

Machine learning–enabled ECG arrhythmia classification: a systematic and educational study from signal processing to decision support

Received: 17 March 2026

Accepted: 3 June 2026

Published online: 10 June 2026

Cite this article as: Melek N. Machine learning–enabled ECG arrhythmia classification: a systematic and educational study from signal processing to decision support. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-56828-4>

Negin Melek

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Machine Learning–Enabled ECG Arrhythmia Classification: A Systematic and Educational Study from Signal Processing to Decision Support

Negin Melek^{1*}

^{1*}Faculty of Engineering and Natural Sciences, Gümüşhane University, Gümüşhane, 29000, Turkey.

Corresponding author(s). E-mail(s): negin.melek@gumushane.edu.tr;

Abstract

Electrocardiogram (ECG) signals play a critical role in the early detection of cardiac arrhythmias, which remain a major cause of morbidity and mortality worldwide. While deep learning approaches have achieved high classification accuracy, their increasing complexity often limits interpretability and practical applicability. This study presents a systematic and interpretable framework for multi-class ECG arrhythmia classification, examining the effects of signal processing, feature extraction, feature selection, and evaluation strategies on classification performance. Experiments were conducted on the MIT-BIH Arrhythmia Database using two feature representations: (i) morphological and temporal features, and (ii) a compact wavelet-based representation. Sequential Forward Feature Selection (SFFS) revealed that classification performance saturates at approximately 15 features, indicating that most discriminative information is captured within a compact subset. Using this feature space, the support vector machine (SVM) achieved the best overall performance, reaching 98.54% accuracy with stable results across different configurations. The wavelet-based representation further improved performance balance, yielding lower Golden Distance (GD) values (as low as 0.0249), indicating more consistent behavior across evaluation metrics. Overall, the results demonstrate that carefully designed feature extraction and selection enable classical machine learning methods, particularly SVM, to achieve high and reliable performance, providing an interpretable alternative to more complex black-box models in ECG arrhythmia classification.

Keywords: ECG signal processing, feature engineering, classical machine learning, arrhythmia detection, model evaluation

1 Introduction

The human heart is a vital organ responsible for continuous blood circulation throughout the body via the cardiovascular system, ensuring the delivery of oxygen and nutrients to tissues and the removal of metabolic waste products [1–3]. This

process is supported by coordinated cardiac structures and regulated blood flow, enabling efficient and stable circulation [4–6].

The coordinated function of the heart is governed by electrical impulses that control rhythmic contraction and relaxation [7]. Cardiac rhythm is initiated by the sinoatrial (SA) node, which acts as the natural pacemaker, and propagates through

the cardiac conduction system to ensure synchronized contraction of the heart muscle [8, 9]. These electrical activities, including depolarization and repolarization processes governed by action potentials, are essential for maintaining stable cardiac function, and abnormalities in these processes may lead to arrhythmias [10–13].

The electrocardiogram (ECG) is a non-invasive technique widely adopted to record the electrical activity of the heart and evaluate cardiac rhythm [14–16]. Variations in ECG waveforms, including the P wave, QRS complex, and T wave, provide essential information for identifying arrhythmias and other cardiac disorders.

Cardiac signal analysis is a fundamental component of modern cardiology and biomedical engineering, enabling the assessment of heart rate variability (HRV), QT interval monitoring, arrhythmia detection, and diagnosis of various cardiac conditions [17–19]. With the advancement of wearable devices and continuous monitoring systems, ECG analysis is increasingly used for real-time health assessment and early diagnosis [20–24].

In ECG signal analysis, morphological characteristics and temporal patterns, such as the QRS complex and R–R intervals, are extensively used to extract meaningful features for classification tasks [25, 26]. These features play a crucial role in distinguishing normal and abnormal cardiac rhythms and support the development of automated diagnostic systems.

Several publicly available ECG databases are widely used for developing and evaluating classification algorithms. Among them, the MIT-BIH Arrhythmia Database is one of the most commonly used benchmark datasets for arrhythmia detection and classification [27]. Other datasets, such as the Apnea-ECG Database and the Kaggle ECG dataset, further support the development of machine learning-based approaches [28, 29].

Despite significant progress in ECG-based arrhythmia classification, several challenges remain. Many existing studies focus primarily on maximizing classification accuracy using complex models, often overlooking the interpretability of feature representations and the contribution of individual processing stages. In addition, the relative importance of feature selection versus classifier choice is not systematically analyzed, limiting the understanding of model behavior in practical applications and

reducing the transparency of decision-making processes [30, 31].

Motivated by these challenges, this study adopts a structured and interpretable approach to ECG classification. The aim of this study is to develop a systematic and interpretable framework for multi-class ECG arrhythmia classification, focusing on the interaction between signal processing, feature extraction, feature selection, and evaluation strategies. The main contributions of this study are summarized as follows:

- A systematic and educational framework for ECG arrhythmia classification is proposed, covering signal preprocessing, feature extraction, feature selection, and evaluation.
- Morphological, temporal, and time–frequency (wavelet) features are analyzed to assess their impact on classification performance.
- The effectiveness of wrapper-based Sequential Forward Feature Selection (SFFS) in reducing redundant features and highlighting feature dominance over classifier choice is demonstrated.
- A multidimensional evaluation framework based on the Golden Distance (GD) approach alongside conventional metrics is introduced, enabling robust performance assessment in imbalanced multi-class problems [32, 33].
- An instructive guide for future researchers is provided, detailing preprocessing, feature extraction, feature selection, and classifier evaluation steps.

Unlike conventional studies that primarily focus on maximizing classification accuracy, this work systematically analyzes the interaction between feature extraction, feature selection, and classifier behavior, demonstrating that feature representation and selection have a more significant impact on performance than classifier choice.

1.1 Related Work

Recent research on ECG-based arrhythmia classification has been dominated by deep learning (DL) approaches, particularly convolutional neural networks (CNN) and hybrid architectures [31, 34, 35]. These models can achieve high classification performance by learning feature representations directly from raw ECG signals. In addition, recent studies have extended machine learning and deep

learning methods to broader clinical and diagnostic problems in cardiovascular signal analysis [36]. However, despite these advances, the limited interpretability of such models remains an important concern, especially in clinical settings where transparent and explainable decisions are required [30, 37]. This limitation restricts their practical adoption in scenarios where model transparency and clinical validation are essential.

As a result, classical machine learning methods, such as support vector machines (SVM) and k-nearest neighbors (k-NN), are still widely used [38, 39]. These methods rely on handcrafted features derived from ECG signals, including morphological, temporal, and frequency-domain characteristics. Since these features are closely related to physiological indicators such as QRS duration and RR interval dynamics, they provide a more interpretable basis for clinical decision-making. Nevertheless, their performance is highly dependent on the quality and relevance of the extracted features, making feature design a critical and non-trivial task.

In addition, recent studies have demonstrated the effectiveness of statistical and temporal feature-based approaches combined with feature selection techniques for arrhythmia classification. For example, Kaya [40] proposed a machine learning-based framework for the detection of Bundle Branch Block (BBB) using higher-order statistical and temporal features, where feature selection methods such as forward selection, backward elimination, and genetic algorithms were employed. Comparative evaluations using k-NN, SVM, and neural networks showed that carefully selected feature subsets can achieve very high classification accuracy on the MIT-BIH dataset. These findings further support the importance of feature quality and selection strategies in improving classification performance, while also highlighting the dependency of results on the chosen feature subset.

Despite this advantage, handcrafted feature-based approaches also face important challenges. One of the most common issues is the curse of dimensionality, where large feature sets introduce redundancy and may reduce classification performance while increasing computational cost [41]. To address this problem, many studies combine signal processing techniques, such as the discrete

wavelet transform (DWT), with feature selection methods to obtain more compact and informative feature sets [42, 43]. Recent work has also explored advanced signal enhancement and denoising strategies to improve ECG processing pipelines [44]. However, determining an optimal balance between feature richness and model simplicity remains an open challenge.

Among these methods, wrapper-based feature selection approaches, particularly Sequential Forward Floating Selection (SFFS), have been widely used due to their ability to dynamically include and exclude features during the selection process [41]. This flexibility helps overcome the limitations of simple greedy methods and enables the identification of more representative feature subsets. Nevertheless, these methods may still involve increased computational cost, especially when applied to large feature spaces.

Another important issue in ECG classification studies is model evaluation. Since class imbalance is common in ECG datasets, relying on a single metric may lead to incomplete or misleading conclusions. For this reason, recent studies have increasingly emphasized the use of more comprehensive evaluation measures, including balanced accuracy, precision, recall, F1-score, and Cohen's kappa [25]. In this context, the Golden Distance (GD) metric has been proposed as a comprehensive performance measure that combines multiple evaluation criteria into a single index [25]. This allows a more balanced comparison of classification models by considering different aspects of performance simultaneously, although its adoption in the literature is still relatively limited.

Finally, although many existing studies report strong classification results, fewer studies provide a systematic explanation of how preprocessing, feature extraction, feature selection, and evaluation stages interact within a unified framework. This gap limits the interpretability and reproducibility of existing approaches and highlights the need for more structured and educational frameworks in ECG-based classification studies.

2 Materials and Methods

The proposed methodology is structured as a four-stage pipeline, as illustrated in Figure 1. In the first stage, ECG signals are acquired from the MIT-BIH Arrhythmia Database and imported

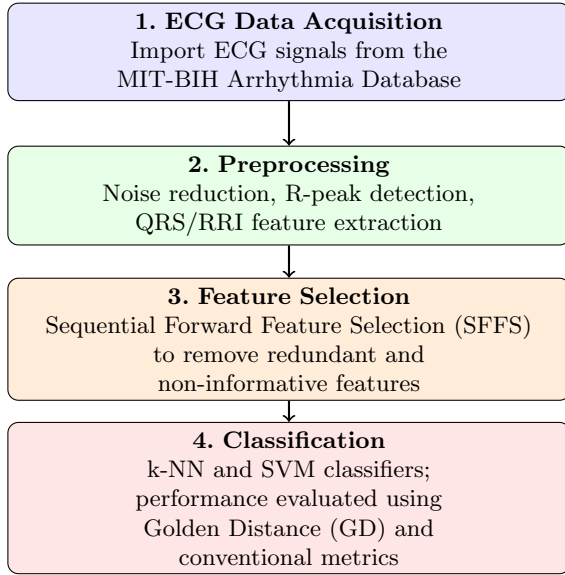


Fig. 1: Block diagram of the proposed ECG arrhythmia classification framework, showing the sequential stages from data acquisition to classification.

into the analysis environment. In the second stage, preprocessing is performed to enhance signal quality, including noise reduction and extraction of QRS complexes and R-R intervals. In the third stage, feature selection is conducted using the Sequential Forward Feature Selection (SFFS) algorithm, which iteratively selects the most informative and non-redundant features based on validation performance. This step reduces feature dimensionality while preserving discriminative information. In the final stage, the selected feature vectors are classified using k-nearest neighbor (k-NN) and support vector machine (SVM) classifiers. Model performance is evaluated using both conventional classification metrics and the Golden Distance (GD) metric, which provides a multidimensional assessment of performance.

2.1 MIT-BIH Arrhythmia Database

The MIT-BIH Arrhythmia Database, hosted on PhysioNet and maintained by MIT's Computational Physiology Laboratory, is a widely used benchmark dataset for studies on cardiac arrhythmias [45]. The database contains two synchronously recorded ECG channels, Modified Lead II (MLII) and V1. In this study, only the

MLII channel was utilized, as it provides more distinct QRS complexes and is commonly used as a reference in arrhythmia classification, ensuring signal quality and comparability with previous studies.

The database consists of 48 recordings, each approximately 30 minutes in duration, provided as raw `.dat` files. In addition to the signal data, auxiliary files are available for each recording: `.atr` files contain heartbeat annotations, `.hea` files provide header information, and `.xws` files include waveform-related data (available for some recordings). This structure enables access to both raw ECG signals and corresponding annotation data for detailed analysis.

The `.dat` files in the MIT-BIH Arrhythmia Database serve as the primary data source in this study, containing two-channel ECG recordings of approximately 30 minutes in duration. The signals are sampled at 360 Hz, resulting in 648,000 samples per channel, as given in Equation 1:

$$30 \times 60 \times 360 = 648,000 \quad (1)$$

Accordingly, each recording forms a $648,000 \times 2$ data matrix due to the simultaneous acquisition of two channels. Auxiliary files, including `.atr`, `.hea`, and `.xws`, are used to provide annotation and metadata information required for further analysis.

The `.atr` files in the MIT-BIH Arrhythmia Database contain annotations for each QRS complex, manually labeled by expert cardiologists to indicate the corresponding arrhythmia type. These annotations serve as reliable ground truth data for supervised ECG analysis and classification tasks. Table 1 [27] summarizes the annotation symbols and their corresponding heartbeat classes.

The `.hea` files in the MIT-BIH Arrhythmia Database provide metadata for each recording, including signal length, sampling parameters, and channel information. In this study, the primary data sources are the `.dat` and `.atr` files, while `.hea` files are used for data verification and organization. R-wave positions and arrhythmia class labels are directly obtained from the `.atr` annotations and utilized throughout the pipeline for segmentation, feature extraction, and performance evaluation.

The MIT-BIH Arrhythmia Database serves as a standard benchmark for the development

Table 1: Heartbeat Symbols and Groups

Group	Description	Sub symbols
N (Normal)	Normal or non-ectopic beats	N, L (LBBB), R (RBBB), j (nodal escape), e (atrial escape)
S (Supraventricular)	Supraventricular premature beats	A (aberrant atrial), a (atrial premature), S (supraventricular premature), J (nodal premature)
V (Ventricular)	Ventricular premature beats	V (ventricular escape), E (premature ventricular contraction)
F (Fusion)	Normal and ectopic fusion beats	F (fusion ventricular-normal)
U (Unidentified)	Uncertain or unclassifiable	Q (unclassifiable), / (paced), p (paced), U (unknown), f (fusion paced-normal)

LBBB and RBBB denote Left Bundle Branch Block and Right Bundle Branch Block, respectively.

and evaluation of ECG-based arrhythmia classification methods. In this study, heartbeat annotations are grouped into five classes following common practice in the literature, enabling consistent evaluation and comparison of classification performance [28].

R peaks were extracted from the annotation files, and QRS segments were generated using 20 samples before and 30 samples after each R peak, ensuring full coverage of the QRS complex given the MIT-BIH sampling frequency. The segments were grouped into five classes commonly used in the literature—normal (N), supraventricular (S), premature ventricular (V), faulty (F), and unknown/other (U)—by combining the corresponding MIT-BIH labels.

The extracted QRS segments were stored as class-specific matrices, and the first 100 segments per class were selected and plotted. The resulting representations illustrate amplitude variations across segment samples and highlight morphological differences among heartbeat classes, providing a visual reference for morphological diversity and supporting subsequent classification and feature extraction stages (Figure 2).

The primary objective of this study is to develop an automated system for the classification of cardiac arrhythmias based on ECG signals, where each QRS complex is assigned to one of five predefined classes (N, S, V, F, or U). For a typical 30-minute recording, all detected QRS complexes are labeled according to this classification scheme.

Such a system facilitates efficient analysis of long-term ECG recordings by providing rapid and consistent classification results. To train the proposed model, a representative subset of QRS segments from each class is selected, enabling the model to learn discriminative patterns and accurately classify QRS complexes into the defined categories.

Prior to classification, a preprocessing stage is applied to accurately detect QRS complexes and isolate them from other ECG signal components. The training dataset is constructed using cardiologist-annotated signals provided in the database.

Since the original annotation scheme contains a larger set of labels, these labels are reorganized into five standard classes (N, S, V, F, U) by mapping the corresponding MIT-BIH annotations, following common practices in the literature [29]. This relabeling step ensures consistency and enables effective multi-class classification.

2.2 Educational Illustrations: Demonstration of the Pan-Tompkins Algorithm

Although expert annotations from the MIT-BIH database are used for QRS detection in the main analysis, the Pan-Tompkins algorithm—a well-established method for real-time QRS detection—is presented here for illustrative and educational purposes. A representative ECG segment from Record 105 (MLII channel, first 6,000 samples) is used to demonstrate the key processing stages of the algorithm.

The algorithm detects QRS complexes through a sequence of processing stages. First, band-pass filtering (cascaded high- and low-pass filters) is applied to remove baseline drift and high-frequency noise. The signal is then differentiated to emphasize rapid slope changes, squared to enhance high-amplitude variations, and smoothed using moving-window integration to obtain an energy envelope corresponding to the QRS duration. Finally, adaptive thresholding is employed to identify R-wave locations [32]. This processing pipeline enables accurate real-time QRS detection and provides a reliable basis for subsequent feature extraction.

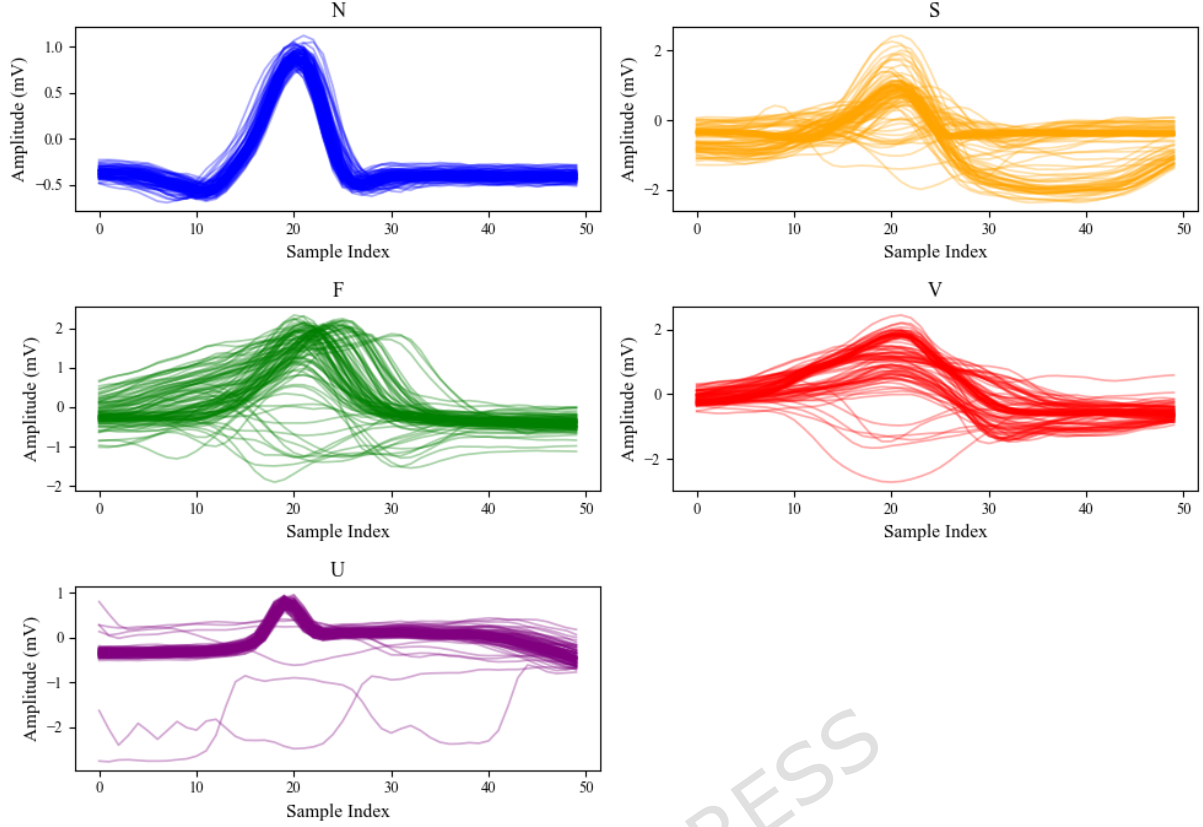


Fig. 2: Representative QRS segments illustrating morphological variations across different heartbeat classes in the MIT-BIH Arrhythmia Database.

For visualization purposes, a single-channel ECG segment (Record 105, MLII channel) was selected as a representative example. A 0.15-second Hanning window was applied during the moving-window integration stage, approximating the typical QRS duration while reducing edge effects. In addition, the minimum distance between consecutive R peaks was set to 0.6 s (corresponding to approximately 100 bpm) to prevent false detections.

Figure 3 illustrates the integrated Pan-Tompkins workflow.

Figure 4 summarizes the preprocessing and R-peak detection results for the first 6,000 samples (approximately 16.7 s, $f_s = 360$ Hz). The six subplots present the original signal, filtered signal, derivative, squared signal, moving-window integration output, and detected R-peaks, respectively. The red triangles in the final subplot

indicate the detected R-peak locations, confirming reliable QRS isolation and accurate detection performance.

R-peaks play a central role in ECG analysis [33], serving as reference points for cardiac cycle timing and signal segmentation. The intervals between consecutive R-peaks, known as R-R intervals (RRI), are fundamental for heart rate and rhythm analysis, including the detection of arrhythmia, tachycardia, and bradycardia. In this study, RR intervals and RRI are used interchangeably to represent the time difference between consecutive R-peaks. In addition, R-peaks provide temporal reference points for identifying P, Q, S, and T waves, supporting both morphological and dynamic ECG analysis.

Accurate R-peak detection is therefore essential for reliable feature extraction [45], segmentation, and subsequent classification tasks. Segmentation enables the isolation of individual cardiac

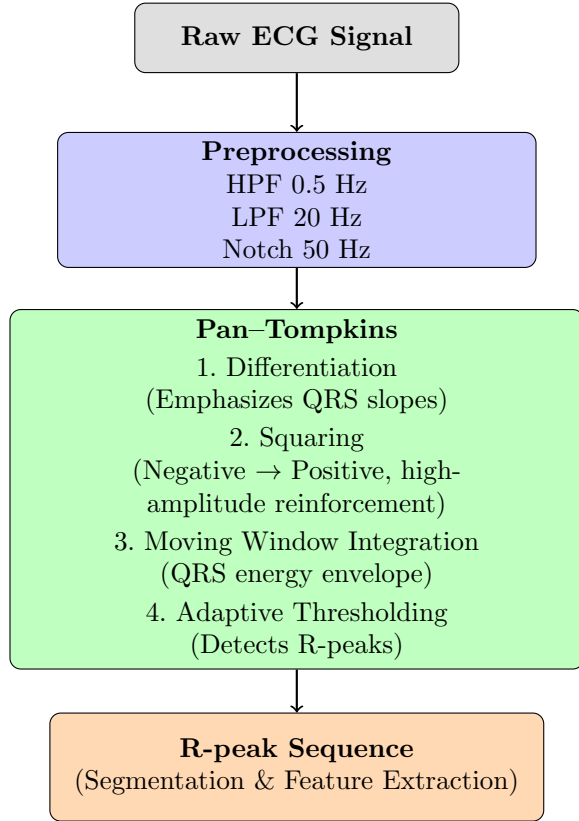


Fig. 3: Integrated workflow of the Pan-Tompkins algorithm for ECG signal processing and R-peak detection.

cycles for detailed analysis, whereas inaccurate segmentation may lead to distorted waveform interpretation and reduced classification performance [46].

During segmentation, each R-peak defines the onset of a cardiac cycle, partitioning the ECG signal into individual heartbeat units for morphological analysis, feature extraction, and classification [47]. The resulting R-R intervals (RRI), defined as the temporal differences between consecutive R-peaks, provide key information for heart rate and rhythm analysis, including the detection of tachycardia, bradycardia, and irregular arrhythmias.

Heart rate (HR) is computed from R-R intervals as their inverse in beats per minute, enabling dynamic assessment of cardiac activity. HR analysis captures both instantaneous rate and rhythm

stability, forming the basis for heart rate variability (HRV) analysis and providing temporal references for P, Q, S, and T waves [48].

Segmentation based on R-peaks and RRI supports systematic analysis of both morphological and temporal ECG features, forming a foundation for monitoring and classification. In this study, RRI values were calculated from consecutive R-peak positions using both direct index differences and the NumPy `diff` function, yielding identical results; the `diff` function was adopted for computational efficiency.

The `locs` variable obtained from R-peak detection represents the sample indices of detected R-peaks, providing temporal reference points for each cardiac cycle. The R-R interval (RRI) is defined as the difference between consecutive R-peak locations, as given in Equation 2.

$$\text{RRI}_i = \text{locs}[i + 1] - \text{locs}[i] \quad (2)$$

By converting sample indices to time using the sampling frequency f_s , R-R interval (RRI) values can be expressed in seconds. The instantaneous heart rate (HR) is then computed from RRI as shown in Equation 3 [49]:

$$\text{HR}(t) = \frac{60}{\text{RRI}(t)} \quad (3)$$

The mean heart rate over the observation period is obtained by averaging all instantaneous HR values, as given in Equation 4:

$$\mu_{\text{HR}} = \frac{1}{N} \sum_{i=1}^N \text{HR}(i) \quad (4)$$

This formulation enables the evaluation of both instantaneous heart rate variability and overall cardiac behavior. The mean HR is a commonly reported parameter in ECG analysis.

After preprocessing and R-peak detection, QRS segmentation was performed to decompose the ECG signal into individual cardiac cycles. For illustration, a single channel from record 100 of the MIT-BIH Arrhythmia Database (first 6000 samples, ≈ 16.7 s, $f_s = 360$ Hz) was analyzed. Segments were generated around each R-peak, including 20 samples (≈ 55 ms) before and 80 samples (≈ 222 ms) after the peak, thereby isolating the QRS morphology.

The resulting superimposed segments (Figure 5) highlight the repetitive structure and

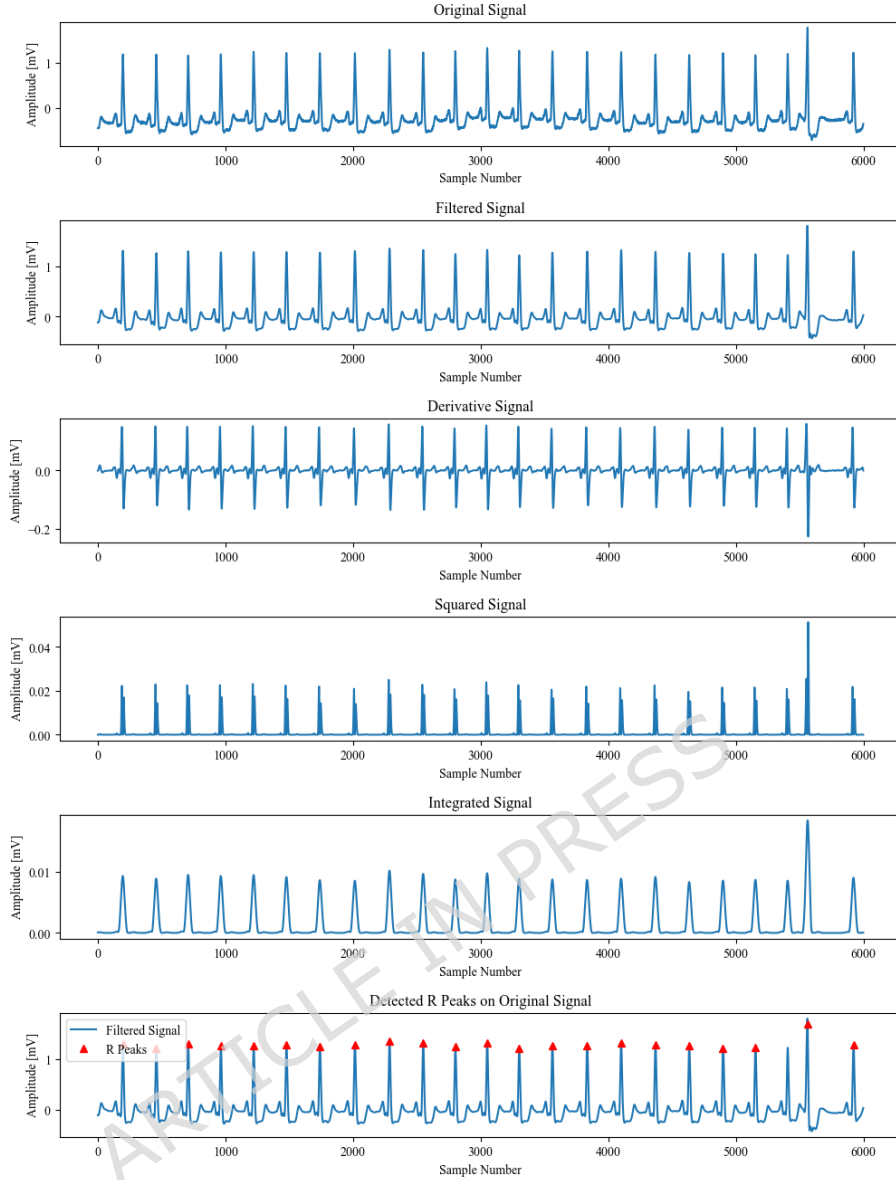


Fig. 4: Preprocessing and R-peak detection for the first 6,000 samples (≈ 16.7 s, $f_s = 360$ Hz). The six subplots show: (a) original signal, (b) filtered signal, (c) derivative, (d) squared signal, (e) moving-window integration, and (f) detected R-peaks (red triangles). Horizontal and vertical axes represent sample number and amplitude [mV], respectively.

morphological consistency of QRS complexes. This observation confirms that R-peak-centered segmentation provides accurate and reliable reference points for subsequent analysis, feature extraction, and classification tasks.

2.3 Working Framework

All analyses were conducted in Python (version 3.11.7) using the Anaconda Navigator environment, with Visual Studio Code employed as the development platform. Signal processing and feature extraction were implemented using standard scientific computing libraries, including NumPy

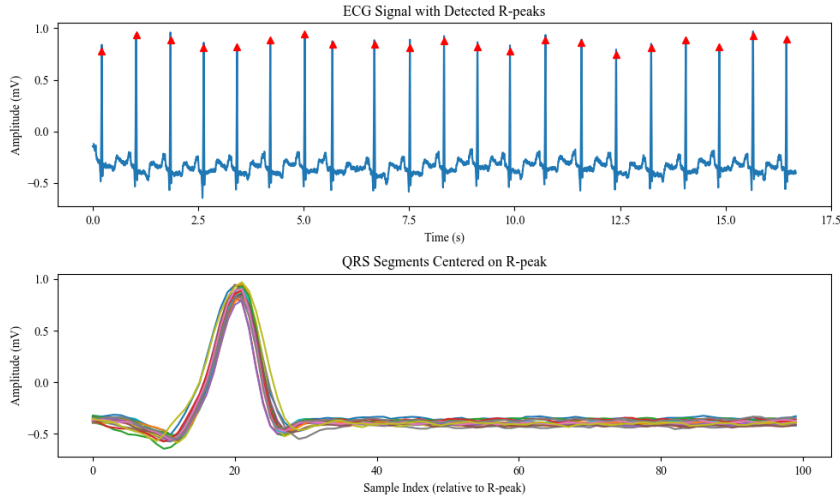


Fig. 5: QRS segments centered on R-peaks in a 6000-sample ECG segment (≈ 16.7 s) from record 100 of the MIT-BIH Arrhythmia Database, extracted using a sample window of $[-20, +80]$.

and SciPy, while the WFDB Python package was utilized for handling the MIT-BIH Arrhythmia dataset.

2.4 Preprocessing and Noise Management in Cardiac Signal Processing

ECG analysis requires preprocessing prior to segmentation or feature extraction. This stage aims to remove unwanted components and noise, thereby improving the accuracy and reliability of subsequent analyses.

In the signal processing literature, noise is defined as any unwanted component added to a signal during recording, transmission, or processing that obscures the true characteristics of the signal [50]. In ECG signals, such noise can distort waveform morphology and lead to inaccurate analysis, misinterpretation, and incorrect detection of cardiac events [51].

The main types of noise affecting ECG signals include low-frequency baseline drift, high-frequency muscle activity (EMG) noise, powerline interference, and electrode motion artifacts. Baseline drift, typically occurring in the 0–0.5 Hz range, arises from respiratory movements, patient posture, or electrode placement, and can interfere with accurate ST-segment analysis. Muscle activity noise is characterized by high-frequency and irregular components caused

by patient movements, tremors, or involuntary contractions, which can hinder the detection of P and T waves. Powerline interference is a sinusoidal disturbance at 50 or 60 Hz, depending on the region, and complicates the analysis of high-frequency components. Electrode motion artifacts result from abrupt changes in electrode–skin contact, adversely affecting signal segmentation and waveform detection [52].

In preprocessing, the Fourier transform (FT) is employed to analyze the frequency components of ECG signals and to identify different types of noise [53]. In the frequency domain [54], low-frequency baseline drift can be mitigated using high-pass or band-pass filters, while powerline interference, which appears as a peak at a specific frequency, can be effectively suppressed using a notch filter. High-frequency, irregular muscle noise can be attenuated using a low-pass filter.

In this study, a high-pass filter with a cutoff frequency of 0.5 Hz was applied to suppress baseline drift, while a low-pass filter with a cutoff frequency of 20 Hz was used to reduce high-frequency muscle noise while preserving the QRS morphology. Additionally, a notch filter in the 49–51 Hz range was employed to eliminate mains interference. All signals were processed in millivolts (mV), as provided by the database.

In addition to conventional filtering approaches, advanced data-driven denoising techniques have recently gained increasing attention

in ECG signal processing [44]. Methods such as wavelet-based denoising, adaptive filtering, and deep learning-based models (e.g., diffusion probabilistic models) have been shown to provide improved noise suppression while preserving subtle morphological characteristics of ECG signals. However, in this study, a relatively simple and well-established filtering strategy was preferred to maintain interpretability, reproducibility, and consistency with the proposed framework. Nevertheless, the integration of advanced denoising techniques represents a promising direction for future research.

This preprocessing stage produced cleaner and more reliable signals, facilitating the accurate determination of characteristic features. In particular, QRS complex detection is critical for identifying heartbeat timing and serves as the foundation for subsequent feature extraction. In this study, QRS locations were not estimated automatically; instead, they were obtained directly from expert annotations in the MIT-BIH Arrhythmia Database, thereby ensuring high reliability for subsequent analysis and evaluation.

The MIT-BIH Arrhythmia Database records used in this study were sampled at 360 Hz with 11-bit resolution and a ± 10 mV dynamic range. Filtering was performed using third-order Butterworth high-pass and low-pass filters, applied with the forward-backward zero-phase (`filtfilt`) method to prevent phase distortion.

2.5 Feature Extraction in Cardiac Signal Analysis

Clinical information derived from electrocardiography (ECG) signals extends beyond direct inspection of the raw waveform and is significantly enhanced through feature extraction across multiple domains. Feature extraction constitutes a fundamental stage in biomedical signal processing and directly influences the performance of classification and diagnostic algorithms. Typically, features are categorized into three groups: time-domain, frequency-domain, and time-frequency-domain features.

In the time domain, the morphological characteristics of the ECG signal—including the amplitudes, durations, and intervals of the P, QRS, and T waves—can be directly measured and provide critical information about cardiac function [38].

Frequency-domain analysis examines the distribution of signal energy to extract spectral features associated with physiological processes. However, since ECG signals are inherently non-stationary, analyses limited to either the time or frequency domain may be insufficient. In such cases, time-frequency methods enable the simultaneous examination of temporal and spectral components, providing a more comprehensive representation of the signal.

In addition to time-domain analyses, frequency-domain methods play a crucial role in feature extraction from ECG signals. One of the most widely used approaches is the Fast Fourier Transform (FFT) [53], which computes the frequency spectrum of the signal and enables the analysis of its energy distribution across different frequency bands. The FFT of a discrete signal $x(n)$ is defined as $X(f) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi fn/N}$, which provides a representation of the signal in the frequency domain. The Power Spectral Density (PSD), derived from the FFT, quantitatively represents how the signal's energy is distributed among its frequency components.

Relative power values—particularly within the very low frequency (VLF), low frequency (LF), and high frequency (HF) bands—are widely used in clinical and analytical studies, as they facilitate comparisons across individuals and signals of varying lengths. However, a fundamental limitation of FFT-based analysis is that the resulting spectrum reflects an average frequency distribution over the entire signal. Consequently, short-term or transient events may be overlooked, limiting the ability of FFT to fully capture the dynamic characteristics of non-stationary biological signals.

Because FFT-based methods are limited in handling non-stationary signals, the Short-Time Fourier Transform (STFT) was developed to analyze time-varying frequency components [55]. In this approach, the signal is segmented into windows of a specified length, and a Fourier transform is applied to each segment to obtain a time-frequency representation. The STFT of a signal $x(t)$ is defined as $STFT\{x(t)\}(t, f) = \int x(\tau)w(\tau - t)e^{-j2\pi f\tau}d\tau$, which enables the analysis of localized frequency content over time. This allows the observation of temporal variations in the frequency content of the ECG signal and facilitates the analysis of transient events.

A key advantage of STFT is its ability to perform frequency analysis on non-stationary signals [56]. However, its reliance on a fixed window size introduces an inherent limitation: a wide window improves frequency resolution at the expense of time resolution, whereas a narrow window enhances time resolution but degrades frequency resolution. This time–frequency resolution trade-off is particularly critical in biomedical signal analysis, where both low- and high-frequency components must be accurately characterized simultaneously.

The Wavelet Transform (WT) was developed to overcome the time–frequency resolution limitations of traditional methods. By decomposing the signal across multiple scales and shifts, WT provides high temporal resolution for short-duration, high-frequency components and high frequency resolution for long-duration, low-frequency components. This property makes it particularly well-suited for ECG analysis, enabling the effective detection of rapid events such as QRS complexes as well as slower morphological components, including P and T waves. The Continuous Wavelet Transform (CWT) of a signal $x(t)$ is defined as $W(a, b) = \int x(t)\psi^*\left(\frac{t-b}{a}\right)dt$, where a and b denote the scale and translation parameters, respectively.

WT can be implemented as either the Continuous Wavelet Transform (CWT) or the Discrete Wavelet Transform (DWT). Although CWT provides a highly detailed time–frequency representation, its high computational cost limits its practical applicability. In contrast, DWT offers a more compact and computationally efficient representation, making it more suitable for feature extraction and classification tasks [42].

Because wavelet-based features can be high-dimensional and potentially redundant, dimensionality reduction techniques are commonly employed. Principal Component Analysis (PCA) reduces feature dimensionality by retaining components that capture the majority of the variance, thereby enhancing classification performance while reducing computational complexity and mitigating the risk of overfitting [43].

Morphological features extracted from electrocardiography (ECG) signals are derived from the fundamental waveforms of cardiac electrical activity. These features include the amplitudes, durations, and intervals of the P wave, QRS

complex, and T wave, and they provide direct clinical insight into cardiac function. For example, the peak amplitude and width of the QRS complex reflect the characteristics of ventricular depolarization, while the P wave corresponds to atrial depolarization and the T wave represents ventricular repolarization. Consequently, morphological features serve as a fundamental reference for identifying cardiac arrhythmias and structural abnormalities [57].

Temporal features capture timing patterns and sequential relationships within the ECG signal [58]. These features quantify correlations and temporal variations between consecutive QRS complexes. For instance, feature C11 represents the maximum correlation between a QRS complex and the subsequent QRS complex, C22 represents the maximum correlation between a QRS complex and the preceding QRS complex, and Cxy denotes the maximum correlation between a QRS complex and reference (template) QRS complexes. Such temporal analyses play a critical role in detecting arrhythmias, particularly short-term rhythm disturbances, and, when combined with morphological features, enhance the discriminative power of classification algorithms [59].

To further enhance both morphological and temporal representations, wavelet-based decomposition was applied to the ECG signals. This approach provides high temporal resolution for abrupt events such as QRS complexes, as well as high frequency resolution for slower components such as P and T waves.

In this study, three main feature groups were extracted from the ECG signals: morphological, temporal, and time–frequency (wavelet-based) features. Each recording was processed using the `wfdb` library, and R-wave positions and class labels were obtained directly from the annotation files. The signals were preprocessed using high-pass and low-pass filters to suppress baseline drift and high-frequency noise. Forward–backward zero-phase filtering (`filtfilt`) was applied to preserve morphological characteristics while avoiding phase distortion.

As a fundamental step in feature extraction, QRS-centered segments were defined around each R wave. Although the physiological duration of the QRS complex is typically 80–120 ms, a wider window was selected to include buffer regions before and after the QRS complex, ensuring

robust extraction of morphological, temporal, and wavelet-based features. Accordingly, fixed-length segments of 200 samples were extracted, consisting of 80 samples before and 120 samples after each R wave.

The pre-R interval ensures that the onset of the Q wave is captured, while the post-R interval preserves the complete S wave and the early transition toward repolarization. This R-centered segmentation strategy is consistent with commonly used approaches in the literature and provides a stable and uniform window for subsequent feature extraction.

In preliminary experiments, several window lengths within the selected range were evaluated. Minor variations in window size did not significantly affect classification performance; however, overly short windows failed to capture essential waveform components, whereas excessively long windows introduced redundant information. Accordingly, the selected window length provides a balanced and robust trade-off between physiological relevance and classification performance.

Morphological features were extracted using correlation coefficients computed between the current ECG segment and the preceding segment, the succeeding segment, and a reference segment. These coefficients quantify the morphological similarity between consecutive heartbeats. The correlation coefficient between two signals x and y is defined as given in Equation 5.

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5)$$

where $\bar{x}$ and $\bar{y}$ denote the mean values of the respective signals.

Temporal features were derived from sequences of consecutive RR intervals. Specifically, three consecutive intervals (RR0, RR1, RR2), their mean value (RRm), interval ratios (RR0/RR1, RR2/RR1, RRm/RR1), and the mean of the preceding 20 RR intervals were used as time-based features. These features capture both short- and long-term rhythm variations associated with arrhythmias and enhance the discriminative capability of the classification algorithm.

In addition to morphological and temporal features, a discrete wavelet transform (DWT)

was applied to each QRS segment. Four levels of decomposition were performed using the Daubechies 6 (db6) wavelet, producing detail coefficients that capture short-term, abrupt changes and approximation coefficients that represent slower signal variations. These wavelet coefficients were combined with morphological correlation features to construct the morphological feature vector.

Finally, a unified feature vector was constructed for each R wave by combining morphological, temporal, and wavelet-based features, with corresponding class labels representing the five heartbeat classes. All features were extracted on a per-beat (R-wave-centered) basis, without record-level aggregation or averaging prior to classification. This strategy enhances classification performance by jointly representing morphological characteristics, temporal dynamics, and time-frequency information. The resulting high-dimensional feature set was subsequently optimized using feature selection techniques and employed for model training.

2.6 Feature Conditioning

Before being fed into the classifier, the extracted features were subjected to a feature conditioning stage. The main objective of this step is to optimize the feature space, improve classification performance, reduce the risk of overfitting, and ensure computational efficiency [39].

Feature conditioning generally involves two main approaches: feature selection and feature mapping. In feature selection, features are chosen according to predefined criteria without altering their original structure, and redundant and non-informative features are removed. In contrast, feature mapping transforms the original features into a new space; for example, Principal Component Analysis (PCA) linearly combines features to maximize variance, while Linear Discriminant Analysis (LDA) focuses on maximizing class separability [60]. These methods enable high-dimensional data to be represented in a more compact and discriminative form.

In this study, rather than explicitly mapping features into a new space using PCA or LDA, feature scaling and feature selection were employed. Specifically, Sequential Forward Feature Selection (SFFS) was applied to identify an informative

subset of features while preserving their original representation.

After feature extraction, feature scaling is a critical step to ensure effective classifier performance. ECG-derived features often differ in units and magnitude; for instance, morphological correlation coefficients and RR intervals can vary substantially in scale. Without scaling, features with larger numerical ranges may dominate the learning process, leading to biased model behavior, particularly in distance-based and gradient-based classifiers.

In this study, all features were normalized using Z-score standardization [61]. Each feature value x was transformed as

$$z = \frac{x - \mu}{\sigma} \quad (6)$$

where μ and σ denote the mean and standard deviation, resulting in zero mean and unit variance. This normalization ensures equal feature weighting and improves numerical stability during model training. Importantly, the mean and standard deviation were computed exclusively from the training data and subsequently applied to the validation and test sets, preventing data leakage between training and evaluation stages and enabling an unbiased assessment of model performance.

2.7 Feature Selection and SFFS

In ECG-based classification studies, feature extraction methods such as wavelet transforms, time-frequency analysis, and morphological measurements typically produce a high-dimensional feature space. This space often includes redundant, highly correlated, or weakly informative features, which can increase sensitivity to noise and lead to overfitting, thereby degrading generalization to unseen data. Feature selection addresses this issue by systematically identifying features that effectively discriminate between classes while preserving complementary information when used jointly [54].

Unlike filter-based methods, which assess features individually using statistical criteria, wrapper-based approaches integrate the classifier into the selection process and explicitly evaluate feature interactions and their combined contribution to classification performance. To leverage this

advantage, Sequential Forward Feature Selection (SFFS) was employed in this study [41]. Formally, SFFS aims to identify the optimal feature subset given by

$$S^* = \arg \max_{S \subseteq F} J(S) \quad (7)$$

where F denotes the full feature set and $J(S)$ represents the validation performance obtained using the selected subset S . Accordingly, SFFS incrementally constructs an optimal feature subset by eliminating redundant and non-informative features while maximizing classifier performance, resulting in a compact and discriminative feature set with improved generalization capability [62].

The puzzle metaphor used in this study illustrates the relationship between information integrity and redundancy in feature selection. The structure on the right represents complementary features that collectively form a coherent and meaningful whole, each contributing distinct information to the classification model. In contrast, the structure on the left depicts repeated instances of the same feature type, highlighting redundancy caused by duplicated information. Such repetitive features do not introduce new information, increase computational cost, and may negatively affect model performance. As shown in Figure 6, treating features as “puzzle pieces” schematically demonstrates how redundancy disrupts information integrity, whereas complementary features strengthen the overall model.

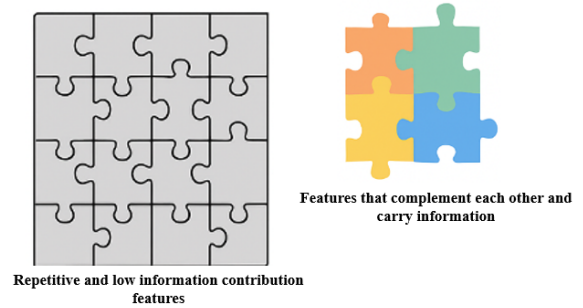


Fig. 6: The concept of information integrity and redundancy in feature selection.

Figure 7 illustrates the key differences between filter- and wrapper-based feature selection approaches [63]. Filter methods rank features using statistical criteria independent of the

classifier, offering high computational efficiency but ignoring inter-feature dependencies. In contrast, wrapper methods evaluate feature subsets directly based on classifier performance, explicitly accounting for interactions between features. Using sequential search strategies such as SFFS, wrapper approaches identify feature combinations that maximize classification performance, albeit at the cost of increased computational complexity.

In the proposed study, the SFFS method starts with an empty feature set. At each iteration, features that have not yet been selected are evaluated by adding them individually to the current set. For each candidate addition, the model is trained and the resulting classification accuracy on the validation set is computed from the confusion matrix. The feature whose inclusion yields the greatest improvement in validation accuracy is then added to the selected set and removed from the search space. This procedure is repeated until the predefined `max_iter` criterion is reached. Consequently, the model incrementally identifies the features that contribute most to validation performance, and the final feature subset used for classification consists of those that achieve the highest accuracy.

2.8 Arrhythmia Classification Pipeline

For formal clarity, the proposed arrhythmia classification framework can be summarized as follows. Let $x \in \mathbb{R}^n$ denote a preprocessed ECG segment centered on an R peak. After feature extraction, each segment is represented by a feature vector

$$f = \phi(x) \in \mathbb{R}^d,$$

where $\phi(\cdot)$ denotes the feature extraction mapping and d is the total number of extracted features. Then, SFFS selects an informative subset of feature indices

$$I^* \subseteq \{1, 2, \dots, d\},$$

corresponding to the optimal subset defined in Equation 7. The reduced feature vector f_{I^*} is subsequently used for classification, and the predicted class label is obtained as

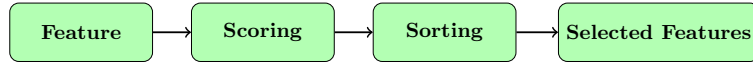
$$\hat{y} = h(f_{I^*}),$$

where $h(\cdot)$ denotes the classifier, implemented in this study using either k-NN or SVM.

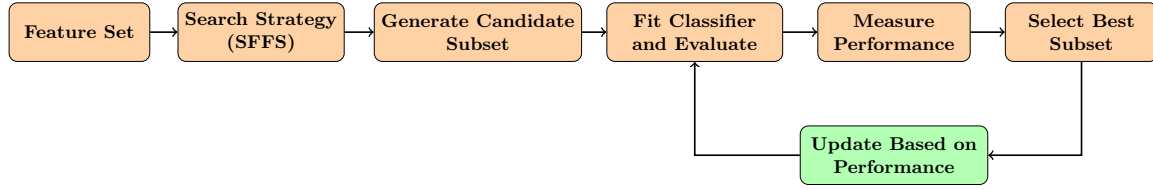
The k-Nearest Neighbor (k-NN) algorithm is a heuristic classification method that assigns class labels based on geometric proximity in the feature space [64]. Its fundamental assumption is that samples with similar characteristics are located close to each other. The class of a new sample is determined by the majority label among its k nearest neighbors in the training set. In this study, Euclidean distance was used, as it is widely preferred for continuous and normalized biomedical features. Since k-NN does not involve an explicit training phase, model performance is directly influenced by the structure of the feature space and the selected features. Accordingly, the compact and information-dense feature subsets obtained via SFFS enhance k-NN performance by clarifying neighborhood relationships. For this reason, k-NN was adopted as a simple yet effective benchmark classifier for pattern-oriented tasks such as ECG arrhythmia classification [65].

Support Vector Machines (SVMs) are margin-based classifiers that aim to determine the optimal decision boundary (hyperplane) that maximizes the separation between classes [66]. Through kernel functions, SVMs can transform nonlinear classification problems into linearly separable ones in a higher-dimensional feature space. In this study, the radial basis function (RBF) kernel was employed, as it provides a smooth and flexible decision surface capable of modeling local similarities and within-class variations commonly observed in biomedical signals. SVMs are inherently resistant to overfitting because the decision boundary is determined only by a subset of critical samples known as support vectors. This property, combined with the information-dense feature subsets selected by SFFS, enables an effective representation of morphological and time-frequency differences between classes [67]. Accordingly, SVM was used as a complementary classifier to contrast margin-based decision behavior with the distance-based k-NN approach.

The k-NN and SVM models were employed as complementary classification approaches because they rely on fundamentally different decision mechanisms. The k-NN classifier assigns class labels based on geometric proximity in the feature space; therefore, the number of neighbors, which is its primary hyperparameter, was initially



Filter approach: Features are evaluated independently; the classifier is not involved in the process.



Wrapper approach: Feature combinations are evaluated through the classifier.

Fig. 7: Comparison of filter and wrapper feature selection approaches.

set to $k = 3$, a commonly adopted baseline in studies involving small and heterogeneous biomedical datasets. Euclidean distance was chosen as the similarity metric because it effectively reflects neighborhood relationships in normalized continuous feature spaces. The influence of varying the number of neighbors within the range $k = 3$ to $k = 11$ on classification performance is examined separately in the Results section [68]. In addition, the compact and information-dense feature subsets obtained through SFFS improve the definition of neighborhood structures in the feature space, thereby enhancing the spatial discrimination capability of the k-NN classifier.

SVM learns a decision boundary that maximizes the margin between classes, and the RBF kernel was used to provide a flexible representation for nonlinear patterns. The kernel parameter γ was set to `auto`, and the regularization coefficient C was set to 1000 as initial values, based on ranges commonly reported in the literature. The impact of a wider hyperparameter range on model performance is further evaluated in the Results section [69]. The decision mechanism of SVM, which relies only on support vectors, combined with informative feature subsets selected by SFFS, enables a more accurate representation of morphological and time-frequency characteristics across classes. Together with k-NN, this approach allows the data space to be analyzed from both geometric proximity and margin-based separation perspectives, providing a comprehensive view of the multiclass arrhythmia classification problem.

The performance of all classification models was evaluated using a repeated holdout validation scheme [70]. Data splitting was performed at the beat level, where ECG beats were randomly divided into training and test sets without explicit patient-level grouping. Although beat-level splitting may introduce subject overlap between training and test sets, this study focuses on evaluating feature representation and selection behavior under a controlled experimental setting rather than subject-independent generalization. The dataset was initially partitioned into training (70%) and test (30%) subsets. To guide the SFFS procedure, the training set was further divided into an internal training subset (70% of the training data) and a validation subset (30%). Feature selection was conducted exclusively using this internal train-validation split. The optimal feature subset was then used to retrain the model on the full training set. Final performance evaluation was performed on the held-out test set, which remained completely unseen during feature selection and model optimization.

This validation strategy was adopted to ensure a clear separation between feature selection, model development, and final testing, thereby preventing information leakage during the SFFS process. Although cross-validation is widely used, it was not preferred in this study because the objective is to provide a clear and reproducible comparison under a controlled experimental setup. In addition, repeated cross-validation combined

with wrapper-based feature selection would significantly increase computational cost without providing additional meaningful insight into model behavior.

The computational complexity of the proposed framework is primarily governed by three stages: feature extraction, feature selection, and classification. Feature extraction is performed on a per-beat basis and scales linearly with the number of ECG beats and the number of extracted features. The SFFS stage represents the most computationally demanding component, as it iteratively evaluates candidate feature subsets using an internal validation set. To control this cost, the maximum number of iterations and the step size were deliberately constrained. The final classification stage using k-NN and SVM introduces relatively low computational overhead compared to feature selection, particularly during inference. Overall, the proposed pipeline exhibits manageable computational complexity and is well suited for offline analysis and educational benchmarking purposes.

2.9 Performance Evaluation

Because the MIT-BIH Arrhythmia Database exhibits a highly imbalanced class distribution, performance assessment was not based solely on overall accuracy. Instead, balanced accuracy, Cohen's kappa, and macro-averaged performance metrics were emphasized to ensure fair evaluation across both majority and minority arrhythmia classes. Within this framework, k-nearest neighbor (k-NN) and support vector machine (SVM) classifiers were implemented and analyzed to examine how different decision mechanisms affect classification performance when identical feature representations are used for ECG-based arrhythmia detection.

In the first stage, a confusion matrix was constructed to analyze correct and incorrect classifications in the test set on a per-class basis. Using this matrix, accuracy, precision, recall, and F1-score were computed for each class [71], and their macro-averaged counterparts (Macro-Precision, Macro-Recall, and Macro-F1) were obtained to summarize class-wise performance. Given the pronounced class imbalance in the MIT-BIH Arrhythmia Database, balanced accuracy was adopted to

ensure equal weighting of sensitivity and specificity across classes, thereby preventing dominant majority classes from masking minority-class performance. In addition, Cohen's kappa coefficient was employed to account for agreement occurring by chance, providing a more reliable measure of classifier consistency beyond raw accuracy [72]. Together, these metrics enable a comprehensive evaluation of both overall performance and class discrimination capability.

To obtain a more robust and reliable estimate of model generalization, repeated holdout validation was adopted using 10 independent random splits. For each split, the complete pipeline, including feature selection, classifier training, and evaluation, was re-executed. The resulting performance metrics were then aggregated and reported as mean $\pm$ standard deviation. This approach explicitly quantifies the variability introduced by data partitioning and provides a more reliable assessment of model stability and generalization.

Although multiple classifier configurations are compared, formal statistical significance testing was not included. Instead, model behavior was interpreted through repeated holdout validation, class-balanced performance metrics, confusion matrix analysis, and the Golden Distance (GD) framework.

The mathematical definitions of the performance metrics are organized into two tables. Table 2 lists the symbolic definitions of parameters shared across all metrics, while Table 3 presents the main evaluation metrics, including accuracy, precision, recall, F1 score, balanced accuracy, macro-averaged metrics, and Cohen's kappa, together with their formulas. This structure avoids repeated symbol definitions and improves clarity. The term *Macro-Metric* denotes the arithmetic mean of class-wise Precision, Recall, and F1 values, yielding Macro-Precision, Macro-Recall, and Macro-F1 scores that are independent of class distribution and assign equal weight to all classes.

In the second stage of performance analysis, the Golden Distance (GD) method was employed to condense multiple evaluation metrics into a single summary value. Metrics defined on the same scale are mapped into a multidimensional space using polar coordinates to form a performance polygon. From this polygon, two quantities are derived: the normalized area value (NAV), representing overall performance, and the distance

Table 2: Parameters Used in Performance Metrics

Parameter	Definition
TP_i	Number of true positives for class i
TN_i	Number of true negatives for class i
FP_i	Number of false positives for class i
FN_i	Number of false negatives for class i
K	Number of classes
κ	Kappa coefficient
p_o	Observed accuracy
p_i	Expected accuracy

of the centroid from the origin (DCO), reflecting the balance among metrics. A higher NAV indicates stronger performance, whereas a lower DCO indicates a more balanced distribution.

To eliminate the dependency on metric ordering, NAV and DCO are averaged over all possible permutations of the evaluation metrics. The Golden Distance (GD) is then defined as the Euclidean distance between the averaged (NAV, DCO) coordinates and the ideal reference point (1, 0):

$$GD = \sqrt{(1 - \overline{NAV})^2 + (\overline{DCO})^2} \quad (8)$$

A lower GD value indicates a model that achieves both high overall performance and a balanced distribution across evaluation metrics.

This evaluation enables an objective comparison of the effects of SFFS-selected feature subsets on the classifiers and provides a multidimensional perspective on model strengths and weaknesses. The Results section presents the detailed metric values and corresponding GD outputs and discusses the observed performance differences. Since GD is constructed from multiple normalized metrics and reflects their joint geometric structure, small variations in individual metrics result in gradual rather than abrupt changes in GD. It is important to note that GD is intended for relative comparisons within the same experimental setup, and direct comparison of absolute GD values across different studies is not recommended unless identical datasets, metrics, and normalization procedures are employed.

3 Results

This section presents the experimental results of the proposed ECG arrhythmia classification framework, emphasizing how preprocessing, feature representation, feature selection, and classifier design collectively influence classification performance. Rather than focusing solely on maximizing accuracy, the analysis aims to provide a comprehensive understanding of how methodological choices at each stage affect class discrimination, robustness, and overall model behavior.

The proposed framework was evaluated using two different feature representations: (i) a hand-crafted feature set composed of morphological and temporal descriptors, and (ii) a compact representation in which morphological features were transformed into a wavelet-based form while preserving temporal descriptors. For both representations, the most informative feature subsets were selected using Sequential Forward Feature Selection (SFFS), and classifier performance was evaluated. The evaluation metrics include total accuracy, balanced accuracy, macro precision, macro recall, macro F1-score, Cohen's kappa, confusion matrices, ROC analysis, and the Golden Distance (GD) metric as a holistic indicator of overall performance.

The first feature representation consists of 203 morphological and 8 temporal features, capturing QRS morphology, amplitude variations, duration characteristics, and temporal dynamics. In the second representation, the morphological component is transformed into a wavelet-based form, reducing it to 47 features while preserving the same 8 temporal descriptors. This transformation compresses the high-dimensional and partially correlated morphological feature space into a more compact and informative representation, reducing dimensionality and computational complexity while retaining discriminative time-frequency characteristics.

To identify compact and informative feature subsets, Sequential Forward Feature Selection (SFFS) was applied using validation-set accuracy as the selection criterion. The impact of subset size was analyzed progressively, and the optimal operating point was determined based on the saturation behavior of validation performance.

Table 3: Performance Metrics and Their Mathematical Definitions Used in This Study.

Metric	Formula	Explanation
Accuracy	$\frac{\sum_{i=1}^K TP_i}{\sum_{i=1}^K (TP_i + FN_i)}$	Overall success rate
Precision	$\frac{TP_i}{TP_i + FP_i}$	Positive predictive accuracy
Recall (Sensitivity)	$\frac{TP_i}{TP_i + FN_i}$	True positive rate
Specificity	$\frac{TN_i}{TN_i + FP_i}$	True negative rate
F1-score	$2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of precision and recall
Macro-Metric	$\frac{1}{K} \sum_{i=1}^K \text{Metric}_i$	Unweighted average across all classes
Balanced Accuracy	$\frac{\text{Recall} + \text{Specificity}}{2}$	Accuracy sensitive to class imbalance
Cohen's Kappa Coefficient	$\kappa = \frac{p_o - p_i}{1 - p_i}$	Agreement corrected for chance

As shown in Tables 4 and 5, both classifiers exhibit a rapid increase in validation accuracy during the early iterations, followed by a saturation-like behavior as the number of selected features increases. For the k-NN classifier, validation accuracy rises sharply from 70.16% to 98.34% within the first 15 features, after which only minor improvements are observed.

Similarly, the SVM classifier reaches 98.57% accuracy at 15 features. Beyond this point, increasing the number of features leads to small fluctuations rather than consistent improvement, with a peak of 98.93% observed at 27 features, followed by a slight decrease at 30 features.

These findings indicate that most discriminative information is captured within the first 15 selected features, while additional features provide limited and inconsistent gains and may introduce redundancy. Accordingly, a 15-feature subset (`numf = 15`) was selected as a balanced and computationally efficient configuration for subsequent experiments.

Classifier performance was evaluated across different feature representations and hyperparameter configurations to assess the impact of model complexity and feature design on classification outcomes. For the k-NN classifier, both Morphological+Temporal and Morphological+Temporal+Wavelet feature sets were examined using $k = 3, 5$, and 7 , while for the SVM classifier, multiple combinations of the regularization parameter C and kernel parameter γ were systematically explored. The resulting performance

metrics, including total accuracy, balanced accuracy, macro precision, macro recall, macro F1-score, Cohen's kappa, and GD, are reported in Tables 6–9.

To obtain a more robust estimate of generalization performance, repeated holdout validation was conducted using 10 independent random splits. For each split, the entire training, feature selection, and evaluation pipeline was repeated, and the resulting metrics were aggregated as mean $\pm$ standard deviation. This procedure enables the variability induced by data partitioning to be explicitly quantified and provides a more reliable assessment of model stability.

Tables 6 and 7 summarize the repeated-holdout performance of the k-NN classifier for the Morphological+Temporal and Morphological+Temporal+Wavelet feature sets, respectively. In both representations, the best overall performance is obtained at $k = 3$. As k increases to 5 and 7, a slight but consistent decline is observed in total accuracy, macro-level metrics, and Cohen's kappa, suggesting mild oversmoothing in the decision boundaries.

Comparing the two feature representations, the inclusion of wavelet-based features leads to a marginal improvement in total accuracy at $k = 3$, along with a noticeable reduction in GD (from 0.0398 to 0.0363), indicating a more balanced overall performance. While differences in macro-level metrics remain limited, these results suggest that the wavelet-based representation

Table 4: Validation performance of SFFS-selected feature subsets for the k-NN classifier ($k = 3$, Euclidean metric).

Iteration	Number of Selected Features	Newly Added Features (Indices)	Validation Accuracy (%)	Δ Accuracy
0	3	54, 49, 46	70.16	—
1	6	16, 13, 15	96.55	+26.40
2	9	52, 18, 50	97.50	+0.95
3	12	14, 30, 48	97.98	+0.48
4	15	10, 32, 33	98.34	+0.36
5	18	29, 25, 45	98.34	0
6	21	51, 40, 19	98.34	0
7	24	24, 12, 31	98.69	+0.36
8	27	11, 53, 21	98.69	0
9	30	36, 22, 44	98.69	0

Table 5: Validation performance of SFFS-selected feature subsets for the SVM classifier (RBF kernel, $C=1000$, $\gamma = \text{auto}$).

Iteration	Number of Selected Features	Newly Added Features (Indices)	Validation Accuracy (%)	Δ Accuracy
0	3	54, 46, 49	73.84	—
1	6	17, 16, 53	95.24	+21.40
2	9	14, 40, 45	97.38	+2.14
3	12	10, 31, 32	98.22	+0.83
4	15	20, 21, 50	98.57	+0.36
5	18	52, 15, 19	98.81	+0.24
6	21	33, 18, 37	98.81	0.00
7	24	12, 44, 43	98.81	0.00
8	27	51, 47, 11	98.93	+0.12
9	30	34, 26, 48	98.81	-0.12

provides complementary discriminative information by capturing time-frequency characteristics beyond purely morphological and temporal descriptors.

The SVM results, summarized in Tables 8 and 9, show that classification performance remains consistently high across the tested hyperparameter configurations. Variations in C and γ result in only minor changes in total accuracy, macro-level metrics, and Cohen's kappa, indicating that the classifier is relatively robust to moderate hyperparameter variations within the selected feature space.

Comparing the two feature representations, the inclusion of wavelet-based features leads to slightly higher accuracy and notably lower GD values, indicating a more balanced and stable performance. In addition, the reduced standard deviation observed in the wavelet-based configuration reflects improved robustness across different data partitions.

Taken together, the k-NN and SVM results indicate that both feature quality and classifier design contribute to classification performance. While well-selected features enable high performance across different classifiers, SVM generally

achieves slightly higher accuracy and more stable results, particularly when combined with wavelet-based representations.

To examine class-level behavior beyond aggregate metrics, the confusion matrix of the best-performing configuration, namely the SVM classifier with $C = 1000$ and $\gamma = \text{scale}$ using the 15-dimensional Morphological+Temporal+Wavelet feature subset selected by SFFS, was analyzed. As shown in Figure 8, the strong dominance of diagonal elements and minimal off-diagonal errors indicate effective inter-class separation.

This observation is consistent with the high overall accuracy, near-perfect macro-level metrics, and low GD value reported for this configuration, confirming reliable and well-balanced classification performance across all classes.

This result should be interpreted within the scope of the present study. The observed performance reflects the effectiveness of the selected feature representation under a deterministic classification setting. While the SVM model demonstrates stable and reliable behavior with the proposed features, more complex or stochastic

Table 6: Overall performance of the k-NN classifier using the Morphological+Temporal feature set under repeated-holdout evaluation (mean $\pm$ standard deviation).

Metric	k=3	k=5	k=7
Total Accuracy (%)	98.02 $\pm$ 0.28	97.80 $\pm$ 0.25	97.73 $\pm$ 0.46
Balanced Accuracy	0.99 $\pm$ 0.002	0.99 $\pm$ 0.002	0.99 $\pm$ 0.003
Macro Precision	0.98 $\pm$ 0.003	0.98 $\pm$ 0.003	0.98 $\pm$ 0.005
Macro Recall	0.98 $\pm$ 0.003	0.98 $\pm$ 0.002	0.98 $\pm$ 0.005
Macro F1-score	0.98 $\pm$ 0.003	0.98 $\pm$ 0.003	0.98 $\pm$ 0.005
Cohen's Kappa	0.98 $\pm$ 0.003	0.97 $\pm$ 0.003	0.97 $\pm$ 0.006
GD	0.0398	0.0405	0.0407

Table 7: Overall performance of the k-NN classifier using the Morphological+Temporal+Wavelet feature set under repeated-holdout evaluation (mean $\pm$ standard deviation).

Metric	k=3	k=5	k=7
Total Accuracy (%)	98.05 $\pm$ 0.386	97.73 $\pm$ 0.369	97.27 $\pm$ 0.359
Balanced Accuracy	0.99 $\pm$ 0.002	0.99 $\pm$ 0.002	0.98 $\pm$ 0.002
Macro Precision	0.98 $\pm$ 0.004	0.98 $\pm$ 0.004	0.97 $\pm$ 0.004
Macro Recall	0.98 $\pm$ 0.004	0.98 $\pm$ 0.004	0.97 $\pm$ 0.004
Macro F1-score	0.98 $\pm$ 0.004	0.98 $\pm$ 0.004	0.97 $\pm$ 0.004
Cohen's Kappa	0.98 $\pm$ 0.005	0.97 $\pm$ 0.005	0.97 $\pm$ 0.004
GD	0.0363	0.0407	0.0551

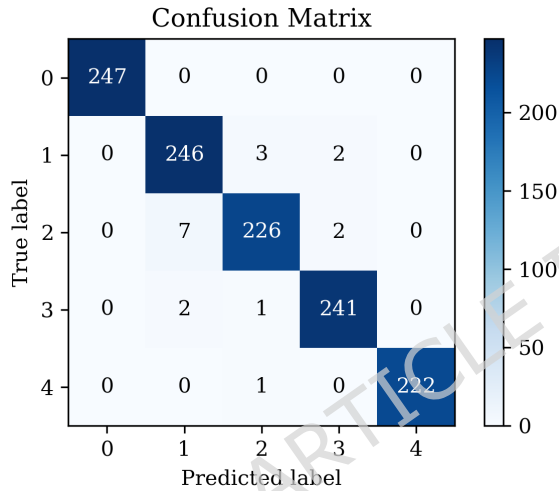


Fig. 8: Confusion matrix of the best-performing SVM configuration ($C = 1000$, $\gamma = \text{scale}$) using the 15-dimensional Morphological+Temporal+Wavelet feature subset selected by SFFS.

approaches, such as deep learning or ensemble-based methods, may introduce additional variability due to differences in model architecture and optimization dynamics.

The holistic performance of the wavelet-enhanced model was further analyzed using the

GD framework. Figure 9(a) illustrates the irregular hexagonal structure formed from six normalized performance metrics and its centroid. The near-symmetrical shape indicates that no single metric dominates the evaluation disproportionately.

Figure 9(b) combines the normalized area value (NAV) and centroid-to-origin distance (DCO) to compute GD as the Euclidean distance from the ideal point ($\text{NAV} = 1, \text{DCO} = 0$). To provide a conservative evaluation, the visualization corresponds to the k-NN configuration with the highest GD value (0.0551) among the tested settings. Despite representing a less optimal case, the resulting structure remains relatively balanced, indicating that the proposed feature representation maintains consistent performance across evaluation metrics even under less favorable conditions.

In contrast, the best-performing configuration, namely the SVM classifier with the Morphological+Temporal+Wavelet feature set, achieves a substantially lower GD value (0.0249), further confirming the robustness and effectiveness of the proposed framework.

In addition to GD-based evaluation, discriminative performance was also assessed using ROC analysis. Figure 10 shows the macro-averaged ROC curve for the SVM classifier using the Morphological+Temporal+Wavelet feature set. The

Table 8: Overall Performance Metrics of SVM Classifier for Morphological + Temporal Feature Set Performances under Different Hyperparameter Configurations with Robust Evaluation (Mean $\pm$ Std).

Metric	$C = 1000, \gamma = \text{auto}$	$C = 500, \gamma = \text{auto}$	$C = 1000, \gamma = \text{scale}$	$C = 500, \gamma = \text{scale}$
Total Accuracy (%)	98.38 ± 0.208	98.42 ± 0.083	97.88 ± 0.125	98.42 ± 0.083
Balanced Accuracy	0.99 ± 0.001	0.99 ± 0.001	0.99 ± 0.001	0.99 ± 0.001
Macro Precision	0.98 ± 0.002	0.98 ± 0.001	0.98 ± 0.002	0.98 ± 0.001
Macro Recall	0.98 ± 0.002	0.98 ± 0.001	0.98 ± 0.002	0.98 ± 0.001
Macro F1-score	0.98 ± 0.002	0.98 ± 0.001	0.98 ± 0.002	0.98 ± 0.001
Cohen's Kappa	0.98 ± 0.003	0.98 ± 0.001	0.97 ± 0.002	0.98 ± 0.001
GD	0.0392	0.0351	0.0402	0.0351

Table 9: Overall Performance Metrics of SVM Classifier for Morphological + Temporal + Wavelet Feature Set Performances under Different Hyperparameter Configurations with Robust Evaluation (Mean $\pm$ Std).

Metric	$C = 1000, \gamma = \text{auto}$	$C = 500, \gamma = \text{auto}$	$C = 1000, \gamma = \text{scale}$	$C = 500, \gamma = \text{scale}$
Total Accuracy (%)	98.54 ± 0.042	98.42 ± 0.083	98.54 ± 0.042	98.42 ± 0.083
Balanced Accuracy	0.99 ± 0.000	0.99 ± 0.001	0.99 ± 0.000	0.99 ± 0.001
Macro Precision	0.99 ± 0.000	0.98 ± 0.001	0.99 ± 0.000	0.98 ± 0.001
Macro Recall	0.99 ± 0.000	0.98 ± 0.001	0.99 ± 0.000	0.98 ± 0.001
Macro F1-score	0.99 ± 0.000	0.98 ± 0.001	0.99 ± 0.000	0.98 ± 0.001
Cohen's Kappa	0.98 ± 0.001	0.98 ± 0.001	0.98 ± 0.001	0.98 ± 0.001
GD	0.0249	0.0351	0.0249	0.0351

high AUC value indicates strong class separability across different decision thresholds. Together with the confusion matrix and GD analysis, this result further confirms that the proposed feature space enables robust and well-balanced multiclass classification.

Overall, the results lead to three main conclusions. First, most of the discriminative information is captured within a compact 15-feature subset selected by SFFS. Second, the inclusion of wavelet-based features provides a more balanced and effective representation compared to using only morphological and temporal features. Third, while both k-NN and SVM achieve high performance, SVM consistently provides slightly higher accuracy and more stable results, particularly when combined with the wavelet-based feature representation.

4 Discussion

In this study, the multi-class ECG arrhythmia classification problem is addressed within a holistic framework that aims not only to achieve high accuracy but also to provide a clear and systematic understanding of how signal processing, feature extraction, feature selection, and model

evaluation stages influence classification performance. Although very high accuracy rates have been reported on the same datasets in the literature, many of these approaches exhibit reduced interpretability and methodological transparency as model complexity increases [34].

In contrast, the proposed framework prioritizes methodological clarity, analytical consistency, and interpretability alongside performance, offering a balanced approach to ECG classification.

The results indicate that morphological and temporal features alone provide strong discriminative capability. However, incorporating wavelet-based representations, which transform morphological information into a more compact and decorrelated form, leads to a significant reduction in feature dimensionality and a more consistent and balanced classification performance. This observation is in agreement with previous studies demonstrating the effectiveness of time-frequency representations in capturing the complex structure of ECG signals [35].

The SFFS-based feature selection results in this study show that classification performance saturates at approximately 15 features for both classifiers, indicating that increasing feature dimensionality beyond this point does not yield

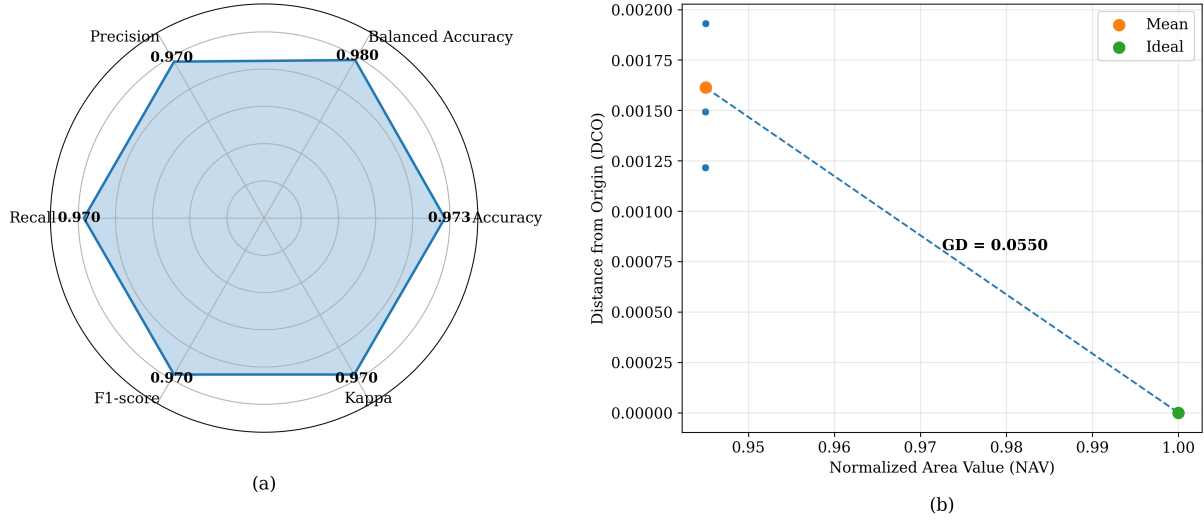


Fig. 9: GD analysis: (a) Irregular hexagonal structure created based on six performance metrics and the location of the center of gravity; (b) Distribution of the distance from the center of gravity to the origin (DCO) with the Normalized Area Value (NAV) and the GD representation according to the ideal performance point.

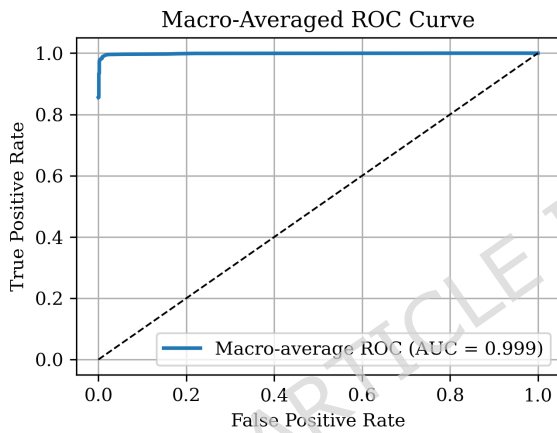


Fig. 10: Macro-averaged ROC curve of the SVM classifier ($C = 1000$, $\gamma = \text{scale}$) using the 15-dimensional Morphological+Temporal+Wavelet feature subset selected by SFFS.

meaningful performance gains and may introduce redundancy. This observation is consistent with previous studies on the MIT-BIH dataset, which report that a compact subset of morphological and temporal features can achieve high accuracy and stable performance [73].

Although feature selection stability across multiple data splits was not explicitly analyzed, the consistent saturation behavior observed for both

k-NN and SVM suggests that the selected feature dimensionality reflects an inherent structure of the feature space rather than being dependent on a specific partitioning of the data.

Furthermore, the results indicate that the majority of discriminative information is concentrated within a relatively small subset of features, supporting the effectiveness of SFFS in identifying informative and non-redundant feature combinations. This compact representation not only reduces computational complexity but also contributes to improved generalization by avoiding overfitting associated with high-dimensional feature spaces. In this context, the observed stability across repeated holdout evaluations further reinforces the reliability of the selected feature subset.

Comparative analysis of the k-NN and SVM classifiers shows that both achieve high and stable performance on the SFFS-selected feature space. For k-NN, the best results are obtained at a low neighborhood size ($k = 3$), while increasing k leads to a gradual decline in total accuracy, macro-level metrics, and Cohen's kappa. This behavior suggests that larger neighborhood sizes introduce oversmoothing in the decision boundaries, thereby reducing the influence of locally discriminative patterns. Consequently, preserving local structure

appears to be critical for maintaining inter-class separability in the proposed feature space.

In contrast, SVM demonstrates consistently high performance across different hyperparameter configurations, with only minor variations in accuracy and macro-level metrics. This behavior suggests that the learned decision boundaries are relatively robust to moderate changes in model parameters. Such stability can be attributed to the compact and well-separated structure of the selected feature space, which enables SVM to effectively exploit global class boundaries.

Although both classifiers achieve comparable performance levels, SVM generally provides slightly higher accuracy and improved stability, particularly when combined with the wavelet-enhanced feature representation. This suggests that, while local neighborhood information is sufficient for effective classification, margin-based methods such as SVM can more effectively utilize the global geometry of the feature space, leading to more stable and reliable decision boundaries.

For SVM, variations in the regularization parameter C and the kernel parameter γ result in only minor changes in performance, indicating that the SFFS-selected feature space provides strong intrinsic discrimination among arrhythmia classes. The consistently high balanced accuracy, macro F1-score, and Cohen's kappa further suggest that the performance is not significantly affected by class imbalance or chance agreement.

Overall, these results demonstrate that the proposed feature extraction and selection strategy produces a compact and well-separated feature space, enabling both distance-based and margin-based classifiers to achieve reliable and stable classification performance.

A key observation is that the confusion matrix obtained using the SFFS-selected wavelet-based feature subset exhibits a strong diagonal structure with minimal off-diagonal errors. This indicates that the selected feature representation effectively preserves inter-class separability.

These findings suggest that, once a compact and well-separated feature space is established, classification performance is largely driven by the discriminative power of the features. Under such conditions, different classical classifiers can achieve comparable decision behavior, although slight differences in performance and stability may still arise.

The holistic evaluation using the GD metric further confirms that the models achieve not only high accuracy but also balanced and stable performance across multiple evaluation criteria. Unlike many studies that focus primarily on overall accuracy or F1-score, the GD framework provides a multidimensional and more comprehensive assessment of classifier performance. These findings demonstrate that classical machine learning approaches remain a strong and interpretable option for arrhythmia classification on the MIT-BIH dataset.

To position the proposed framework within the current literature, recent studies on the MIT-BIH dataset were analyzed, as summarized in Table 10.

Most recent approaches rely on deep learning architectures, including CNNs, attention mechanisms, and hybrid models, achieving accuracy values exceeding 99%. While these models demonstrate superior performance, they typically involve complex architectures, high computational cost, and limited interpretability.

In contrast, classical machine learning approaches, as demonstrated by Singh et al. [80], can achieve competitive accuracy when supported by effective feature design. The proposed framework follows this paradigm by combining morphological, temporal, and wavelet-based features with Sequential Forward Feature Selection (SFFS), resulting in a compact subset of only 15 features. Despite its simplicity, the proposed method achieves competitive performance while maintaining full interpretability and low computational complexity.

Compared to recent deep learning approaches such as hybrid models [78], attention-based architectures [77], and explainable deep learning frameworks [74], the proposed method offers a fundamentally different trade-off. Instead of maximizing accuracy at the expense of complexity, it focuses on achieving a balanced performance across multiple evaluation metrics, as confirmed by the Golden Distance (GD) analysis.

These results highlight that carefully designed feature-based approaches remain a strong and practical alternative for ECG arrhythmia classification, particularly in scenarios where interpretability, efficiency, and reproducibility are critical.

Despite the promising results, several limitations of this study should be acknowledged. First,

Table 10: Comparison of recent ECG arrhythmia classification studies on the MIT-BIH dataset

Study	Method	Type	Acc. (%)	Key Contribution
Kavitha & Shakkeera (2026) [74]	Cycle-CMCHA-GAN + SHAP	DL	99.87	Explainable deep learning framework with SHAP-based interpretability
Cui et al. (2025) [75]	Interpretable KAN	DL	99.08	Interpretable neural architecture with learnable activation functions
Elsheikhy et al. (2025) [76]	Lightweight CNN + Attention	DL	99.18	Multi-lead lightweight architecture for efficient classification
Shah et al. (2025) [77]	Self-attention Autoencoder	DL	99.71	Attention-based representation learning with enhanced preprocessing
Zabihi et al. (2024) [78]	ResNet-LSTM + RF	Hybrid	99.26	Hybrid deep and classical learning framework with feature fusion
Amhia et al. (2024) [79]	SLT + CNN	DL	99.40	Time-frequency analysis using Superlet Transform with deep CNN
Singh et al. (2019) [80]	ANN / SVM / RF	ML	99.59	Comparative machine learning study with handcrafted features
Proposed	SFFS + SVM / k-NN	ML	98.54	Compact, interpretable, and low-complexity feature-based framework

the evaluation was conducted on a single benchmark dataset (MIT-BIH Arrhythmia Database), which may limit the generalizability of the findings to datasets with different acquisition conditions or class distributions. Nevertheless, the proposed feature extraction and SFFS-based selection pipeline is not inherently dataset-specific and can be adapted to other annotated ECG datasets with appropriate preprocessing.

Second, although a repeated holdout strategy with an internal validation split was employed to reduce bias, alternative validation approaches such as k -fold cross-validation could further mitigate dependency on specific data partitions, albeit at the cost of increased computational complexity.

Third, feature extraction and selection were performed in an offline setting, and real-time implementation aspects were not considered. Finally, while classical machine learning models were preferred to emphasize interpretability and methodological transparency, comparisons with deep learning-based approaches were not included and represent an important direction for future research.

5 Future Work

The proposed framework offers several advantages, including interpretability through handcrafted and wavelet-based features, a compact and computationally efficient representation enabled by SFFS, and stable classification performance

across different classifiers and evaluation metrics. However, certain limitations should also be acknowledged. The evaluation was conducted on a single benchmark dataset, and the generalizability of the results to different datasets or real-world clinical settings remains to be further validated. In addition, comparisons with deep learning-based approaches were not included, and real-time implementation aspects were not considered. These limitations motivate several directions for future research.

The findings of this study highlight the critical role of feature extraction and selection in determining classification performance. Future research may focus on enriching the feature space by incorporating additional descriptors such as entropy, fractal dimension, and nonlinear complexity measures to better capture the intrinsic dynamics of ECG signals. Integrating these descriptors with morphological, temporal, and wavelet-based features may further enhance discrimination, particularly for arrhythmia classes with subtle inter-class differences.

A systematic comparison between SFFS and alternative feature selection strategies, including heuristic, evolutionary, and embedded approaches, would provide a more comprehensive understanding of the trade-offs between performance, dimensionality, and computational complexity. In particular, investigating embedded feature selection mechanisms within classifiers may offer deeper

insight into the interaction between feature relevance and decision boundary formation.

In addition, evaluating lightweight deep learning architectures on the SFFS-selected feature space would enable a direct comparison between classical machine learning and deep learning paradigms. Approaches such as CNNs and RNNs have demonstrated promising results in ECG arrhythmia classification; however, they typically require larger datasets and higher computational resources. Future studies could explore the application of such models on diverse ECG datasets to enable automatic feature extraction while maintaining interpretability and reproducibility.

Finally, validating the proposed framework on multiple ECG datasets with varying acquisition conditions and class distributions is essential to assess robustness and generalizability. Incorporating practical considerations such as computational cost, memory requirements, and decision latency will also be important for evaluating the clinical applicability of the approach in real-world deployment scenarios.

6 Conclusion

This study presents a holistic and interpretable framework for multi-class ECG arrhythmia classification, emphasizing the combined impact of signal processing, feature extraction, feature selection, and evaluation strategies on overall performance. Rather than focusing solely on accuracy, the proposed approach demonstrates how methodological choices at each stage influence classification outcomes.

The results show that morphological and temporal features provide strong baseline discrimination, while wavelet-based representations enable a more compact and decorrelated feature space, leading to improved balance and stability with reduced dimensionality. The SFFS-based feature selection further reveals that a small subset of features is sufficient to achieve high and stable performance, indicating that increasing dimensionality does not necessarily improve generalization.

Comparative analysis of k-NN and SVM indicates that, once a well-structured feature space is established, classical classifiers achieve reliable and consistent performance, with SVM providing slightly improved stability. The use of complementary evaluation metrics, including balanced

accuracy, Cohen's kappa, and the Golden Distance (GD), enables a comprehensive assessment beyond overall accuracy.

Overall, the findings demonstrate that classical machine learning approaches, when supported by carefully designed feature engineering and selection, remain a powerful, transparent, and computationally efficient solution for ECG arrhythmia classification. The proposed framework also offers a reproducible and interpretable reference for future studies aiming to balance performance, efficiency, and methodological clarity. Future work will focus on enriching the feature space with additional descriptors and validating the proposed framework across diverse ECG datasets to further assess its robustness and generalizability.

Funding

No funding was received for this study.

Data Availability

The dataset analyzed during the current study is publicly available in the PhysioNet repository. The MIT-BIH Arrhythmia Database can be accessed at <https://physionet.org/content/mitdb/1.0.0/>. No new datasets were generated during this study.

References

- [1] Buckberg, G.D., Nanda, N.C., Nguyen, C., Kocica, M.J.: What is the heart? anatomy, function, pathophysiology, and misconceptions. *Journal of Cardiovascular Development and Disease* **5** (2018) <https://doi.org/10.3390/jcdd5020033>
- [2] Torrent-Guasp, F., Buckberg, G.D., Clemente, C., Cox, J.L., Coghlan, H.C., Gharib, M.: The structure and function of the helical heart and its buttress wrapping. i. the normal macroscopic structure of the heart. *Seminars in Thoracic and Cardiovascular Surgery* **13**, 301–319 (2001) <https://doi.org/10.1053/stcs.2001.29953>
- [3] Anderson, R., Chikwe, J.: The heart and its valves. In: *OSH Cardiac Anaesthesia*. Oxford University Press, 2010, pp. 3–12. <https://doi.org/10.1093/MED/9780199209101.003.01>

- [4] Cleveland Clinic: Heart: Anatomy & Function [Internet]
- [5] Kember, G., Armour, J.A., Zamir, M.: Neural control of heart rate: The role of neuronal networking. *Journal of Theoretical Biology* **277**, 41–47 (2011) <https://doi.org/10.1016/j.jtbi.2011.02.013>
- [6] Cleveland Clinic: Heart Conduction System (Cardiac Conduction) [Internet]
- [7] Bonke, F.I.M., Kirchhof, C.J.H.J., Allessie, M.A., Wit, A.L.: Impulse propagation from the sa-node to the ventricles. *Experientia* **43**, 1044–1049 (1987) <https://doi.org/10.1007/BF01956037/METRICS>
- [8] Kamil, M.R.: Physiology of the heart. In: *Heart Transplantation*, pp. 19–48. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-17311-0_2
- [9] Joukar, S.: A comparative review on heart ion channels, action potentials and electrocardiogram in rodents and human: extrapolation of experimental insights to clinic. *Laboratory Animal Research* **37**, 1–15 (2021) <https://doi.org/10.1186/S42826-021-00102-3/TABLES/5>
- [10] Madona, P., Basti, R.I., Zain, M.M.: Pqrst wave detection on ecg signals. *Gaceta Sanitaria* **35**, 364–369 (2021) <https://doi.org/10.1016/J.GACETA.2021.10.052>
- [11] Cook, D.A., Oh, S.Y., Pusic, M.V.: Accuracy of physicians' electrocardiogram interpretations: A systematic review and meta-analysis. *JAMA Internal Medicine* **180**, 1461–1471 (2020) <https://doi.org/10.1001/jamainternmed.2020.3989>
- [12] Barral, J.-P., Croibier, A.: Circulatory Physiology. In: *Visceral Vascular Manipulations*. Churchill Livingstone, 2011, pp. 27–45. <https://doi.org/10.1016/B978-0-7020-4351-2.00002-8> . <https://www.sciencedirect.com/science/chapter/monograph/abs/pii/B9780702043512000028>
- [13] Chen, T.E., Yang, S.I., Ho, L.T., Tsai, K.H., Chen, Y.H., Chang, Y.F.: S1 and s2 heart sound recognition using deep neural networks. *IEEE Transactions on Biomedical Engineering* **64**, 372–380 (2017) <https://doi.org/10.1109/TBME.2016.2559800>
- [14] Lu, G., Yang, F., Taylor, J.A., Stein, J.F.: A comparison of photoplethysmography and ecg recording to analyse heart rate variability in healthy subjects. *Journal of Medical Engineering & Technology* **33**, 634–641 (2009) <https://doi.org/10.3109/03091900903150998>
- [15] Hnatkova, K., Malik, M.: Sources of qtc variability: Implications for effective ecg monitoring in clinical practice. *Annals of Noninvasive Electrocardiology* **25**, 12730 (2019) <https://doi.org/10.1111/anec.12730>
- [16] Ansari, Y., Mourad, O., Qaraqe, K., Serpedin, E.: Deep learning for ecg arrhythmia detection and classification: an overview of progress for period 2017–2023. *Frontiers in Physiology* **14** (2023) <https://doi.org/10.3389/fphys.2023.1246746>
- [17] Sinha, A., Dutta, U., Demir, O.M., De Silva, K., Ellis, H., Belford, S.: Rethinking false positive exercise electrocardiographic stress tests by assessing coronary microvascular function. *Journal of the American College of Cardiology* **83**, 291–299 (2024) <https://doi.org/10.1016/j.jacc.2023.10.034>
- [18] Bahrami, M., Forouzanfar, M.: Sleep apnea detection from single-lead ecg: A comprehensive analysis of machine learning and deep learning algorithms. *IEEE Transactions on Instrumentation and Measurement* **71** (2022) <https://doi.org/10.1109/TIM.2022.3151947>
- [19] Park, D., So, K., Prabhakar, S.K., Kim, C., Lee, J.J., Sohn, J.H.: Early warning score and feasible complementary approach using artificial intelligence-based bio-signal monitoring system: a review. *Biomedical Engineering Letters* **15**, 717–734 (2025) <https://doi.org/10.1007/s13534-025-00486-4>
- [20] Attachaipanich, T., Attachaipanich, S., Kaewboot, K.: Fragmented qrs complex as a predictor of in-hospital

- life-threatening arrhythmias in myocardial infarction patients: A systematic review and meta-analysis. *Indian Heart Journal* **77**, 142–150 (2025) <https://doi.org/10.1016/j.ihj.2025.04.003>
- [21] Faust, O., Kareem, M., Ali, A., Ciaccio, E.J., Acharya, U.R.: Automated arrhythmia detection based on rr intervals. *Diagnostics* **11** (2021) <https://doi.org/10.3390/diagnostics11081446>
- [22] PhysioNet: MIT-BIH Arrhythmia Database v1.0.0 [Internet]
- [23] PhysioNet: Apnea-ECG Database v1.0.0 [Internet]
- [24] Kaggle: ECG dataset [Internet]
- [25] Melek, M., Melek, N.: Golden distance: A new and comprehensive metric definition study facilitating classification performance evaluations. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering* **49**(3), 1111–1123 (2025) <https://doi.org/10.1007/s40998-025-00870-x>
- [26] Moody, G.B., Mark, R.G.: The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* **20**, 45–50 (2001) <https://doi.org/10.1109/51.932724>
- [27] Park, J., Kang, K.: Pchd: Personalized classification of heartbeat types using a decision tree. *Computers in Biology and Medicine* **54**, 79–88 (2014) <https://doi.org/10.1016/j.compbimed.2014.08.013>
- [28] Mathunjwa, B.M., Lin, Y.T., Lin, C.H., Abbod, M.F., Shieh, J.S.: Ecg arrhythmia classification by using a recurrence plot and convolutional neural network. *Biomedical Signal Processing and Control* **64**, 102262 (2021) <https://doi.org/10.1016/j.bspc.2020.102262>
- [29] Kuila, S., Dhanda, N., Joardar, S.: Ecg signal classification to detect heart arrhythmia using elm and cnn. *Multimedia Tools and Applications* **82**, 29857–29881 (2022) <https://doi.org/10.1007/s11042-022-14233-9>
- [30] Kiranyaz, S., Ince, T., Gabbouj, M.: Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering* **63**(3), 664–675 (2016) <https://doi.org/10.1109/TBME.2015.2468589>
- [31] Nguyen, Q.H., Nguyen, B.P., Nguyen, T.B., Do, T.T.T., Mbinta, J.F., Simpson, C.R.: Stacking segment-based cnn with svm for recognition of atrial fibrillation from single-lead ecg recordings. *Biomedical Signal Processing and Control* **68**, 102672 (2021) <https://doi.org/10.1016/j.bspc.2021.102672>
- [32] Thurner, T., Hintermueller, C., Blessberger, H., Steinwender, C.: Complex-pan-tompkins-wavelets: Cross-channel ecg beat detection and delineation. *Biomedical Signal Processing and Control* **66**, 102450 (2021) <https://doi.org/10.1016/j.bspc.2021.102450>
- [33] Merah, M., Abdelmalik, T.A., Larbi, B.H.: R-peaks detection based on stationary wavelet transform. *Computer Methods and Programs in Biomedicine* **121**, 149–160 (2015) <https://doi.org/10.1016/j.cmpb.2015.06.003>
- [34] Khan, F., Yu, X., Yuan, Z., Rehman, A.: Ecg classification using 1-d convolutional deep residual neural network. *PLOS ONE* **18**, 0284791 (2023) <https://doi.org/10.1371/journal.pone.0284791>
- [35] Pandey, S.K., Shukla, A., Bhatia, S., Gadekallu, T.R., Kumar, A., Mashat, A.: Detection of arrhythmia heartbeats from ecg signal using wavelet transform-based cnn model. *International Journal of Computational Intelligence Systems* **16**, 80 (2023) <https://doi.org/10.1007/s44196-023-00256-z>
- [36] Ebrahimipour Moghaddam Tasouj, P., Soysal, G., Eroğul, O., Yetkin, S.: Ecg signal analysis for detection and diagnosis of post-traumatic stress disorder: Leveraging deep learning and machine learning techniques. *Diagnostics* **15**, 1414 (2025) <https://doi.org/10.3390/diagnostics15111414>

- [37] Fuadah, Y.N., Lim, K.M.: Advances in cardiovascular signal analysis with future directions: A review of machine learning and deep learning models for cardiovascular disease classification based on ecg, pcg, and ppg signals. *Biomedical Engineering Letters* **15**, 619–660 (2025) <https://doi.org/10.1007/s13534-025-00473-9>
- [38] Li, H., Yuan, D., Wang, Y., Cui, D., Cao, L.: Arrhythmia classification based on multi-domain feature extraction for an ecg recognition system. *Sensors* **16** (2016) <https://doi.org/10.3390/s16101744>
- [39] Mar, T., Zaunseder, S., Martínez, J.P., Llamado, M., Poll, R.: Optimization of ecg classification by means of feature selection. *IEEE Transactions on Biomedical Engineering* **58**, 2168–2177 (2011) <https://doi.org/10.1109/TBME.2011.2113395>
- [40] Kaya, Y.: Detection of bundle branch block using higher order statistics and temporal features. *The International Arab Journal of Information Technology* **18**(3), 279–285 (2021) <https://doi.org/10.34028/iajit/18/3/3>
- [41] Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature selection. *Pattern Recognition Letters* **15**, 1119–1125 (1994) [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9)
- [42] Khorrami, H., Moavenian, M.: A comparative study of dwf, cwt and dct transformations in ecg arrhythmias classification. *Expert Systems with Applications* **37**, 5751–5757 (2010) <https://doi.org/10.1016/j.eswa.2010.02.033>
- [43] Gupta, V., Mittal, M., Mittal, V.: A novel frwt based arrhythmia detection in ecg signal using ywara and pca. *Wireless Personal Communications* **124**, 1229–1246 (2021) <https://doi.org/10.1007/s11277-021-09403-1>
- [44] Hassoon, B.A., Xiong, S., Hasson, M.A., Abdulsalami, A.O.: Leveraging conditional diffusion and pruning for optimal ecg signal denoising. *Computers in Biology and Medicine* **193**, 110341 (2025) <https://doi.org/10.1016/j.compbimed.2025.110341>
- [45] Gupta, V., Mittal, M.: R-peak detection in ecg signal using yule-walker and principal component analysis. *IETE Journal of Research* **67**, 921–934 (2021) <https://doi.org/10.1080/03772063.2019.1575292>
- [46] Niroshana, S.M.I., Kuroda, S., Tanaka, K., Chen, W.: Beat-wise segmentation of electrocardiogram using adaptive windowing and deep neural network. *Scientific Reports* **13**, 11039 (2023) <https://doi.org/10.1038/s41598-023-37773-y>
- [47] Remy, J., Velmani, P., Rajakumar, T.C.: A novel approach to perform ecg signal identification and segmentation based on pan-tompkins and hamilton-tompkins algorithm. *International Journal of Pharmaceutical Research* **13**, 5126 (2021) <https://doi.org/10.31838/ijpr/2021.13.01.695>
- [48] Kamata, K., Fujiwara, K., Kinoshita, T., Kano, M.: Missing rri interpolation algorithm based on locally weighted partial least squares for precise heart rate variability analysis. *Sensors* **18**, 13870 (2018) <https://doi.org/10.3390/s18113870>
- [49] Malik, M., Camm, A.J., Bigger, J.T., Breithardt, G., Cerutti, S., Cohen, R.J.: Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation* **93**, 1043–1065 (1996) <https://doi.org/10.1161/01.cir.93.5.1043>
- [50] Elsevier: Analog Electronics. Elsevier, ??? (2002). <https://doi.org/10.1016/B978-0-7506-5095-3.X5000-4>
- [51] Kumar, P., Sharma, V.K.: Detection and classification of ecg noises using decomposition on mixed codebook for quality analysis. *Healthcare Technology Letters* **7**, 18 (2020) <https://doi.org/10.1049/htl.2019.0096>
- [52] Chatterjee, S., Thakur, R.S., Yadav, R.N., Gupta, L., Raghuvanshi, D.K.: Review of noise removal techniques in ecg signals. *IET Signal Processing* **14**, 569–590 (2020) <https://doi.org/10.1049/iet-spr.2020.0104>

- [53] Kumar, M.A., Chakrapani, A.: Classification of ecg signal using fft based improved alexnet classifier. *PLOS ONE* **17**, 0274225 (2022) <https://doi.org/10.1371/journal.pone.0274225>
- [54] Darmawahyuni, A., Nurmaini, S., Tutuko, B., Rachmatullah, M.N., Firdaus, F., Sapitri, A.I.: An improved electrocardiogram arrhythmia classification performance with feature optimization. *BMC Medical Informatics and Decision Making* **24**, 412 (2024) <https://doi.org/10.1186/s12911-024-02822-7>
- [55] Cao, M., Zhao, T., Li, Y., Zhang, W., Benharash, P., Ramezani, R.: Ecg heartbeat classification using deep transfer learning with convolutional neural network and stft technique. *Journal of Physics: Conference Series* **2547**, 012031 (2023) <https://doi.org/10.1088/1742-6596/2547/1/012031>
- [56] Huang, J., Chen, B., Yao, B., He, W.: Ecg arrhythmia classification using stft-based spectrogram and convolutional neural network. *IEEE Access* **7**, 92871–92880 (2019) <https://doi.org/10.1109/ACCESS.2019.2928017>
- [57] Ramezani Moghadam, S., Asl, B.M.: Automatic diagnosis and localization of myocardial infarction using morphological features of ecg signal. *Biomedical Signal Processing and Control* **83**, 104671 (2023) <https://doi.org/10.1016/j.bspc.2023.104671>
- [58] Din, S., Qaraqe, M., Mourad, O., Qaraqe, K., Serpedin, E.: Ecg-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial–temporal and long-range dependency features. *Artificial Intelligence in Medicine* **150**, 102818 (2024) <https://doi.org/10.1016/j.artmed.2024.102818>
- [59] Dey, M., Omar, N., Ullah, M.A.: Temporal feature-based classification into myocardial infarction and other cvds merging cnn and bi-lstm from ecg signal. *IEEE Sensors Journal* **21**, 21688–21695 (2021) <https://doi.org/10.1109/JSEN.2021.3079241>
- [60] Xia, Y., Wang, W., Wang, K.: Ecg signal generation based on conditional generative models. *Biomedical Signal Processing and Control* **82**, 104587 (2023) <https://doi.org/10.1016/j.bspc.2023.104587>
- [61] Park, J., Kimata, C., Young, J., Perry, J.C., Bratincsak, A.: Fine tuning ecg interpretation for young athletes: Ecg screening using z-score-based analysis. *Sports Medicine - Open* **10** (2024) <https://doi.org/10.1186/s40798-024-00775-9>
- [62] Narmadha, G., Deivasigamani, S., Velaisamy, M., Freitas, L.I., Ahmad, B.R., Sakthivel, B.: Detection of human stress using optimized feature selection and classification in ecg signals. *Mathematical Problems in Engineering* **2023**, 3356347 (2023) <https://doi.org/10.1155/2023/3356347>
- [63] Patro, K.K., Jaya Prakash, A., Jayamanmadha Rao, M., Rajesh Kumar, P.: An efficient optimized feature selection with machine learning approach for ecg biometric recognition. *IETE Journal of Research* **68**, 2743–2754 (2022) <https://doi.org/10.1080/03772063.2020.1725663>
- [64] Manshouri, N.: Classification of human vision discrepancy during watching 2d and 3d movies based on eeg signals. *International Journal of Computer Science and Information Security* **15**, 430–436 (2017)
- [65] Saini, I., Singh, D., Khosla, A.: Qrs detection using k-nearest neighbor algorithm (knn) and evaluation on standard ecg databases. *Journal of Advanced Research* **4**, 331–344 (2013) <https://doi.org/10.1016/J.JARE.2012.05.007>
- [66] Aydin, S., Melek, M., Gökrem, L.: A safe and efficient brain–computer interface using moving object trajectories and led-controlled activation. *Micromachines* **16** (2025) <https://doi.org/10.3390/mi16030340>
- [67] Zhang, H., Pan, X., Zhou, S., Zhang, W., Chen, J., Pan, L.: Advanced ecg feature extraction and svm classification for predicting defibrillation success in ohca. *Frontiers in*

- Cardiovascular Medicine **12**, 1550422 (2025) <https://doi.org/10.3389/fcvm.2025.1550422>
- [68] Hu, J., Peng, H., Wang, J., Yu, W.: knn-p: A knn classifier optimized by p systems. Theoretical Computer Science **817**, 55–65 (2020) <https://doi.org/10.1016/j.tcs.2020.01.001>
- [69] Oluchi Anyanwu, G., Nwakanma, C.I., Lee, J.M., Kim, D.S.: Optimization of rbf-svm kernel using grid search algorithm for ddos attack detection in sdn-based vanet. IEEE Internet of Things Journal **10**, 8477–8490 (2023) <https://doi.org/10.1109/JIOT.2022.3199712>
- [70] Melek, M., Manshour, N., Kayikcioglu, T.: Low-cost brain-computer interface using the emotiv epoc headset based on rotating vanes. Traitement du Signal **37**, 831–837 (2020) <https://doi.org/10.18280/ts.370516>
- [71] DeVries, Z., Locke, E., Hoda, M., Moravek, D., Phan, K., Stratton, A.: Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and fl-score for the assessment of prognostic capability. Spine Journal **21**, 1135–1142 (2021) <https://doi.org/10.1016/j.spinee.2021.02.007>
- [72] Ojha, M.K., Wadhwani, S., Wadhwani, A.K., Shukla, A.: Automatic detection of arrhythmias from an ecg signal using an auto-encoder and svm classifier. Physical and Engineering Sciences in Medicine **45**, 665–674 (2022) <https://doi.org/10.1007/s13246-022-01119-1>
- [73] Chaubey, K., Saha, S.: Electrocardiogram morphological arrhythmia classification using fuzzy entropy-based feature selection and optimal classifier. Biomedical Physics & Engineering Express **9** (2023) <https://doi.org/10.1088/2057-1976/acf222>
- [74] Kavitha, P., Shakkeera, L.: An explainable deep learning framework for accurate and automated cardiac arrhythmia classification using electrocardiogram signals. Biomedical Signal Processing and Control **117**, 109590 (2026) <https://doi.org/10.1016/j.bspc.2026.109590>
- [75] Cui, H., Ning, S., Wang, S., Zhang, W., Peng, Y.: Ecg signal classification using interpretable kan: Towards predictive diagnosis of arrhythmias. Algorithms **18**(2), 90 (2025) <https://doi.org/10.3390/a18020090>
- [76] Elsheikhy, D.H., Hassan, A.S., Yhlea, N.M., Fareed, A.M., Rashed, E.A.: Lightweight deep learning architecture for multi-lead ecg arrhythmia detection. Sensors **25**(17), 5542 (2025) <https://doi.org/10.3390/s25175542>
- [77] Shah, A., Singh, D., Mohamed, H.G., Bhargava, S., Rehman, A.U., Hussien, S.: Electrocardiogram analysis for cardiac arrhythmia classification and prediction through self attention based auto encoder. Scientific Reports **15**, 9230 (2025) <https://doi.org/10.1038/s41598-025-93906-5>
- [78] Zabihi, F., Safara, F., Ahadzadeh, B.: An electrocardiogram signal classification using a hybrid machine learning and deep learning approach. Healthcare Analytics **6**, 100366 (2024) <https://doi.org/10.1016/j.health.2024.100366>
- [79] Amhiah, H., Sarkar, V., Pandey, L., Shakuntala, M., Anudeep, S., Singh, S.: Advanced algorithm for cardiac arrhythmia classification and efficient detection of abnormalities in ecg signals. In: 2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG), pp. 1–6 (2024). <https://doi.org/10.1109/ICTBIG64922.2024.10911394>
- [80] Singh, V., Tewary, S., Sardana, V., Sardana, H.K.: Arrhythmia detection: A machine learning based comparative analysis with mit-bih ecg data. In: 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1–5 (2019). <https://doi.org/10.1109/I2CT45611.2019.9033665>