corresponded to approximately 2.78 seconds of ECG activity. This duration was selected because it was sufficient to capture multiple cardiac cycles, including diagnostically important waveform components such as the P-wave, QRS complex, and T-wave, while maintaining computational efficiency for subsequent feature extraction and classification processes. Following the fixed-length windowing procedure employed in recent ECG classification studies, the i-th ECG segment is defined in Eq. (3) [22], [24]:

$$s_i = x[i : i + L] \tag{3}$$

where s_i represents the i-th ECG segment, x denotes the preprocessed ECG signal, and L is the window length of 1000 samples. To improve temporal continuity and increase the number of training samples, a 50% overlap was applied between consecutive segments. The overlap length is specified in Eq. (4) [22], [24]:

$$O = \frac{L}{2} \tag{4}$$

where O denotes the overlap length and L represents the segment window size. Consequently, the step size between two consecutive segments was 500 samples. The total number of segments generated from all ECG recordings was calculated using Eq. (5) [22]:

$$N_s = \sum_{r=1}^{N_r} n_r \tag{5}$$

where N_s denotes the total number of generated segments, N_r is the number of ECG recordings, and n_r represents the number of segments obtained from the r-th recording. To prevent data leakage caused by overlapping segments originating from the same patient, patient-wise cross-validation was employed throughout the experimental process. All segments generated from a single patient record were assigned exclusively to either the training or testing fold, ensuring that patient-specific ECG morphology did not simultaneously appear during training and evaluation stages. This strategy provides a more realistic assessment of model generalization capability and reduces optimistic bias in classification performance evaluation. Overall, the segmentation stage plays a crucial role in preserving localized ECG temporal patterns, increasing data availability, and ensuring

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

robust and leakage-free model evaluation prior to hybrid time-frequency feature extraction.

D. Hybrid Time-Frequency Feature Extraction

ECG signals are inherently non-stationary and contain complex spectral-temporal characteristics that vary across different cardiac conditions. Therefore, relying on a single feature extraction technique may limit the ability of the classification system to capture comprehensive ECG morphology. To address this limitation, this study proposed a hybrid time-frequency feature extraction framework integrating three complementary representations: Mel-Spectrogram, Short-Time Fourier Transform (STFT), and Discrete Wavelet Transform (DWT). These transformations were selected because they capture complementary ECG characteristics across multiple temporal and frequency resolutions [25]. Accordingly, for each segmented ECG signal $x(t)$, the proposed hybrid feature set comprising three parallel time-frequency representations is defined in Eq. (6) [25]:

$$F = \{F_{Mel}, F_{STFT}, F_{DWT}\} \quad (6)$$

where F_{Mel}, F_{STFT}, and F_{DWT} denote the Mel-Spectrogram, STFT, and DWT feature representations, respectively.

E. Short-Time Fourier Transform (STFT)

The Short-Time Fourier Transform converts ECG signals into a fixed-resolution time-frequency representation by applying Fourier analysis within short overlapping windows. STFT is particularly effective for capturing global rhythmic patterns and stable frequency distributions in ECG signals. The STFT transformation was computed using Eq. (7) [21]:

$$STFT\{x(t)\}(\tau, f) = \int_{-\infty}^{\infty} x(t) w(t-\tau)\, e^{-j2\pi f t} dt \quad (7)$$

where $x(t)$ denotes the ECG signal, $w(t-\tau)$ represents the sliding window function centered at time τ, and f denotes the frequency component. The resulting STFT spectrogram provides stable time-frequency representations capable of highlighting rhythmic ECG patterns associated with arrhythmic abnormalities. Due to its fixed-frequency resolution, STFT is effective for capturing long-duration spectral structures in ECG recordings.

F. Discrete Wavelet Transform (DWT)

Unlike STFT, the Discrete Wavelet Transform performs multi-resolution analysis, enabling simultaneous localization in both the time and frequency domains. This capability is particularly important for ECG analysis because arrhythmia-related abnormalities often appear as transient morphological variations within short temporal intervals. Previous ECG classification research has also demonstrated that DWT-based representations can preserve diagnostically relevant cardiac patterns for downstream deep-learning classification [26]. The discrete wavelet coefficients are computed using Eq. (8) [21], [26], [27]:

$$DWT(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-b}{a}\right) dt \quad (8)$$

where a denotes the scaling parameter, b represents the translation parameter, and ψ is the mother wavelet function. The multi-resolution capability of DWT allows effective extraction of localized ECG features such as abrupt QRS complex variations, transient waveform irregularities, and short-duration cardiac abnormalities. This makes DWT highly suitable for arrhythmia classification tasks involving non-stationary ECG signals.

G. Mel-Spectrogram

The Mel-Spectrogram transforms ECG signals into perceptually scaled frequency representations by mapping linear frequencies onto the Mel scale. The Mel-frequency conversion is defined in Eq. (9) [28]:

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (9)$$

where f represents the frequency in Hertz. The Mel-Spectrogram produces two-dimensional time-frequency images that preserve spatial spectral information suitable for convolutional neural networks. Although Mel-scale representations are widely used in speech processing, their effectiveness for ECG analysis remains challenging because ECG morphology relies heavily on subtle high-frequency transient characteristics that may be compressed during Mel-frequency scaling. Unlike conventional approaches that rely on a single representation, the proposed framework preserves representation diversity by processing Mel-Spectrogram, STFT, and DWT features independently. This strategy enables the framework to capture complementary ECG characteristics across multiple spectral and temporal resolutions. Specifically, STFT captures stable rhythmic frequency structures, DWT focuses on transient morphological variations, and Mel-Spectrogram provides structured spatial spectral representations. The outputs of these three transformations were subsequently combined into a unified hybrid feature representation, as shown in Fig. 2. This integration enables the model to leverage complementary information from different domains, including global spectral patterns and localized temporal variations. By systematically combining these heterogeneous features, the proposed approach enhances feature diversity and improves the robustness of the classification system compared to single-representation methods.

H. Feature Representation

Following the hybrid time-frequency feature extraction stage, the resulting ECG representations were transformed into model-specific feature spaces

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

according to their structural characteristics. Unlike conventional approaches that apply a single classifier uniformly across all feature representations, this study introduced a feature–model compatibility strategy that explicitly aligned each representation with the classifier most suitable for its structural properties. This strategy was designed to maximize representation learning effectiveness while preserving the unique spectral-temporal information captured by each transformation method [29]. The hybrid feature set is restated in Eq. (10) [21]:

$$F = \{F_{Mel}, F_{STFT}, F_{DWT}\} \tag{10}$$

where F_{Mel}, F_{STFT}, and F_{DWT} denote the Mel-Spectrogram, STFT, and DWT feature representations, respectively.

1. Mel-Spectrogram Representation For CNN

The Mel-Spectrogram representation produced two-dimensional time-frequency images that preserve spatial spectral patterns over time. These representations are particularly suitable for Convolutional Neural Networks (CNNs), which are capable of learning hierarchical spatial features through convolutional operations. Image-based ECG classification studies have also demonstrated the capability of CNN architectures such as EfficientNet and InceptionNet to identify discriminative patterns from two-dimensional arrhythmia representations [30]. The Mel-Spectrogram is represented as a two-dimensional matrix in Eq. (11) [28], [30]:

$$M \in \mathbb{R}^{m \times n} \tag{11}$$

where m and n represent the frequency and temporal dimensions of the spectrogram image. CNN architectures are highly effective for extracting local spatial correlations from image-based representations. Therefore, Mel-Spectrogram features were paired with CNN to enable automatic extraction of spatial ECG

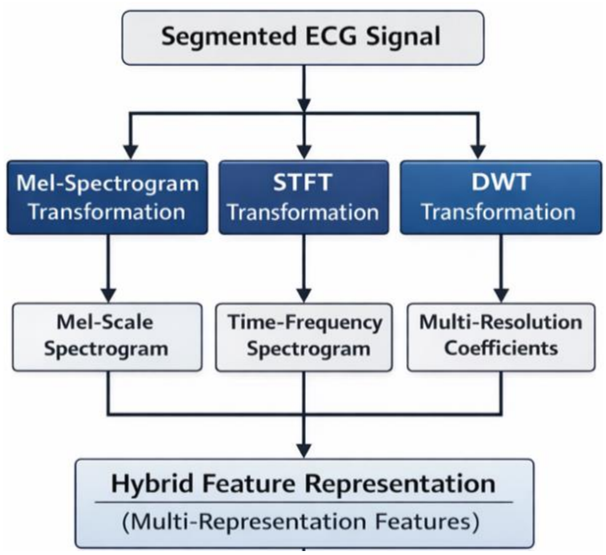


Fig. 2. Hybrid Time-Frequency Feature Extraction

patterns from the generated time-frequency images.

2. STFT Feature Representation for Random Forest

The STFT representation generated stable fixed-resolution time-frequency distributions that preserve rhythmic spectral structures within ECG signals. To enable machine learning-based classification, the STFT spectrograms were reshaped into one-dimensional feature vectors in Eq. (12) [14]:

$$V_{STFT} \in \mathbb{R}^{d} \tag{12}$$

where d denotes the dimensionality of the flattened STFT feature vector. These vectorized representations are well suited for Random Forest classifiers because ensemble decision trees can effectively partition high-

Algorithm 1. Proposed Hybrid ECG Arrhythmia Classification Framework

Input: ECG recordings
Output: Arrhythmia classification result
1: Preprocess ECG signals using Butterworth filtering
2: Normalize ECG amplitudes using z-score normalization
3: Segment ECG into overlapping windows
4: Split data using patient-wise cross-validation
5: Extract DWT, STFT, and Mel-Spectrogram features
6: Train SVM using DWT features
7: Train RF using STFT features
8: Train CNN using Mel-Spectrogram features
9: Generate classifier outputs
10: Apply stacking, hard voting, and soft voting
11: Obtain final prediction
12: Evaluate using Accuracy, Precision,
Recall, and F1-score
13: Perform Friedman and Wilcoxon tests
14: Report classification performance

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1277-1295 **e-ISSN: 2656-8632**

dimensional feature spaces while maintaining robustness against noise and overfitting. The ensemble structure of Random Forest also enables stable classification performance across varying ECG morphological patterns.

3. DWT Coefficient Representation for SVM

The Discrete Wavelet Transform produces multi-resolution coefficients capable of capturing localized transient ECG characteristics across multiple frequency scales. The resulting wavelet coefficients are represented in Eq. (13) [27]:

$$W = \{w_1, w_2, \dots, w_n\} \tag{13}$$

where w_i denotes the wavelet coefficients obtained from different decomposition levels. These coefficients preserve both temporal and frequency-localized ECG morphology, making them highly suitable for Support Vector Machine (SVM) classifiers equipped with non-linear kernels such as the Radial Basis Function (RBF). SVM is particularly effective for separating complex arrhythmic patterns in high-dimensional feature spaces generated by wavelet decomposition.

The proposed feature–model compatibility strategy ensures that each ECG representation was processed using the classifier most capable of exploiting its structural characteristics. Specifically, CNN was utilized for spatial image-based Mel representations, Random Forest was employed for structured STFT feature vectors, and SVM was applied to multi-resolution DWT coefficients. Unlike generalized classification frameworks that apply identical learning mechanisms across heterogeneous features, the proposed approach preserves representation diversity and enables representation-specific learning. This design allows complementary ECG characteristics extracted from different time-frequency domains to be learned more effectively, thereby improving classification robustness and overall arrhythmia detection performance. The resulting model outputs were subsequently integrated within the ensemble decision fusion stage to produce the final arrhythmia classification. As shown in Fig. 3, the proposed mapping strategy aligns each feature representation with its most suitable classifier, improving learning effectiveness and overall classification robustness. To improve methodological clarity and reproducibility, the overall procedure of the proposed hybrid ECG arrhythmia classification framework is summarized in Algorithm 1. The algorithm describes the complete workflow, including ECG preprocessing, segmentation, hybrid time-frequency feature extraction, feature–model compatibility mapping, model training, ensemble decision fusion, and statistical evaluation. As shown in Algorithm 1, the proposed framework performs parallel feature extraction and classifier training using representation-specific learning models. The resulting predictions were subsequently integrated through multiple decision fusion strategies to obtain the final arrhythmia classification.

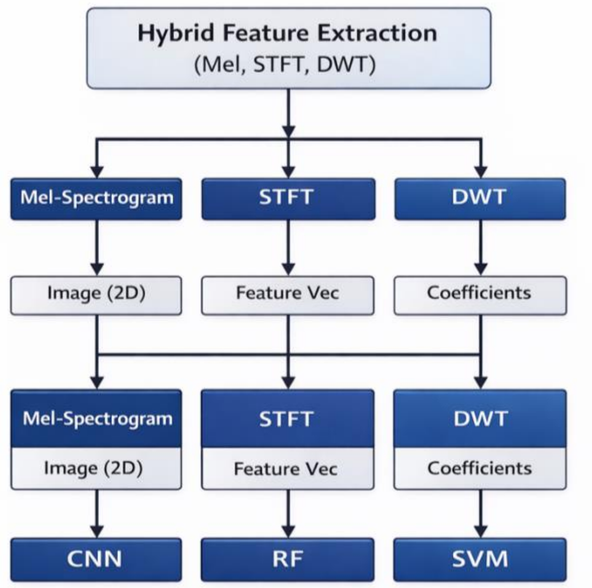


Fig. 3. Mapping of hybrid time-frequency features to classification models

I. Model Training

After transforming the extracted features into their respective representations, the classification models were trained in parallel using their corresponding inputs. This parallel learning strategy enables each model to independently capture discriminative patterns from its associated feature space, thereby maximizing the benefits of heterogeneous representations [31], [32]. The representation–model pairs used in the proposed framework are defined in Eq. (14) [29]:

$$R_i = \{(F_{mel}, CNN), (F_{stft}, RF), (F_{dwt}, SVM)\} \tag{14}$$

Each classifier learns a representation-specific mapping function, as expressed in Eq. (15) [29]:

$$\begin{aligned} f_{cnn} &: F_{mel} \rightarrow y_{cnn} \\ f_{rf} &: F_{stft} \rightarrow y_{rf} \\ f_{svm} &: F_{dwt} \rightarrow y_{svm} \end{aligned} \tag{15}$$

where y_{cnn}, y_{rf}, y_{svm} denote the predicted outputs from each classifier. The empirical classification loss for each model is computed using Eq. (16) [33]:

$$L_m = \frac{1}{N} \sum_{i=1}^{N} l(y_i, f_m(x_i)) \tag{16}$$

where $m \in \{cnn, rf, svm\}, y_i$ the ground-truth label, and $l(\cdot)$ is the loss function. For binary arrhythmia classification, the CNN was optimized using the binary cross-entropy loss defined in Eq. (17) [33]:

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

$$L_{cnn} = \sum_{i=1}^{N} y_i \log(y_{cnn,i)} \quad (17)$$

which is optimized using gradient-based learning. The Random Forest class probability is obtained by averaging the probability estimates generated by its T constituent decision trees, as expressed in Eq. (18) [21]:

$$y_{rf} = \frac{1}{T}\sum_{i=1}^{T} T_j(F_{stft}) \quad (18)$$

where T is the number of trees. The Support Vector Machine (SVM) model learns a decision boundary by maximizing the margin between classes through kernel transformation. The resulting set of trained prediction functions is summarized in Eq. (19) [29]:

$$F_{train} = \{f_{cnn}, f_{rf}, f_{svm}\} \quad (19)$$

This parallel formulation ensures that each classifier specializes in its respective feature domain without interference from other representations. By allowing independent optimization across models, the framework preserves representation-specific learning while enabling complementary knowledge extraction. The resulting model outputs serve as inputs for the subsequent decision fusion stage, where they are combined to produce the final classification.

J. Cross-Validation

To ensure the robustness and generalizability of the proposed classification framework, a k-fold cross-validation strategy was employed during model training and evaluation. This approach is particularly important given the limited number of subjects, as it reduces overfitting and provides a more reliable estimation of model performance. Formally, the dataset D was partitioned into k mutually exclusive subsets according to Eq. (20) [34]:

$$D = \bigcup_{i=1}^{k} D_i, D_i \cap D_j = \emptyset \; for \; i \neq j \quad (20)$$

At each iteration i, one subset D_i was used for testing, while the remaining subsets constituted the training set, as defined in Eq. (21) [34]:

$$D_{train}^{(i)} = D | D_i, D_{val}^{(i)} = D_i \quad (21)$$

The model was trained and evaluated across all k folds, ensuring that each subset was used exactly once for validation. The overall cross-validation performance was then calculated as the mean of the fold-level evaluation scores using Eq. (22) [33], [34]:

$$\varepsilon = \frac{1}{k}\sum_{i=1}^{k} E_i \quad (22)$$

where E_i denotes the evaluation metric (e.g., accuracy or F1-score) for fold i. To improve the reliability of the evaluation, all models were trained and validated consistently across the same fold partitions. This ensures a fair comparison between different feature–model combinations and decision fusion strategies. Furthermore, the use of cross-validation enables the framework to be evaluated on varying data distributions, thereby improving result stability and reducing bias. This is particularly important for ECG datasets with limited subject diversity, where data variability is inherently constrained. To ensure a rigorous and leakage-free evaluation, a five-fold patient-wise cross-validation strategy was employed throughout all experiments. The selected 22 MIT-BIH records (11 normal and 11 arrhythmia records) were partitioned at the patient level into five mutually exclusive folds. Unlike conventional segment-wise validation, where ECG segments from the same patient may appear simultaneously in both training and testing sets, the proposed protocol strictly assigns all segments originating from a given MIT-BIH record to a single fold only. During each iteration, one fold was used as the testing set while the remaining four folds were used for training. Consequently, approximately 80% of the patient records were used for training and 20% for testing in each fold. This procedure was repeated five times so that every patient record served as testing data exactly once. The patient-wise partitioning strategy prevents patient-specific ECG morphology from leaking into the training process and provides a more realistic assessment of model generalization capability. Table 2 summarizes the distribution of MIT-BIH records across the five patient-wise cross-validation folds used in this study.

As shown in Table 2, all ECG segments generated from a specific MIT-BIH record were maintained within the same fold throughout the experimental process. Therefore, no ECG segment originating from the same patient appeared simultaneously in both training and testing sets, eliminating potential data leakage and ensuring fair model evaluation.

Table 2. Patient-Wise Cross-Validation Folds

Fold	Normal Records	Arrhythmia Records	Total Records
Fold 1	100, 101	200, 201	4
Fold 2	103, 105	202, 203	4
Fold 3	115, 117	205, 207	4
Fold 4	121, 122	208, 209	4
Fold 5	123, 124, 230	210, 213, 214	6

K. Decision Fusion

To improve classification performance and enhance system reliability, a decision fusion strategy was employed to combine the outputs of multiple classifiers [35], [36]. Given the trained models f_{cnn}, f_{rf}, and f_{svm}, each model produced a prediction for a given input defined in Eq. (23) [35], [36]:

$$y_{cnn}, y_{rf}, y_{svm} \quad (23)$$

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

These outputs were aggregated into a decision vector shown in Eq. (24) [35]:

$$\mathbf{z} = [\hat{y}_{cnn}, \hat{y}_{rf}, \hat{y}_{svm}] \tag{24}$$

The proposed framework primarily adopted a stacking-based fusion strategy, where a meta-learner $g(\cdot)$ is trained to map the decision vector to the final prediction according to Eq. (25) [35]:

$$\hat{y} = g(\mathbf{z}) \tag{25}$$

Unlike conventional ensemble methods, stacking enables the learning of non-linear relationships between model outputs, allowing adaptive weighting of each classifier based on its predictive behavior. This makes stacking particularly effective for combining heterogeneous models trained on different feature representations. For comparison, two alternative fusion strategies were also implemented. In hard voting, the final prediction is determined by majority rule according to Eq. (26) [35]:

$$\hat{y} = \text{mode}(\hat{y}_{cnn}, \hat{y}_{rf}, \hat{y}_{svm}) \tag{26}$$

In soft voting, the predicted probabilities from each model are averaged according to Eq. (27) [35]:

$$\hat{y} = \arg\max\left(\frac{1}{3}\sum_{m\in\{cnn,rf,svm\}} p_m\,(y \mid x)\right) \tag{27}$$

where $p_m(y \mid x)$ denotes the predicted probability of class y from model m.

Unlike voting-based approaches, which assume equal contribution from all models, stacking learns a data-driven fusion function that captures interactions among model outputs. This is particularly important in the proposed framework, where each classifier was trained on a different feature representation and exhibits distinct prediction characteristics. Therefore, the complete model-specific decision-fusion function is summarized in Eq. (28) [35], [36]:

$$\hat{y} = g(f_{cnn}(F_{mel}),\ f_{rf}(F_{stft}),\ f_{svm}(F_{dwt})) \tag{28}$$

This formulation highlights that the proposed framework performs model-specific learning followed by adaptive decision-level integration. By combining complementary predictions from heterogeneous models, the fusion strategy enhances classification robustness and reduces the impact of individual model errors.

L. Evaluation Metrics

To evaluate the performance of the proposed classification framework, several standard evaluation metrics are employed, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the model's ability to distinguish between normal and arrhythmic ECG signals. Let $y$denote the ground-truth label and $\hat{y}$the predicted label. The classification outcomes were defined in terms of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [37]. Accuracy, precision, recall, and F1-score were calculated using Eq. (29) [33], Eq. (30) [33], Eq. (31) [33], and Eq. (32) [33], respectively. The overall classification accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{29}$$

Precision measures the proportion of correctly predicted positive instances:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{30}$$

Recall (or sensitivity) measures the proportion of actual positive instances that are correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{31}$$

To provide a balanced evaluation between precision and recall, the F1-score is defined as:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{32}$$

These evaluation metrics were computed for each individual classification model (CNN, RF, and SVM) as well as for each decision fusion strategy (stacking, hard voting, and soft voting). The final reported performance was obtained by averaging the results across all cross-validation folds. By incorporating both overall accuracy and class-specific metrics, the evaluation provided a balanced and reliable assessment of classification performance, particularly in handling potential class imbalance and model-specific prediction behavior.

M. Statistical Analysis

To ensure that the observed differences in model performance are statistically significant and not due to random variation, a comprehensive statistical analysis was conducted on the cross-validation results [34]. Let $E_i^{(m)}$ denote the evaluation metric (e.g., accuracy or F1-score) obtained by model m on fold i, where $i = 1,2,\ldots,k$. The performance of each model was therefore represented by the vector defined in Eq. (33) [33]:

$$\mathbf{E}^{(m)} = \{E_1^{(m)}, E_2^{(m)}, \ldots, E_k^{(m)}\} \tag{33}$$

Initially, the Shapiro Wilk test was applied to assess the normality of the performance distributions. Since the data do not follow a normal distribution, non-parametric statistical tests are employed. The Friedman test was used as the primary statistical method to compare multiple models across the same folds. The test statistic was computed based on the ranking of model performances across folds, allowing the detection of

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

significant differences among related samples. A significance level of $\alpha = 0.05$ was adopted. When the Friedman test indicates statistically significant differences, post-hoc pairwise comparisons were performed using the Wilcoxon signed-rank test, as defined in Eq. (34) [33]:

$$W = \sum_{i=1}^{k} \text{sign}(d_i) \cdot r_i \tag{34}$$

where d_i is the difference between paired observations

Table 3. Summarizes the optimized hyperparameters used in this study

Model (Input)	Key Hyperparameters	Value
SVM (DWT)	Kernel	RBF
	C (Regularization)	10
	Gamma	0.01
	Optimizer	SMO
Random Forest (STFT)	n_estimators	100
	Criterion	Entropy
	Max Depth	None
	Min Samples Split	2
CNN (Mel-Spectrogram)	Optimizer	Adam (LR = 0.001)
	Batch Size	32
	Dropout	0.5
	Activation	ReLU, Softmax

and r_i is the rank of the absolute differences. Bonferroni correction was applied to control the family-wise error rate. In addition to statistical significance, effect size analysis was conducted using Cohen's d calculated using Eq. (35) [38]:

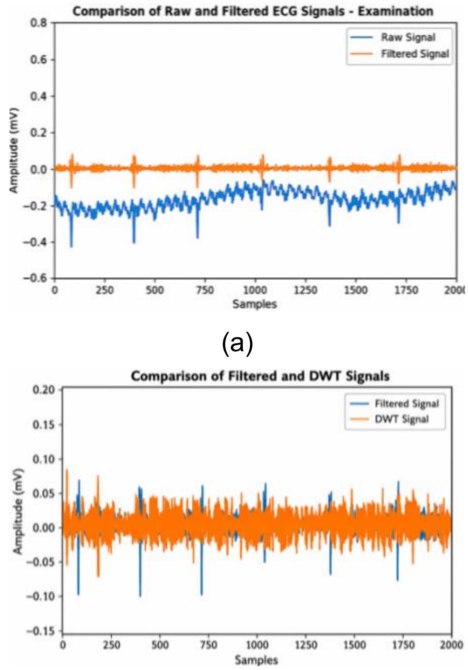
$$d = \frac{\mu_1 - \mu_2}{\sigma_p} \tag{35}$$

where μ_1 and μ_2 are the mean performances of two models, and σ_p is the pooled standard deviation. By combining normality testing, non-parametric comparison, post-hoc analysis, and effect size measurement, this study provided a robust and reliable statistical validation of the proposed hybrid classification framework.

III. Result

A. Preprocessing and Feature Extraction Analysis

The preprocessing stage plays a critical role in enhancing ECG signal quality by removing noise while preserving diagnostically relevant morphological features. As shown in Fig. 4(a), the application of the Butterworth bandpass filter (0.5–40 Hz) effectively eliminates baseline wander and high-frequency noise without distorting the QRS complex. This indicates that the filtering process successfully maintains the structural integrity of key ECG components, which is essential for reliable feature extraction. Furthermore, the effectiveness of the Discrete Wavelet Transform (DWT) in capturing signal characteristics is illustrated in Fig. 4(b). The DWT-extracted signal retains the essential waveform patterns while reducing signal amplitude and noise components. This multi-resolution decomposition enables better localization of transient features, which are critical for identifying arrhythmic events. The comparative visualization of feature extraction methods is presented in Fig. 5. As observed in Fig. 5(a), the DWT spectrogram highlights transient variations across frequency bands, making it particularly effective in capturing irregular and localized signal changes associated with arrhythmia. In contrast, the STFT spectrogram in Fig. 5(b) provides a consistent time-frequency representation with fixed resolution, clearly showing energy concentrations around the 45–50 Hz range during arrhythmic segments. Meanwhile, the Mel-Spectrogram representation shown in Fig. 5(c) emphasizes lower-frequency components due to its perceptual scaling. Although this results in visually smooth and structured representations, it may suppress high-frequency details that are important for detecting subtle abnormalities in ECG signals. This limitation suggests that Mel-based representations may be less effective in capturing fine-grained arrhythmic patterns compared to

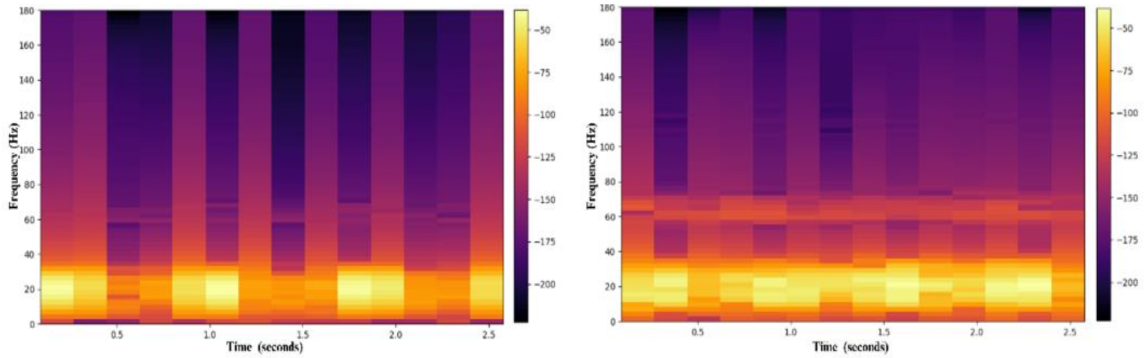


(a)

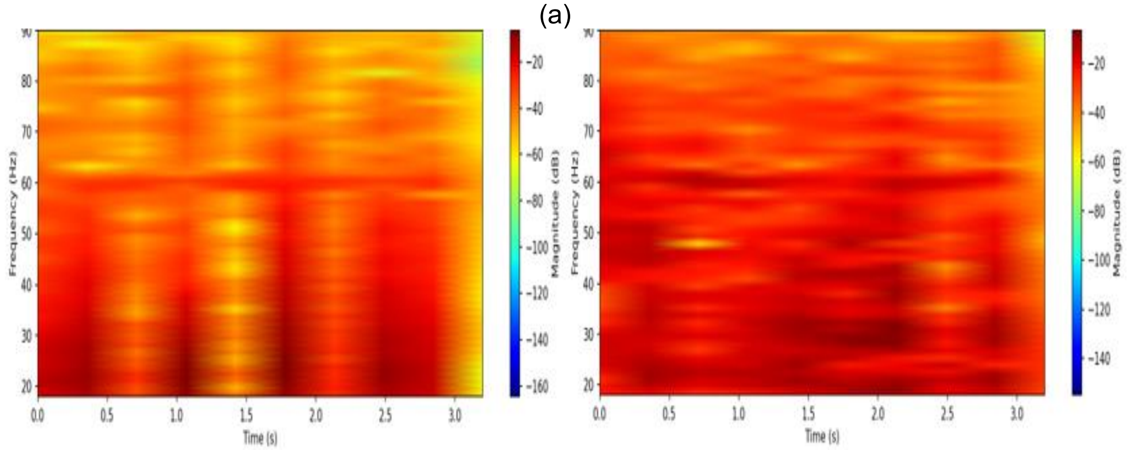
(b)

Fig. 4. Signal processing results: (a) raw vs filtered; (b) filtered vs DWT

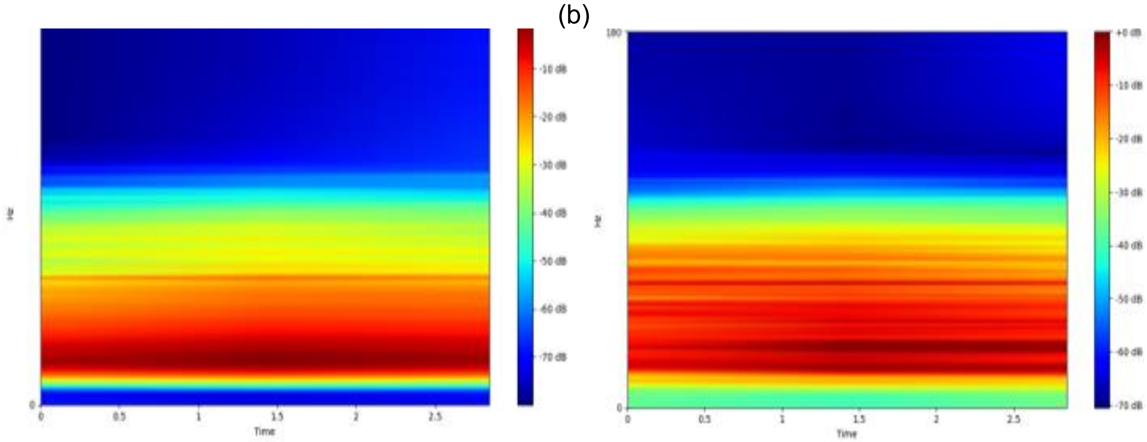
Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730



(a)



(b)



(c)

Fig. 5. Spectrogram comparison of normal (left) and arrhythmia (right) signals: (a) DWT, (b) STFT, (c) Mel-Spectrogram

DWT and STFT. Overall, these observations demonstrate that each feature extraction method captures different aspects of ECG signal characteristics, reinforcing the importance of integrating multiple representations to achieve a more comprehensive analysis.

B. Model Optimization and Performance

To ensure fair and reliable comparison among classification models, hyperparameter optimization was conducted independently for CNN, Random Forest (RF), and Support Vector Machine (SVM). The optimized hyperparameters were selected based on stable convergence behavior and validation performance under patient-wise cross-validation. Table 3 has been expanded to include both the optimized hyperparameters and the optimization strategy used for each model. Hyperparameter tuning was performed using GridSearchCV on the training folds, while the final CNN configuration includes the number of convolutional layers, filter sizes, kernel dimensions, pooling

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1277-1295 e-ISSN: 2656-8632

Table 4. Performance Comparison of Hybrid Methods

Method	Fusion Strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DWT - SVM	Stacking (Epoch 40)	94.82	94.51	94.68	94.59
STFT - RF	Hard Vote (Epoch 40)	94.75	94.42	94.58	94.50
Mel-Spectrogram - CNN	Stacking (Epoch 20)	81.45	81.12	80.89	81.00

Table 5. Statistical Comparison of Classification Models

Comparison	Corrected p-value	Cohen's d	Interpretation
DWT-SVM vs STFT-RF	0.153	0.18	No significant difference
DWT-SVM vs Mel-CNN	< 0.001	2.87	Very large effect
STFT-RF vs Mel-CNN	< 0.001	2.81	Very large effect

operations, dense units, optimizer settings, batch size, and training epochs to ensure reproducibility. The classification performance obtained from different feature–model combinations and ensemble fusion strategies is summarized in Table 4. The results demonstrate that the DWT-SVM model combined with stacking achieved the highest overall classification performance, obtaining 94.82% accuracy and a 94.59% F1-score. This superior performance can be attributed to the capability of DWT to capture localized transient ECG morphology through multi-resolution analysis, while SVM effectively models non-linear decision boundaries within wavelet-based feature spaces. The STFT-RF model also demonstrated highly competitive performance, achieving 94.75% accuracy and a 94.50% F1-score. The stable time-frequency structures generated by STFT, combined with the robustness of ensemble-based Random Forest classification, contributed to strong arrhythmia detection capability and reduced sensitivity to signal noise. In contrast, the Mel-Spectrogram-CNN model produced substantially lower performance, with only 81.45% accuracy. This result suggests that Mel-frequency scaling may suppress high-frequency transient ECG characteristics required for accurate arrhythmia discrimination. In addition, the CNN model exhibited overfitting behaviour after epoch 20, indicating limited representation suitability for ECG morphology analysis. Fig. 6 presents the CNN training and validation performance curves. The validation loss increased after epoch 20 while training accuracy continued improving, indicating overfitting during later training stages.

C. Decision Fusion Effectiveness

The effectiveness of ensemble decision fusion strategies was further evaluated using stacking, hard voting, and soft voting approaches. Experimental results indicate that fusion effectiveness depends strongly on the characteristics of the underlying classifiers and feature representations. Among the evaluated methods, stacking achieved the best overall performance when applied to DWT-SVM classification. The stacking strategy enabled adaptive weighting of heterogeneous classifier outputs through the meta-learner, allowing non-linear relationships among model predictions to be effectively learned. This contributed to improved classification stability and reduced validation error across patient-wise folds. In contrast, the Random Forest classifier achieved optimal performance using hard voting. Since Random Forest itself is inherently ensemble-based, majority voting was sufficient to produce robust classification outputs without requiring additional adaptive fusion mechanisms. These findings demonstrate that ensemble fusion effectiveness is model-dependent rather than universally optimal across all classifiers. The results further confirm the importance of aligning feature representation, classification model, and fusion strategy within a unified ECG classification framework.

D. Statistical Validation

To verify whether the observed performance differences were statistically significant, Friedman testing followed by Wilcoxon signed-rank post-hoc analysis with Bonferroni correction was conducted. The statistical comparison results are summarized in Table 5. The Friedman test indicated statistically significant differences among the evaluated classification frameworks ($p < 0.001$). However, post-hoc analysis revealed no statistically significant difference between DWT-SVM and STFT-RF, suggesting that both approaches provide comparable classification capability for ECG arrhythmia detection. Conversely, both DWT-SVM and STFT-RF significantly outperformed the Mel-Spectrogram-CNN model with

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

very large effect sizes ($d > 2.8$). These findings confirm that DWT and STFT representations are substantially more suitable for ECG arrhythmia analysis than Mel-frequency representations under the evaluated classification framework. Overall, the statistical analysis validates the robustness and reliability of the proposed hybrid time-frequency ECG arrhythmia classification framework.

IV. Discussion

A. Performance Analysis of Feature–Model Compatibility

This study demonstrates that the integration of hybrid time-frequency representations with feature–model compatibility significantly improves ECG arrhythmia classification performance. The experimental results indicate that both DWT-SVM and STFT-RF achieved high and statistically comparable classification performance, while the Mel-Spectrogram-CNN model produced substantially lower accuracy. Specifically, DWT-SVM achieved the highest overall performance with 94.82% accuracy and a 94.59% F1-score, followed closely by STFT-RF with 94.75% accuracy and a 94.50% F1-score. In contrast, Mel-Spectrogram-CNN achieved only 81.45% accuracy and an 81.00% F1-score. These findings highlight the importance of selecting appropriate feature representations and classification models for non-stationary biomedical signal analysis.

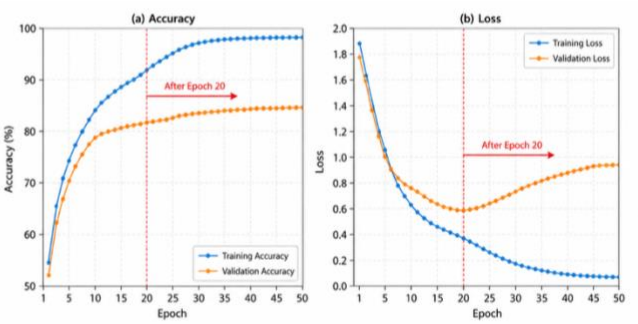


Fig. 6. CNN training and validation performance curves

One of the most important findings of this study is the superior performance achieved by the DWT-SVM framework. The Discrete Wavelet Transform effectively captures localized transient ECG morphology through multi-resolution decomposition, enabling simultaneous analysis in both time and frequency domains. ECG signals inherently contain abrupt waveform variations, particularly within the QRS complex and arrhythmic segments, which are more effectively represented through wavelet coefficients than fixed-resolution spectral methods. When combined with Support Vector Machine using an RBF kernel, the framework becomes capable of modeling complex non-linear decision boundaries within high-dimensional wavelet feature spaces. This combination contributes to robust arrhythmia discrimination and strong generalization capability under patient-wise validation.

The STFT-RF model also demonstrated highly competitive performance, achieving classification results statistically comparable to DWT-SVM. The difference in accuracy between DWT-SVM and STFT-RF was only 0.07%, indicating that both approaches provide similarly effective solutions for ECG arrhythmia classification. The STFT transformation preserves stable rhythmic frequency structures across ECG segments, enabling Random Forest to effectively partition high-dimensional feature distributions through ensemble decision trees. Furthermore, the ensemble nature of Random Forest contributes to improved robustness against signal noise and feature variability, making it highly suitable for practical ECG classification applications.

In contrast, the Mel-Spectrogram-CNN model produced substantially lower performance compared with DWT-SVM and STFT-RF. The performance gap reached 13.37 percentage points in accuracy when compared with DWT-SVM. Although CNN architectures are highly effective for image-based learning tasks, Mel-frequency scaling may not be optimal for ECG morphology analysis. Unlike speech processing, ECG signals rely heavily on subtle transient high-frequency waveform variations associated with cardiac electrical activity. The Mel-scale transformation compresses higher-frequency information while emphasizing lower-frequency components, potentially suppressing diagnostically important ECG characteristics required for accurate arrhythmia discrimination. In addition, the observed overfitting behavior after epoch 20 indicates that the Mel-Spectrogram representation may provide insufficient structural diversity for stable CNN learning in limited biomedical datasets.

B. Effectiveness of Ensemble Decision Fusion

Another important contribution of this study is the investigation of model-dependent ensemble decision fusion strategies. The experimental results demonstrate that fusion effectiveness depends strongly on the characteristics of the underlying classifiers rather than following a universally optimal strategy. Stacking achieved the best overall performance when applied to DWT-SVM, producing 94.82% accuracy and a 94.59% F1-score. The meta-learner was able to adaptively model non-linear relationships among heterogeneous classifier outputs and assign appropriate weights to each prediction source. Consequently, stacking improved classification stability and reduced prediction errors across patient-wise validation folds.

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

Table 6. Comparison with Recent ECG Arrhythmia Classification Studies

Study	Dataset	Method	Validation Strategy	Accuracy (%)	Main Limitation
[39]	MIT-BIH Arrhythmia Database	Hybrid Machine Learning–Deep Learning Framework	Segment-wise	99.26	Potential patient-level data leakage due to segment-wise evaluation
[14]	MIT-BIH + BIDMC	Three-Channel Spectrogram Deep Learning with Feature Fusion (STFT + HOG + LBP + CNN)	Segment-wise	98.90	Relies primarily on spectrogram-based representations
[22]	MIT-BIH Arrhythmia Database	Deep Learning ECG Classification (FADLEC)	Subject-independent	98.21	No ensemble decision fusion mechanism
[25]	MIT-BIH Arrhythmia Database	Hybrid CNN-BLSTM Architecture	Segment-wise	98.74	Limited investigation of feature–model compatibility
[18]	ECG Arrhythmia Dataset	CNN–LSTM–GRU Hybrid Deep Learning Model	Cross-validation	98.60	High computational complexity and deployment cost
Proposed Method	**MIT-BIH Arrhythmia Database**	**Hybrid Time-Frequency + Feature–Model Compatibility + Ensemble Decision Fusion**	**Patient-wise Cross-Validation**	**94.82**	**Binary classification and single-dataset evaluation**

Note: The reported accuracy values should be interpreted cautiously because the compared studies differ in dataset composition, class definitions, ECG segmentation, model architecture, and validation strategy.

In contrast, hard voting proved more suitable for the Random Forest classifier. Since Random Forest itself operates as an ensemble of decision trees, majority-voting aggregation was sufficient to produce robust and stable classification outputs without requiring an additional adaptive fusion layer. These findings indicate that ensemble fusion should not be applied uniformly across different models. Instead, fusion strategies should be optimized according to classifier behavior and feature representation characteristics to maximize classification performance.

C. Comparison with Previous Studies

Table 6 presents a comparison between the proposed framework and several recent ECG arrhythmia classification studies. The comparison highlights differences in datasets, methodologies, validation protocols, classification performance, and study limitations. Table 6 shows that the proposed framework shares several characteristics with previous studies, particularly in the use of automated ECG arrhythmia classification and hybrid or time-frequency feature representations. Similar to the frameworks reported in [39] and [14], the present study combined complementary feature-processing and classification components. However, both studies employed segment-wise validation, whereas the proposed framework used patient-wise cross-validation, ensuring that ECG segments originating from the same MIT-BIH record were not distributed across both the training and testing sets. This stricter separation reduces the risk of patient-level data leakage and provides a more realistic assessment of generalization to unseen records, although it may also produce lower nominal performance than segment-wise evaluation. The proposed study is also similar to FADLEC [22] in adopting a subject-independent evaluation perspective. Nevertheless, unlike [22], the present framework explicitly investigates feature–model compatibility and compares stacking, hard voting, and soft voting as model-dependent decision fusion strategies. Compared with the CNN-BLSTM architecture in [25] and the CNN-LSTM-GRU model in [18], the proposed framework employs heterogeneous classifiers paired with structurally different feature representations, namely DWT-SVM, STFT-RF, and Mel-Spectrogram-CNN, rather than relying exclusively on a single end-to-end deep learning architecture. Although the proposed method achieved a lower

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

reported accuracy of 94.82% compared with the 98.21–99.26% accuracies reported in the previous studies, these values should not be interpreted as direct evidence of methodological inferiority or superiority because the studies differ in record selection, class definitions, segmentation procedures, model complexity, and validation protocols. The main distinction of the proposed framework is therefore not the achievement of the highest raw accuracy, but the demonstration that representation-specific classifier selection and model-dependent decision fusion can maintain strong classification performance under leakage-resistant patient-wise validation. Nevertheless, the proposed method remains limited by its binary classification setting, the use of selected MIT-BIH records, and the absence of external dataset validation.

D. Clinical Implications

From a clinical perspective, the proposed framework demonstrates strong potential for integration into computer-aided cardiac diagnostic systems and real-time ECG monitoring applications. The high classification performance achieved by DWT-SVM and STFT-RF suggests that hybrid time-frequency analysis can effectively support automated arrhythmia screening while reducing clinician workload and inter-observer variability. Furthermore, the relatively lightweight computational requirements of SVM and Random Forest classifiers compared with large deep learning architectures may facilitate deployment in wearable monitoring devices and edge-based healthcare systems. Previous work involving an AD8232-based ECG acquisition system has also demonstrated the feasibility of implementing low-cost ECG monitoring in healthcare-oriented applications [40]. Such acquisition platforms may be integrated with automated classification frameworks to support portable and continuous cardiac monitoring. Consequently, the proposed framework may contribute to improving early diagnosis, reducing diagnostic delays, and supporting clinical decision-making in cardiovascular care.

E. Limitations and Future Work

Despite these promising findings, several limitations should be acknowledged. First, this study utilized a relatively limited subset of records from the MIT-BIH Arrhythmia Database, which may affect the generalizability of the proposed framework across broader patient populations. Second, the study focused exclusively on binary classification between normal and arrhythmic ECG recordings. Although this design reduces class imbalance and facilitates robust abnormality detection, it does not reflect the complexity of real-world multi-class arrhythmia diagnosis. Third, only single-lead ECG signals were evaluated, whereas multi-lead ECG recordings may provide additional spatial cardiac information that could further improve classification performance. Fourth, no external validation dataset was employed, limiting the assessment of model generalizability across different populations and acquisition settings. Fifth, the selection of specific MIT-BIH records may introduce class-selection bias that could influence performance estimation. Sixth, although the proposed framework achieved strong classification performance, explainability remains limited because the classification process primarily relies on data-driven learning mechanisms. Finally, computational complexity, inference latency, and real-time deployment feasibility were not explicitly evaluated in this study. Future work will focus on extending the proposed framework using larger and more diverse multi-lead ECG datasets, investigating multi-class arrhythmia classification scenarios, incorporating explainable artificial intelligence (XAI) techniques to improve clinical interpretability, conducting external validation studies, and optimizing lightweight architectures for real-time edge deployment in portable cardiac monitoring systems.

V. Conclusion

This study developed a hybrid time-frequency ECG arrhythmia classification framework that integrated feature–model compatibility with ensemble decision fusion. Three complementary representations, namely Discrete Wavelet Transform, Short-Time Fourier Transform, and Mel-Spectrogram, were paired with Support Vector Machine, Random Forest, and Convolutional Neural Network classifiers, respectively, and evaluated using patient-wise cross-validation on selected records from the MIT-BIH Arrhythmia Database. The DWT-SVM configuration with stacking achieved the highest overall performance, with an accuracy of 94.82% and an F1-score of 94.59%, while the STFT-RF configuration with hard voting produced statistically comparable results, achieving an accuracy of 94.75% and an F1-score of 94.50%. In contrast, the Mel-Spectrogram-CNN configuration achieved substantially lower performance, indicating that Mel-frequency scaling was less suitable for preserving diagnostically important transient ECG morphology under the evaluated conditions. The statistical analysis confirmed significant overall differences among the evaluated configurations, although no significant difference was observed between DWT-SVM and STFT-RF. These findings demonstrate that classification performance depends not only on the selected feature representation or classifier, but also on their structural compatibility and the decision fusion strategy applied. Nevertheless, the study was limited to binary classification, selected single-lead ECG records, and a

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

single database. Future research should therefore investigate multi-class and multi-lead ECG classification, external validation using independent datasets, explainable artificial intelligence methods, and computationally efficient implementations for real-time cardiac monitoring.

Acknowledgment

The authors would like to express their sincere gratitude to the creators and administrators of the MIT-BIH Arrhythmia Database for providing open access to the high-quality ECG data that supported this study, as well as to the Research and Community Service Institute (LPPM) at Syiah Kuala University. The authors would also like to thank the Department of Electrical and Computer Engineering for its technical assistance, and express their appreciation to all colleagues and contributors for their support, guidance, and valuable constructive feedback throughout this research.

Funding

This work was supported by Kemdiktisaintek under Grant No. 58/UN11.L1/PG.01.03/DPPM/2025, with additional support from the Ministry of Research, Technology, and Higher Education, Republic of Indonesia, and LPPM Universitas Syiah Kuala.

Data Availability

The dataset used in this study is publicly available from the MIT-BIH Arrhythmia Database, which can be accessed through PhysioNet (https://physionet.org/). No new data were generated in this study. All data used are open-access and widely utilized for research purposes.

Author Contribution

J. Prayoga: Conceptualization, methodology, software, data curation, formal analysis, and writing original draft preparation. Melinda Melinda: Supervision, validation, writing review and editing, and project administration. Teuku Yuliar Arif: Methodology support, technical validation, and review of the manuscript. Herlina Dimiati: Clinical validation, data interpretation, and domain expertise in ECG analysis.

Declarations

Ethics Statement

This study utilized the publicly available MIT-BIH Arrhythmia Database obtained from PhysioNet. All ECG recordings were fully anonymized and publicly accessible for research purposes. Therefore, no additional ethical approval or informed consent was required.

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] V. R. Netala, S. K. Teertam, H. Li, and Z. Zhang, "A Comprehensive Review of Cardiovascular Disease Management: Cardiac Biomarkers, Imaging Modalities, Pharmacotherapy, Surgical Interventions, and Herbal Remedies," Sep. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/cells13171471.

[2] H. Velandia, A. Pardo, M. I. Vera, and M. Vera, "Systematic Review of Artificial Intelligence and Electrocardiography for Cardiovascular Disease Diagnosis," *Bioengineering*, vol. 12, no. 11, 2025, doi: 10.3390/bioengineering12111248.

[3] N. Katal, S. Gupta, P. Verma, and B. Sharma, "Deep-Learning-Based Arrhythmia Detection Using ECG Signals: A Comparative Study and Performance Evaluation," *Diagnostics*, vol. 13, no. 24, 2023, doi: 10.3390/diagnostics13243605.

[4] M. R. Khan, Z. M. Haider, J. Hussain, F. H. Malik, I. Talib, and S. Abdullah, "Comprehensive Analysis of Cardiovascular Diseases: Symptoms, Diagnosis, and AI Innovations," *Bioengineering*, vol. 11, no. 12, 2024, doi: 10.3390/bioengineering11121239.

[5] A. J. Prakash and M. Atef, "A lightweight deep learning approach for patient-specific electrocardiogram beat classification using local and long-term dependencies," *Eng. Appl. Artif. Intell.*, vol. 152, p. 110754, 2025, doi: https://doi.org/10.1016/j.engappai.2025.110754.

[6] E. H. Houssein, H. Mostafa, A. A. Ali, H. Shaban, and M. S. Hassanein, "Advances in artificial intelligence techniques for diagnosis of cardiac diseases," *Discover Artificial Intelligence*, vol. 6, no. 1, p. 424, 2026, doi: 10.1007/s44163-026-01323-y.

[7] Z. Wu and C. Guo, "Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management," *Biomed. Eng. Online*, vol. 24, no. 1, p. 23, 2025, doi: 10.1186/s12938-025-01349-w.

[8] M. Melinda, M. Raja, J. Junidar, R. Miftahujjannah, S. Rusdiana, and M. Irhamsyah, "Performance Comparison Analysis of Random Forest, Support Vector Machine, and AdaBoost in Arrhythmia Classification," *Journal of Image and Graphics (United Kingdom)*, vol. 13, no. 5, pp. 540–548, 2025, doi: 10.18178/JOIG.13.5.540-548.

[9] L. C. Nechita *et al.*, "AI-Enhanced ECG Applications in Cardiology: Comprehensive Insights from the Current Literature with a Focus

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

on COVID-19 and Multiple Cardiovascular Conditions," *Diagnostics*, vol. 14, no. 17, 2024, doi: 10.3390/diagnostics14171839.

[10] D. Al-Shammary, M. Radhi, A. H. AlSaeedi, A. M. Mahdi, A. Ibaida, and K. Ahmed, "Efficient ECG classification based on the probabilistic Kullback-Leibler divergence," *Inform. Med. Unlocked*, vol. 47, p. 101510, 2024, doi: https://doi.org/10.1016/j.imu.2024.101510.

[11] Y. Liu *et al.*, "Y-Net-ECG: A Multi-Lead informed and interpretable architecture for ECG segmentation across diverse rhythms," *Expert Syst. Appl.*, vol. 283, p. 127955, 2025, doi: https://doi.org/10.1016/j.eswa.2025.127955.

[12] M. Fira, L. Goraș, L. Fira, R. F. Popa, and H.-N. Costin, "Arrhythmia Classification with Single-Channel Features Extracted from 'A Large-Scale 12-Lead ECG Database for Arrhythmia Study,'" *Sensors*, vol. 25, no. 18, 2025, doi: 10.3390/s25185621.

[13] E. Oporto, D. Mauricio, N. Maculan, and G. Uribe, "Challenges in the Classification of Cardiac Arrhythmias and Ischemia Using End-to-End Deep Learning and the Electrocardiogram: A Systematic Review," *Diagnostics*, vol. 16, no. 1, 2026, doi: 10.3390/diagnostics16010161.

[14] A. Eleyan, F. Bayram, and G. Eleyan, "Spectrogram-Based Arrhythmia Classification Using Three-Channel Deep Learning Model with Feature Fusion," *Applied Sciences*, vol. 14, no. 21, 2024, doi: 10.3390/app14219936.

[15] Y. Yunidar, M. Melinda, Albahri, H. Aulia, H. Dimiati, and N. Basir, "Precise Electrocardiogram Signal Analysis Using ResNet, DenseNet, and XceptionNet Models in Autistic Children," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 4, pp. 1303–1319, Oct. 2025, doi: 10.35882/jeeemi.v7i4.1044.

[16] F. M. Heir, H. Najafzadeh, and S. Erfani, "A hybrid CNN and reinforcement learning framework for speaker identification using Mel-Spectrogram and continuous wavelet transform features," *Sci. Rep.*, vol. 16, no. 1, Dec. 2026, doi: 10.1038/s41598-026-35858-y.

[17] J. Rahul and L. D. Sharma, "Advancements in AI for cardiac arrhythmia detection: A comprehensive overview," *Comput. Sci. Rev.*, vol. 56, p. 100719, 2025, doi: https://doi.org/10.1016/j.cosrev.2024.100719.

[18] A. A. Mohammed *et al.*, "Enhanced arrhythmia diagnosis using a hybrid deep learning model: A CNN-LSTM-GRU approach on ECG data," *Biosens. Bioelectron. X*, vol. 30, p. 100767, 2026, doi: https://doi.org/10.1016/j.biosx.2026.100767.

[19] M. Qi, H. Shao, N. Shi, G. Wang, and Y. Lv, "Arrhythmia classification detection based on multiple electrocardiograms databases," *PLoS One*, vol. 18, no. 9, pp. e0290995-, Sep. 2023, [Online]. Available: https://doi.org/10.1371/journal.pone.0290995

[20] K. Gunasekaran, V. D. Ambeth Kumar, and M. J. A., "Artifact removal from ECG signals using online recursive independent component analysis," *Journal of Computational Mathematics and Data Science*, vol. 13, p. 100102, 2024, doi: https://doi.org/10.1016/j.jcmds.2024.100102.

[21] M. F. Safdar, R. M. Nowak, and P. Pałka, "Pre-Processing techniques and artificial intelligence algorithms for electrocardiogram (ECG) signals analysis: A comprehensive review," *Comput. Biol. Med.*, vol. 170, p. 107908, Mar. 2024, doi: 10.1016/J.COMPBIOMED.2023.107908.

[22] S. Lamba, S. Kumar, and M. Diwakar, "FADLEC: feature extraction and arrhythmia classification using deep learning from electrocardiograph signals," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 82, 2025, doi: 10.1007/s44163-025-00290-0.

[23] C. Qiu, H. Li, C. Qi, and B. Li, "Enhancing ECG classification with continuous wavelet transform and multi-branch transformer," *Heliyon*, vol. 10, no. 5, Mar. 2024, doi: 10.1016/j.heliyon.2024.e26147.

[24] H. Bechinia, D. Benmerzoug, and N. Khlifa, "Approach Based Lightweight Custom Convolutional Neural Network and Fine-Tuned MobileNet-V2 for ECG Arrhythmia Signals Classification," *IEEE Access*, vol. 12, pp. 40827–40841, 2024, doi: 10.1109/ACCESS.2024.3378730.

[25] Y. Ye, K. Chipusu, M. A. Ashraf, B. Ding, Y. Huang, and J. Huang, "Hybrid CNN-BLSTM architecture for classification and detection of arrhythmia in ECG signals," *Sci. Rep.*, vol. 15, no. 1, p. 34510, 2025, doi: 10.1038/s41598-025-17671-1.

[26] M. Irhamsyah, M. Melinda, J. Alifa, J. Prayoga, and Y. H. P. Iskandar, "ECG Atrial Fibrillation Signal Classification Method Based on Discrete Wavelet Transform (DWT) and DenseNet-121," *2024 10th International Conference on Smart Computing and Communication, ICSCC 2024*, pp. 619–624, 2024, doi: 10.1109/ICSCC62041.2024.10690792.

[27] S. Dhyani, A. Kumar, and S. Choudhury, "Analysis of ECG-based arrhythmia detection system using machine learning," *MethodsX*, vol. 10, p. 102195, Jan. 2023, doi: 10.1016/J.MEX.2023.102195.

[28] J. Wang, J. Zang, S. Yao, Z. Zhang, and C. Xue, "Multiclassification for heart sound signals under multiple networks and multi-view feature," *Measurement*, vol. 225, p. 114022, Feb. 2024, doi: 10.1016/J.MEASUREMENT.2023.114022.

[29] M. Fakhry and A. Gallardo-Antolín, "A hybrid neural network with multistage feature fusion for

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

detecting heart failure and murmurs from time–frequency representations of phonocardiograms," *Systems and Soft Computing*, vol. 7, p. 200391, 2025, doi: https://doi.org/10.1016/j.sasc.2025.200391.

[30] J. Prayoga, M. Melinda, T. Y. Arif, and H. Dimiati, "Comparison of Efficientnet B3 and Inception V3 Architecture Performance in Heart Disease Arrhythmia Image Classification," *International Conference on Computer and Communication Engineering, ICCCE*, no. 2025, pp. 357–362, 2025, doi: 10.1109/ICCCE66530.2025.11474189.

[31] M. Dehghanpour, M. Fateh, Z. Mohammadpoory, and S. Ferdowsi, "KhayyamNet: A Parallel Multiscale Feature Fusion Framework for Accurate Diagnosis of Multiple Sclerosis and Myelitis," *Mach. Learn. Knowl. Extr.*, vol. 8, no. 3, 2026, doi: 10.3390/make8030062.

[32] Z. Xu and Z. Yu, "Maximizing Single-Feature Separability for Improving Transfer Learning in Motor Imagery EEG Decoding," *Brain Sci.*, vol. 16, no. 2, 2026, doi: 10.3390/brainsci16020230.

[33] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, pp. 6086-, Mar. 2024, doi: 10.1038/s41598-024-56706-x.

[34] D. Wilimitis and C. G. Walsh, "Practical Considerations and Applied Examples of Cross-Validation for Model Development and Evaluation in Health Care: Tutorial," *JMIR AI*, vol. 2, no. 1, 2023, doi: 10.2196/49023.

[35] A. Alwakeel, M. Alwakeel, M. Hijji, T. J. Saleem, and S. R. Zahra, "Performance Evaluation of Different Decision Fusion Approaches for Image Classification," *Applied Sciences*, vol. 13, no. 2, 2023, doi: 10.3390/app13021168.

[36] F. Zhao *et al.*, "Multi-classifier fusion based on belief-value for the diagnosis of autism spectrum disorder," *Front. Hum. Neurosci.*, vol. Volume 17-2023, 2023, doi: 10.3389/fnhum.2023.1257987.

[37] S. U. Haq *et al.*, "Reseek-Arrhythmia: Empirical Evaluation of ResNet Architecture for Detection of Arrhythmia," *Diagnostics*, vol. 13, no. 18, 2023, doi: 10.3390/diagnostics13182867.

[38] F. Fiel Peres, "Effect sizes for nonparametric tests," *Biochem. Med. (Zagreb).*, vol. 36, no. 1, Feb. 2025, doi: 10.11613/BM.2026.010101.

[39] F. Zabihi, F. Safara, and B. Ahadzadeh, "An electrocardiogram signal classification using a hybrid machine learning and deep learning approach," *Healthcare Analytics*, vol. 6, p. 100366, 2024, doi: https://doi.org/10.1016/j.health.2024.100366.

[40] C. N. Nurbadriani, M. Melinda, Y. Yunidar, and F. Arnia, "Electrocardiogram Detection System of Autistic Children Based on AD8232 for Healthcare," *Proceeding - 2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering: Sustainable Development for Smart Innovation System, COSITE 2023*, pp. 126–131, 2023, doi: 10.1109/COSITE60233.2023.10250117.

Author Biography

J. Prayoga, born in Payakumbuh on June 10, 1990, is a permanent lecturer at the Faculty of Engineering and Computer Science, Dharmawangsa University, Medan. He completed his bachelor's and master's degrees in Information Systems at the University of Putra Indonesia, "YPTK" Padang. His areas of expertise focus on Information Systems, Computer Science, and Computer Applications, encompassing information technology design, software development, and the implementation of practical computing solutions for organizations and society. To date, he has produced 31 scientific publications with over 1,587 reads and 36 citations on ResearchGate, and has authored the academic book Research Methods as a contribution to the development of scientific literature in Indonesia. He can be contacted at email: yoga@dharmawangsa.ac.id

Melinda was born in Bireuen, Aceh, on June 10, 1979. She received a B.Eng degree from the Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, in 2002. She completed her master's degree at the Faculty of Electrical Department, University of Southampton, United Kingdom, with a concentration in the field study of Radio Frequency Communication Systems in 2009. She has already completed her Doctoral degree at the Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, in February 2018. She has been with the Department of Electrical Engineering, Faculty of Engineering, Universitas Syiah Kuala since 2002. She is also a member of IEEE. Her research interests include multimedia signal processing and fluctuation processing. She can be contacted at email: melinda@usk.ac.id

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730

Teuku Yuliar Arif received the B.E. degree from Sepuluh Nopember Institute of Technology (ITS), Surabaya, Indonesia, in 1998, and the M.Com. and Ph.D. degrees from the University of Indonesia (UI), Indonesia, in 2004 and 2013, respectively. In 1999, he joined Syiah Kuala University, Banda Aceh, Indonesia, as a lecturer at the Electrical Engineering Department. His research interests include wireless and mobile networks, vehicular ad hoc networks (VANET), network security, network performance analysis, and mesh networks for Internet of Things (IoT).

Herlina Dimiati was born on March 5, 1964, in Banda Aceh. She graduated as a doctor in 1992 from the Faculty of Medicine at USK and continued her specialization in Pediatrics at the Faculty of Medicine, University of North Sumatra, completing her studies in 2002. In 2006, she continued his pediatric cardiology education at Cipto Mangunkusumo Hospital in Jakarta and graduated in 2008. Continuing doctoral studies at the Faculty of Medicine, Gadjah Mada University in 2011 and completed in 2016. Currently, she has held the title of Professor in child health at the Faculty of Medicine, USK since 2021. She can be reached at herlinadimiati@usk.ac.id.

Manuscript received 6 Mei 2026; Revised 25 June 2026; Accepted 20 July 2026; Available online 30 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1730