Vol. 8, No. 2, April 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Machine Learning Classification of SCD, CHF, and NSR Using 15-Minute ECG-Derived HRV Features

Febriyanti Panjaitan[1]*, Win Ce[2], M. Fajar Ramadhan[3], Winarnie[4], Hery Oktafiandi[5]

[1,3,4]Department of Computer Science, Satu University, Bandung, Indonesia
[2]Information System Department, Bina Nusantara University, Bandung, Indonesia
[5]Information System Department, Satu University, Bandung, Indonesia

Received:
October 3, 2025
Revised:
March 11, 2026
Accepted:
April 11, 2026
Published:
April 26, 2026

Corresponding Author:
Author Name*:
Febriyanti Panjaitan
Email*:
pebrianti.panjaitan@univ.satu.ac.id

DOI:
10.63158/journalisi.v8i2.1557

Abstract. Heart disease remains one of the leading causes of mortality worldwide, making early detection essential for effective intervention. Heart Rate Variability (HRV) is widely used as a non-invasive marker for assessing cardiac conditions, and machine learning has shown potential in classifying heart diseases such as Sudden Cardiac Death (SCD) and Congestive Heart Failure (CHF). This study evaluates the performance of Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighbors (KNN) using 15-minute ECG signals comprising three 5-minute segments. The dataset consists of 53 subjects, generating 159 segments, including SCD, CHF, and Normal Sinus Rhythm (NSR). To prevent data leakage, a subject-wise split (80:20) is applied for training and testing. Two evaluation scenarios are considered: per-segment classification and combined 15-minute classification. Results indicate that SVM and DT achieve consistently high, stable performance with near-perfect accuracy, precision, recall, and F1-score, whereas KNN shows lower, more variable performance, particularly in segment-based analysis. The combined 15-minute approach provides more stable results, suggesting improved HRV representation and class separability. Although the results are promising, further validation with larger, more diverse datasets is required to ensure robustness and generalizability. This study highlights the potential of HRV-based machine learning while emphasizing the importance of appropriate temporal representation and rigorous evaluation design.

Keywords: Heart Rate Variability, ECG classification, Sudden Cardiac Death, Congestive Heart Failure, Machine Learning

1. INTRODUCTION

Heart disease is one of the leading causes of death worldwide [1]. Two types of heart conditions that are frequently the focus of research are Sudden Cardiac Death (SCD) and Congestive Heart Failure (CHF) [2]. SCD is a fatal condition that occurs suddenly due to abnormal heart rhythms, while CHF is characterized by the heart's inability to pump blood efficiently [3]. In addition, normal heart conditions are compared to assess differences in cardiac function characteristics [4].

The electrocardiogram (ECG) is an electrical signal that represents heart activity through various waves, such as the P wave, QRS complex, and T wave [5] (Figure 1). The QRS complex, which reflects ventricular depolarization, is one of the main components of the ECG signal. The interval between R wave peaks, known as the RR interval, can be analyzed to evaluate heart rate variability (HRV) [6]. One of the crucial pieces of information that can be extracted from the ECG is Heart Rate Variability (HRV), which refers to the variability in the intervals between heartbeats [7]. HRV reflects the balance between the sympathetic and parasympathetic nervous systems and can be used to identify various physiological and pathological conditions, including heart disease [8].

This study uses eight time-domain HRV features to describe the characteristics of heart rate variability, providing an in-depth view of heart conditions [9]. According to previous studies, time-domain HRV features have been proven effective in distinguishing between healthy heart conditions and various heart diseases, such as Sudden Cardiac Death (SCD) and Congestive Heart Failure (CHF). Several studies, such as those conducted by Malik et al. (1996) [10], have shown that features such as Mean RR Interval, Standard Deviation of NN Interval (SDNN), Root Mean Square of Successive Differences (RMSSD), Number of successive RR intervals that differ by more than 50 ms (NN50), Percentage of Successive RR Intervals that Differ by More than 50 ms (pNN50), Coefficient of Variation of RR intervals (CVRR), Minimum of RR intervals (Min-RR), and Maximum of RR intervals (Max-RR) can provide strong indications of the balance between the sympathetic and parasympathetic nervous systems, which play a critical role in heart conditions

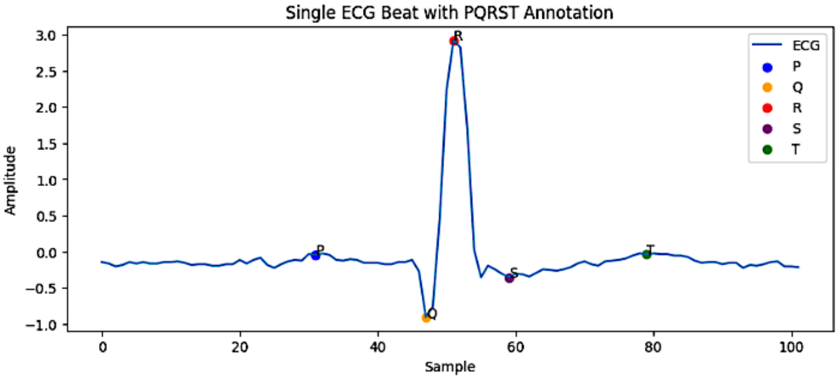


Figure 1. Main Features of the ECG Waveform

Previous studies have shown that HRV analysis using machine learning methods can yield significant results in classifying heart conditions [9]. Several studies have utilized data from PhysioNet MIT-BIH [11] to classify heart conditions such as SCD, CHF, and normal sinus rhythm (NSR) [12]. In a previous study using Support Vector Machine (SVM), an accuracy of up to 87% was achieved [13], while Decision Tree (DT) showed an accuracy of around 73% [14] in classifying heart disease. However, this performance can be improved through proper hyperparameter optimization.

A previous study by Febriyanti et al. [15] employed Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighbors (KNN), achieving an average classification accuracy of over 98% using 5-minute ECG signal segments across multiple cardiac conditions. However, existing studies predominantly rely on short-duration ECG signals and do not systematically examine how temporal representation influences HRV feature stability, class separability, and model generalizability. In addition, limited attention has been given to comparing segment-based classification with combined long-duration approaches, and the transferability of optimized parameters across different signal durations remains unclear. To address these gaps, this study proposes a machine learning-based framework using 15-minute ECG signals, consisting of three 5-minute segments, while adopting optimized parameters from prior work. The use of a longer duration is motivated by the need to capture more comprehensive HRV dynamics, balancing physiological relevance and computational efficiency, and providing a more stable representation of cardiac activity by reducing the influence of transient fluctuations [16]. Importantly, this study explicitly compares segment-based and combined 15-minute classification to investigate how temporal segmentation affects feature stability, robustness, and classification performance. While short segments may capture localized physiological variations, they can introduce variability and reduce consistency, whereas longer combined signals integrate temporal information, potentially improving noise robustness and generalization across subjects. Therefore, this comparison is scientifically significant in determining the most appropriate temporal representation for HRV-based classification. This study focuses on SCD and CHF, along with NSR, to reduce model complexity and mitigate overfitting while maintaining clinical relevance, as both conditions are associated with a high risk of sudden cardiac events [2]. The main contributions of this study are: (1) evaluating the impact of 15-minute ECG duration on HRV-based classification performance, (2) comparing segment-based and combined classification strategies, and (3) assessing the effectiveness of SVM, DT, and KNN models under a subject-wise evaluation framework to provide more reliable and generalizable results.

2. METHODS

The proposed framework consists of sequential steps: (1) ECG data acquisition from PhysioNet, (2) signal preprocessing including denoising and R-peak detection, (3) segmentation into 5-minute segments, (4) HRV feature extraction, (5) dataset construction, (6) subject-wise data splitting, (7) classification using SVM, DT, and KNN, and (8) performance evaluation using confusion matrix metrics. A structured workflow of the proposed methodology is illustrated in Figure 2.

2.1. Dataset

The ECG dataset used in this study was obtained from the PhysioNet MIT-BIH database, consisting of three categories: Sudden Cardiac Death (SCD) [17], Congestive Heart Failure (CHF) [18], and Normal Sinus Rhythm (NSR) [19]. The original ECG file format obtained from PhysioNet is ".DAT," which stores ECG recordings in digital form collected from medical devices. This file contains raw signal data representing the heart's electrical activity and is typically accompanied by a metadata file, such as ".hea" (header file), which contains information about the recording parameters [20].

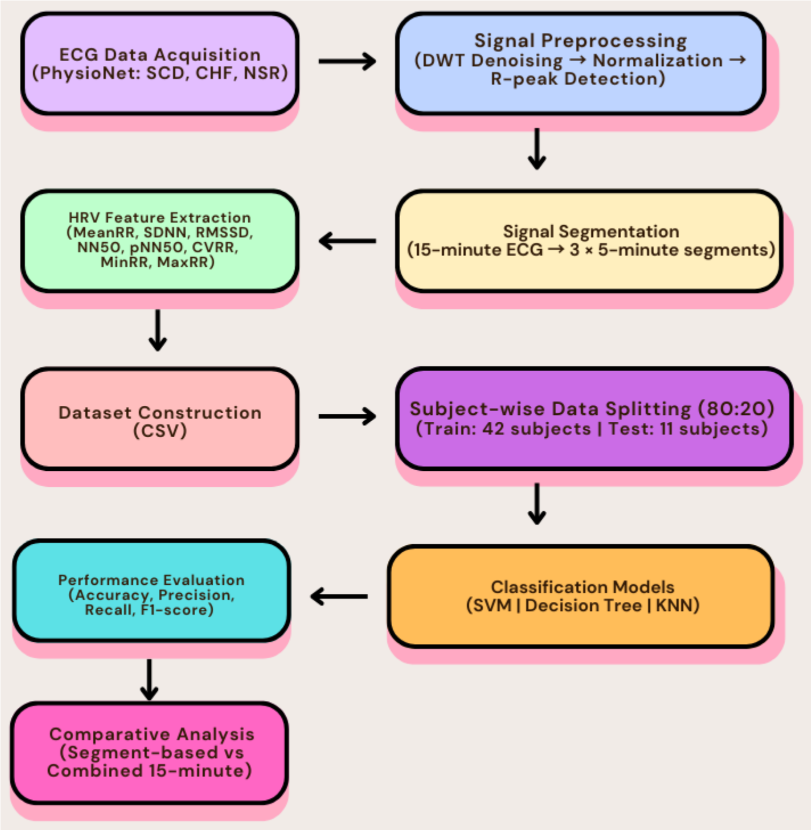


Figure 2. Proposed Workflow for HRV-Based Heart Disease Classification

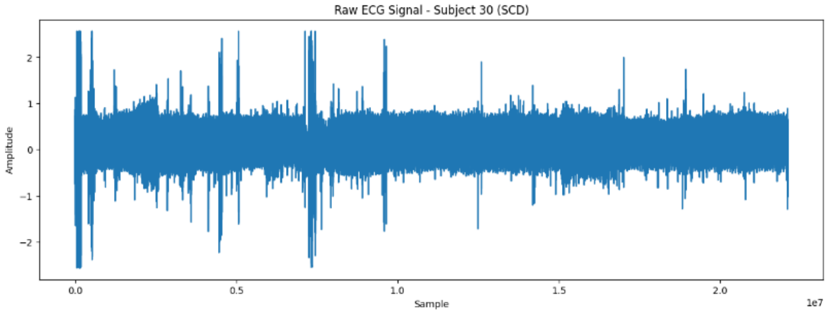


Figure 3. Raw ECG Data of an SCD Subject

2.2. Preprocessing

Before the ECG signals are extracted into HRV features, a preprocessing stage is carried out. This includes trimming ECG recordings longer than one hour down to 15 minutes and then further dividing them into 5-minute segments. This segmentation is based on recommendations from previous studies, which suggest that optimal HRV analysis can be performed on 5-minute segments to enhance accuracy and data processing efficiency

(Electrophysiology, 1996) [21]. Each 5-minute segment is then analyzed to identify and remove any noise that may be present in the ECG signal

The QRS complex in the ECG is the primary element used to obtain RR intervals. These intervals must be detected through R-peaks within the ECG signal duration, as these peaks contain highly relevant information for categorizing heart diseases [28]. Figure 3 illustrates the process of obtaining RR intervals, which are then extracted into HRV features. The first step involves removing noise from the ECG signal using Discrete Wavelet Transform (DWT). The second step is signal normalization using the Normalized Absolute Difference (NADiff). The third step is detecting R-peaks using the Hamilton-Tompkins method. Previous studies have applied these steps using 5-minute signal durations [15].

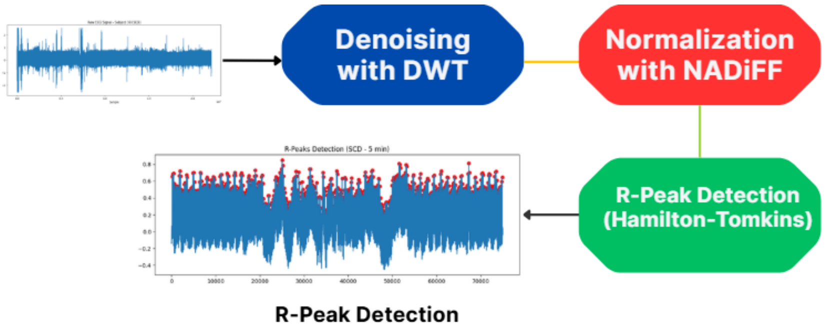


Figure 3. Steps to Obtain RR Intervals

Each subject initially provides a long-duration ECG recording (approximately 24 h 33 min 27 s), as shown in Figure 2. For analysis purposes, a 15-minute segment is extracted from each recording. R-peak detection is then performed on this 15-minute ECG signal following the steps illustrated in Figure 3 to obtain accurate RR intervals. The detected R-peaks are highlighted in red, as shown in Figure 4. After the R-peak detection process, the ECG signal is segmented into three equal segments of 5 minutes each, which are subsequently used for HRV feature extraction. Consequently, each subject contributes three ECG segments for analysis. The detailed distribution of segments by category and subject is presented in Table 1.

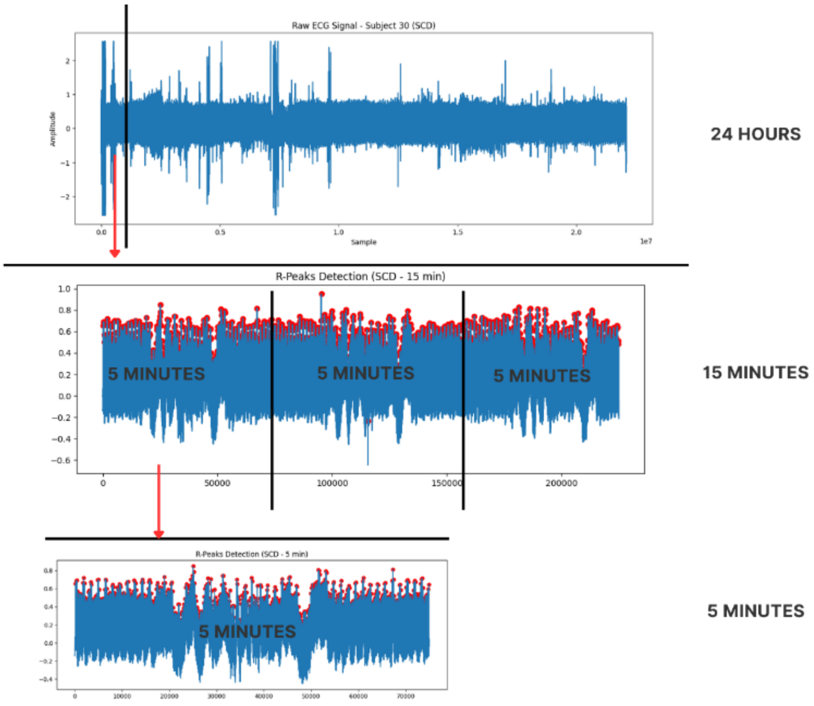


Figure 4. Illustration of ECG Segmentation

Table 1. Segment Distribution

Category	Number of Subject	Subject Duration	Segment per Subject	Total Segments
SCD	20	15 minutes	3	60
CHF	15	15 minutes	3	45
NSR	18	15 minutes	3	54
Total	**53**			**159**

2.3. Feature Extraction

After preprocessing the ECG signal and calculating the RR intervals, the extracted HRV feature data can be stored in CSV format. This process begins once the RR interval time series is obtained and converted into a tabular format. Each row in the table represents one RR interval with eight HRV feature attributes: MeanRR, SDNN, RMSSD, NN50, pNN50, CVRR, Min-RR, and Max-RR. This data is then saved as a CSV file to facilitate further analysis, particularly using machine learning methods [22].

HRV features are extracted from each 5-minute segment, focusing on time-domain features that have been proven relevant in classifying heart conditions [23]. The eight extracted features are:

1) Mean of RR Intervals (MeanRR): Measures the average time interval between two R-peaks in the ECG signal.
2) Standard Deviation of NN Interval (SDNN): This measure measures the overall variability of RR intervals, indicating autonomic balance.
3) Root Mean Square of Successive Differences (RMSSD): Measures short-term fluctuations in RR intervals.
4) Number of Successive RR Intervals that Differ by More Than 50 ms (NN50): Indicates the number of significant changes in RR intervals.
5) Percentage of Successive RR Intervals that Differ by More Than 50 ms (pNN50): The percentage of significant changes in RR intervals.
6) Coefficient of Variation of RR Intervals (CVRR): Measures the consistency of RR intervals within a segment.
7) Minimum of RR Intervals (Min-RR): The shortest RR interval value in the segment.
8) Maximum of RR Intervals (Max-RR): The longest RR interval value in the segment.

After these features are calculated, the data is organized into a table format, with columns representing the HRV features and rows representing individual subjects or measurements.

2.4. Classification

After extracting HRV features, the next step is to classify using three machine learning methods: SVM, DT, and KNN. These methods were selected due to their proven effectiveness in handling high-dimensional data and their ability to provide reliable classification performance across various medical datasets. The parameters used in this study are adopted from previous work [24], where they were optimized using a grid search strategy and demonstrated high classification performance (>98%). Rather than re-optimizing the parameters, this study investigates the generalizability and robustness of these previously optimized configurations when applied to longer ECG signal durations (15 minutes), compared to the 5-minute segments used in prior studies.

This study intentionally avoids re-optimizing hyperparameters to evaluate their transferability and robustness across different data conditions. Unlike prior work that used shorter ECG segments, this study employs longer ECG recordings (15 minutes) and a different class composition (SCD, CHF, and NSR), introducing a domain shift. Therefore, the objective is not to achieve dataset-specific optimal performance, but to assess whether previously optimized configurations remain stable and effective under temporal and distributional variations. This approach provides insight into the generalizability of machine learning models without repeated parameter tuning. Nevertheless, future work will include a comparative analysis with re-optimized parameters to validate performance differences further.

This approach introduces a novel perspective by evaluating the transferability of hyperparameters across different temporal settings of ECG data, specifically from shorter segments to longer-duration signals. This is important to determine whether parameter configurations remain effective under varying data conditions without requiring repeated optimization, while also reducing computational cost. Table 2 presents the parameter configurations adopted from previous research and applied in this study. These configurations are expected to maintain high performance while providing insights into the stability of machine learning models when applied to longer ECG signal durations for heart disease classification.

Table 2. Best Parameters

Methods	Parameter
SVM	C': 1, 'degree': 2, 'gamma': 'scale', 'kernel': 'linear'
DT	'criterion': 'gini', 'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 5
KNN	'metric': 'manhattan', 'n_neighbors': 3, 'weights': 'distance'

This study also evaluates two different classification approaches in processing ECG data to compare the effectiveness of long-duration versus shorter-segmented data. The first approach is segment-based classification, where the ECG signal is divided into three 5-

minute segments (first, second, and third), and classification is performed independently for each segment. This approach allows observation of how HRV dynamics change over time and whether signal segmentation offers advantages in distinguishing different heart conditions. The second approach involves combining the three segments into a single unit, representing a 15-minute ECG signal, and performing classification based on this longer data. By comparing these two approaches, the study aims to determine whether segment-based classification is more effective in distinguishing SCD, CHF, and normal conditions or if the approach using a more extended duration provides more accurate results in heart disease classification.

In this study, the data is split using an 80:20 ratio for training and testing. To prevent data leakage, the splitting is performed at the subject level (subject-wise split) rather than at the segment level [25]. A total of 53 subjects are divided into 42 subjects (80%) for training and 11 subjects (20%) for testing. The data distribution is maintained proportionally across each category. For the SCD category (20 subjects), 16 subjects are used for training (48 segments) and 4 subjects for testing (12 segments). For the CHF category (15 subjects), 12 subjects are used for training (36 segments) and 3 subjects for testing (9 segments). For the NSR category (18 subjects), 14 subjects are used for training (42 segments) and 4 subjects for testing (12 segments).

All segments belonging to a single subject are assigned exclusively to either the training or testing set to ensure independence between datasets. This strategy prevents the model from learning similar patterns from the same subject across both datasets, thereby providing a more reliable evaluation. The 80:20 subject-wise data split ensures that the model does not experience overfitting and can generalize well to unseen data. Tables 3 and 4 present the data distribution for each condition (SCD, CHF, NSR) under both approaches.

Table 3. Training and Testing Data (Combined Segments Approach)

Subject	Number of Subject	Train Subjects	Test Subjects	Train Segments	Test Segments
SCD	20	16	4	48	12
CHF	15	12	3	36	9

Subject	**Number of Subject**	**Train Subjects**	**Test Subjects**	**Train Segments**	**Test Segments**
NSR	18	14	4	42	12
Total	53	42	11	126	33

Table 4. Training and Testing Data (Segment-Based Approach)

Subject	**Training**			**Testing**		
	Seg 1	**Seg 2**	**Seg 3**	**Seg 1**	**Seg 2**	**Seg 3**
SCD	16	16	16	4	4	4
CHF	12	12	12	3	3	3
NSR	14	14	14	4	4	4

2.5. Model Evaluation

After the classification process is completed, evaluation is carried out using a confusion matrix to measure the model's performance [26]. The confusion matrix provides detailed information about the model's predictions, including the number of:

1) True Positives (TP): Correct predictions for a specific category (e.g., an SCD subject correctly classified as SCD).
2) True Negatives (TN): Correct predictions outside a specific category (e.g., a non-SCD subject correctly not classified as SCD).
3) False Positives (FP): Incorrect predictions for a specific category (e.g., a non-SCD subject incorrectly classified as SCD).
4) False Negatives (FN): Incorrect predictions due to failure to classify a subject into the correct category (e.g., an SCD subject incorrectly classified as another category).

The confusion matrix is used for each category, SCD, CHF, and Normal, providing an overview of the classification performance for each condition. From this confusion matrix data, several evaluation metrics are calculated to assess the overall performance of the model:

1) Accuracy: Measures the percentage of correct predictions out of the total number of predictions. The calculation as shown in Equation 1.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

2) Precision: Measures the accuracy of the model's predictions for a specific category (the proportion of correct predictions out of all predictions made for that category). The calculation as shown in Equation 2.

$$\text{Precision}=\frac{TP}{TP+FP} \tag{2}$$

3) Recall (Sensitivity): Measures the model's ability to capture all data in a specific category (the proportion of correct predictions out of the total actual instances for that category). The calculation as shown in Equation 3.

$$\text{Recall}=\frac{TP}{TP+FN} \tag{3}$$

4) F1-Score: Combines Precision and Recall to provide a balanced measure of performance, especially for imbalanced datasets. The calculation as shown in Equation 4.

$$\text{F1score}=\frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{4}$$

This evaluation provides a deeper understanding of the model's strengths and weaknesses in classifying heart conditions. Using these metrics, it can be determined whether the model tends to overfit, is biased toward a particular category, or demonstrates consistent performance across all categories.

To ensure reproducibility, all experiments used a fixed random seed (random_state = 42) consistently across data splitting and model initialization. Additionally, each experiment was repeated 10 times, and the reported results represent the average performance to reduce the effect of randomness. The implementation was carried out in Python using libraries such as NumPy, Pandas, Scikit-learn, WFDB, and NeuroKit2. All preprocessing, feature extraction, and classification steps were executed under consistent experimental settings to ensure reliable and reproducible results.

3. RESULTS AND DISCUSSION

This study uses ECG-based HRV analysis to distinguish between normal heart conditions, SCD, and CHF. This section presents the results and discussion based on the methodology previously described.

3.1. Preprocessing Data

After preprocessing, the ECG signals improved quality with reduced noise and more accurate R-peak detection. Visualizing the preprocessing results clearly illustrates the difference between the raw and processed signals. Figure 4 presents the ECG signal visualization from SCD, CHF, and normal condition subjects, each with a 5-minute duration. The 5-minute duration was selected to ensure clear and detailed ECG signal visualization.

The amplitude in the ECG signal in Figure 5 represents the strength or magnitude of the electrical activity generated by the heart during its cycle. In ECG signal visualization, amplitude reflects the voltage variation (typically in millivolts) measured on the skin surface. This value corresponds to the electrical activity during phases such as depolarization and repolarization of the heart.

Negative amplitudes indicate certain phases of electrical activity, such as repolarization or signals falling below the baseline (zero line). In contrast, positive amplitudes represent depolarization, such as the R-wave peak in the QRS complex, which signifies ventricular contraction [27], [28]. The total number of time values (ms) is 75,000, based on the calculation that each second in an SCD ECG signal is represented by 250 ms [17], multiplied by the total duration of 5 minutes. The same calculation method is applied to CHF and normal condition subjects, with each subject's time representation. CHF subjects have a time resolution of 250 ms per second [18], while normal condition (NSR) subjects are represented with 128 ms per second [19]. Red 'x' markers indicate the R-peaks in the ECG signal.

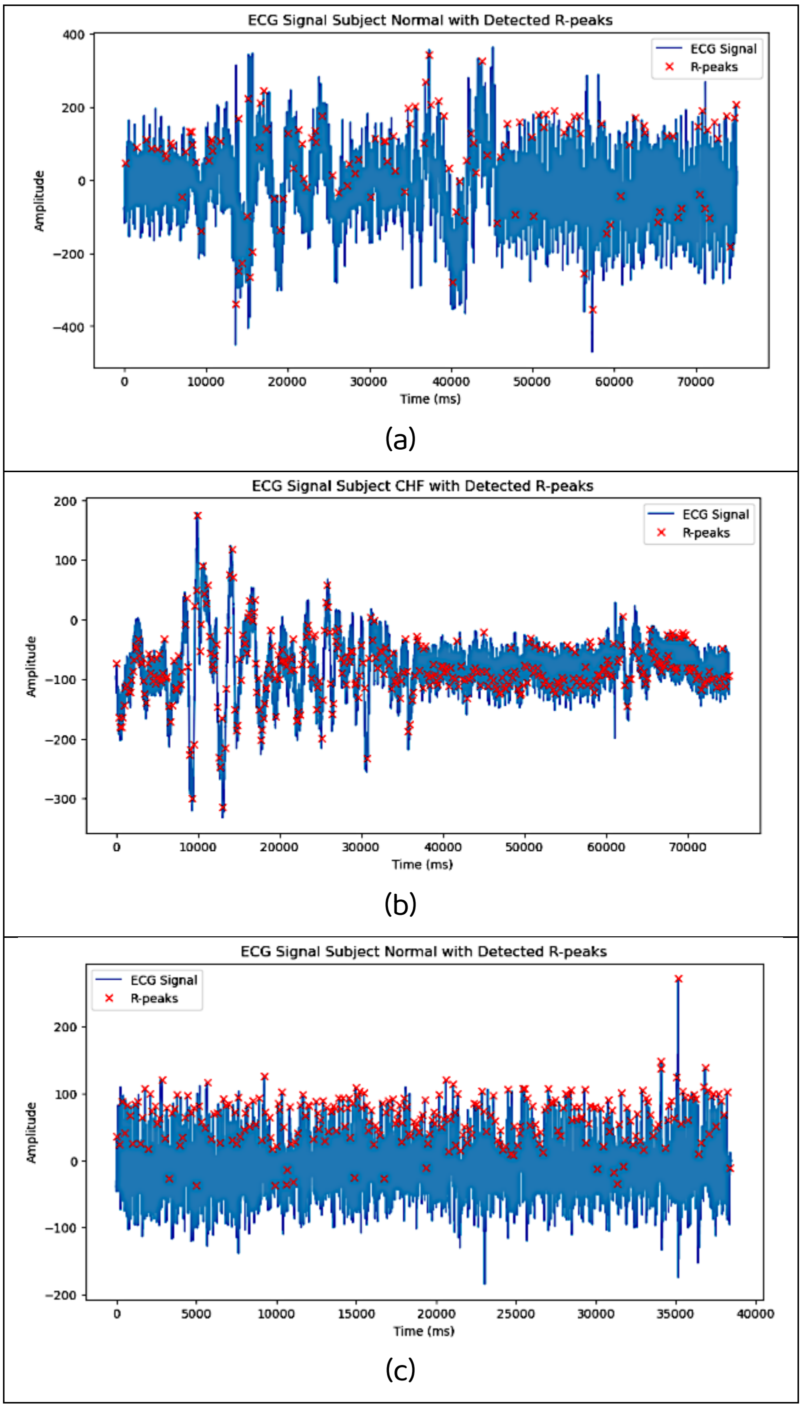


Figure 5. ECG Signals After Preprocessing: (a) SCD, (b) CHF, (c) Normal Subject (NSR)

3.2. HRV Feature Extraction

Each subject has a total of three segments with an overall signal duration of 15 minutes, which has been divided into three 5-minute segments. The data presented in Table 5 was specifically selected to highlight the results from all segments across the three subject groups.

Table 5. HRV Feature Extraction

Subject	Segment	MeanRR	SDNN	NN50	RMSSD	PNN50	CVRR	MinRR	MaxRR
SCD	1	0.736	0.482	0.919	0.632	0.009	0.655	0.007	3.578
CHF	1	0.787	0.425	0.928	0.581	0.009	0.540	0.007	2.617
NSR	1	0.769	0.315	0.586	0.442	0.005	0.409	0.412	2.078
SCD	2	0.765	0.465	0.924	0.622	0.009	0.608	0.007	2.914
CHF	2	0.885	0.601	0.918	0.786	0.009	0.679	0.007	0.796
NSR	2	0.979	0.370	0.419	0.519	0.004	0.378	0.578	2.609
SCD	3	0.533	0.368	0.929	0.571	0.009	0.691	0.007	0.484
CHF	3	0.936	0.757	0.937	0.950	0.009	0.808	0.210	11.843
NSR	3	0.894	0.383	0.496	0.527	0.004	0.428	0.179	0.718

The background section explains the selection of subjects with SCD and CHF conditions, as both conditions are known to have significant clinical impacts and pose a high risk of sudden cardiac death. The first segment represents results from a previous study [29]. In contrast, the second and third segments reflect the HRV dynamics in the later phases of the ECG signal, thus complementing the insights obtained from the first segment. By presenting all three segments, this study expands the scope of HRV analysis to provide deeper insights. This analysis reinforces findings from earlier research and opens opportunities to explore dynamic changes in HRV signals across segments for each subject. Including the second and third segments offers a more comprehensive view of how heart conditions are reflected through heart rate variability over a longer duration.

The data structure from this analysis is organized in CSV format, with the columns and rows arranged as follows:

1) Columns: Represent HRV features such as Mean RR, SDNN, NN50, RMSSD, pNN50, CVRR, min-RR, and max-RR.
2) Rows: Represent different subjects and segments (second and third segments).

With this approach, each segment provides specific and detailed information about heart rate variability over a shorter duration, offering a more comprehensive view of each subject's heart condition

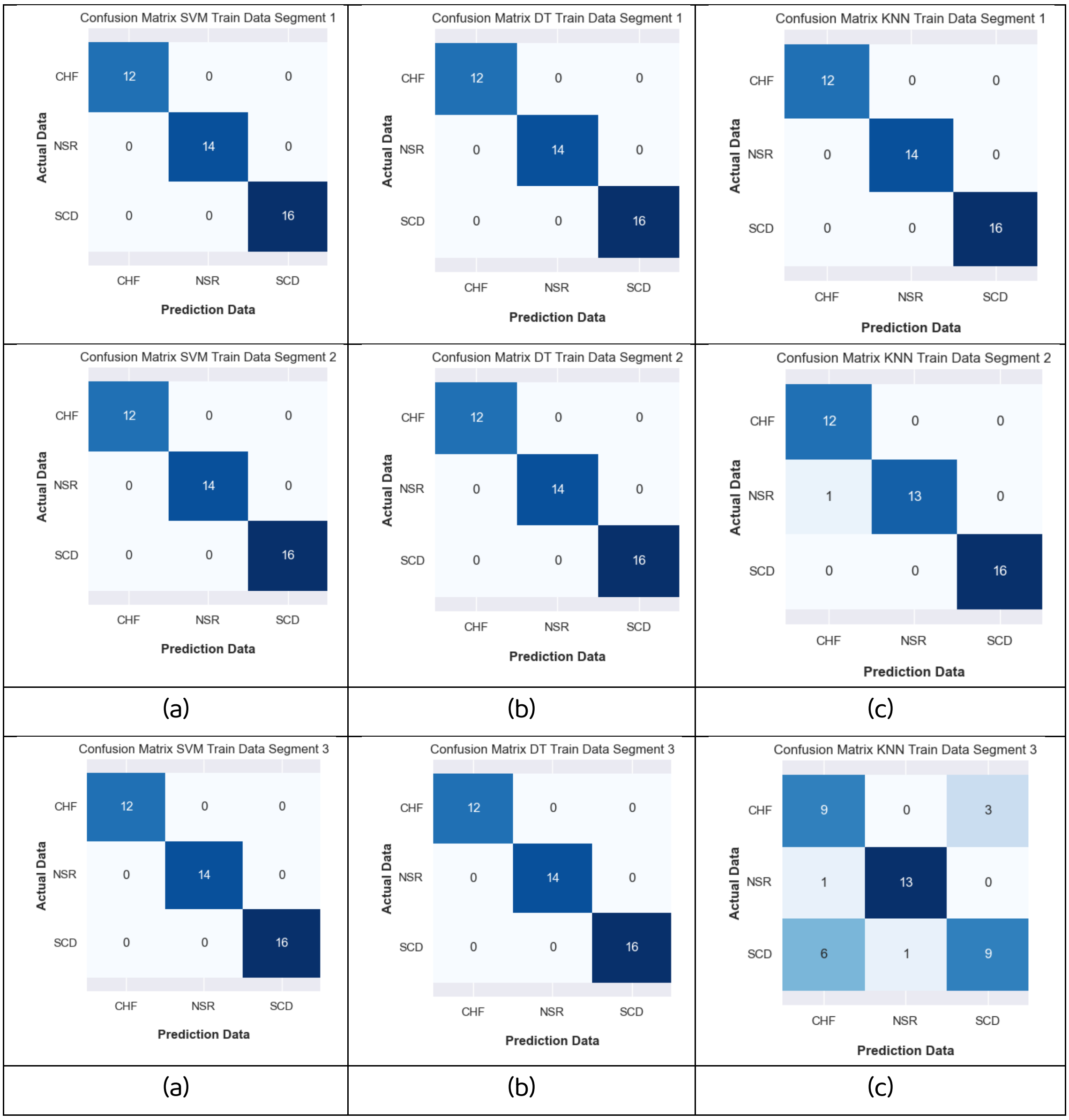
3.3. Classification Evaluation

These classification results are based on evaluation using the confusion matrix, which produces values for accuracy, precision, recall, and F1-score. The results include both per-segment analysis and the overall combined segment data, allowing for a comparison of differences and accuracy in classification.

Figure 6 presents the prediction results for segments 1, 2, and 3 on the training data using the SVM, DT, and KNN methods. In general, the per-segment analysis shows that SVM and DT consistently achieved optimal classification performance across all segments, indicating their robustness in handling variations in HRV features within shorter time windows. In contrast, KNN exhibited performance variability across segments. In segment 2, a minor misclassification was observed in the NSR class, where one instance was incorrectly classified as CHF. This error may be attributed to the similarity of HRV feature patterns between NSR and CHF within the limited time window, combined with the nature of KNN, which relies heavily on distance-based similarity. When class distributions overlap or are not clearly separable, KNN becomes more sensitive to boundary conditions.

The performance degradation of KNN becomes more evident in segment 3, where multiple misclassifications occurred across all classes. Several CHF instances were incorrectly classified as SCD, and a portion of SCD data was misclassified as CHF and NSR. These results indicate that the feature space in segment 3 is more complex and less separable, making it more challenging for distance-based methods such as KNN to perform effectively.

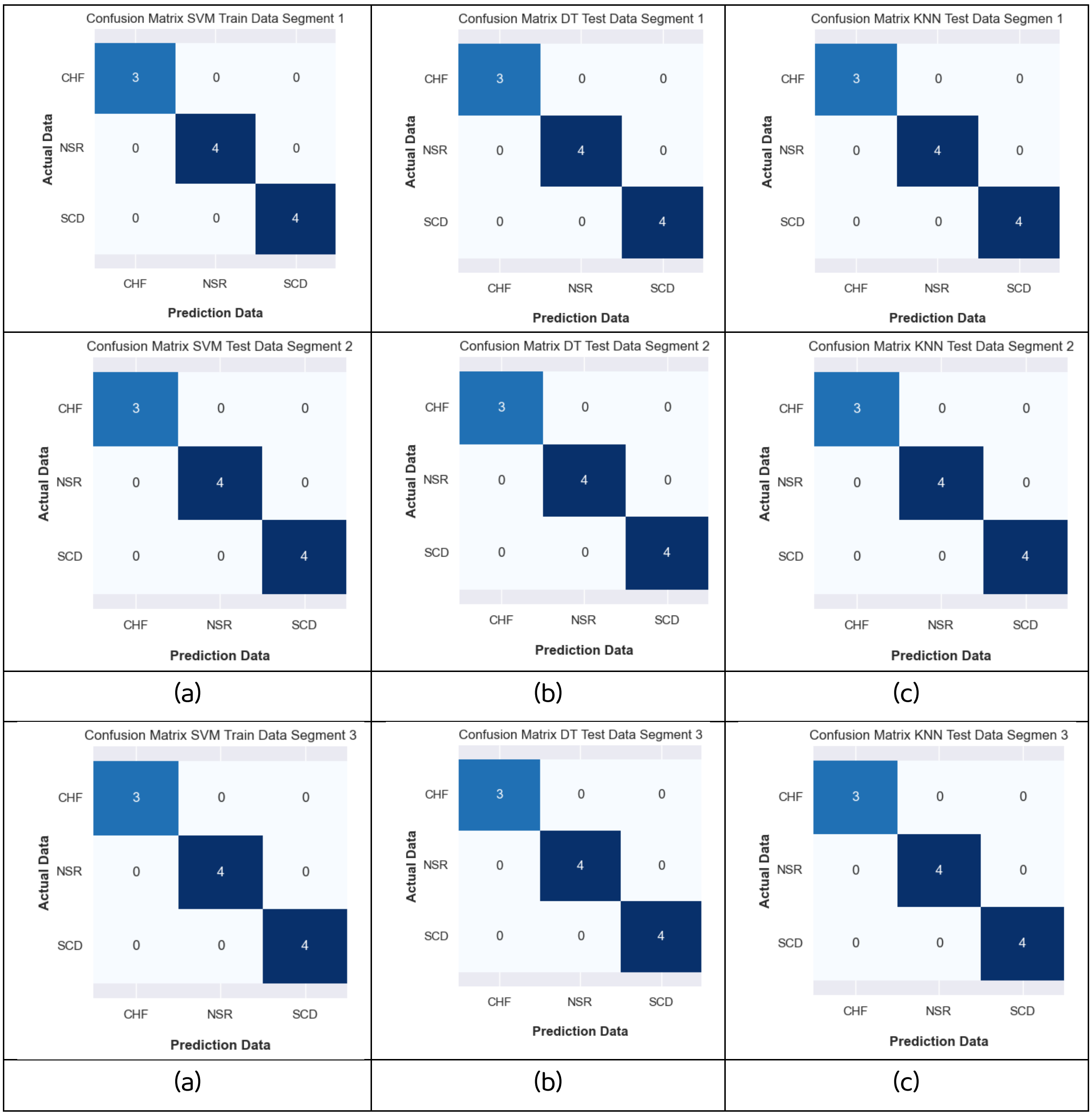
Compared to the per-segment approach, the combined 15-minute analysis provides a more stable and comprehensive representation of HRV characteristics. By integrating information from all segments, the overall feature distribution becomes more consistent, reducing the impact of local variability observed in individual segments. As a result, classification performance across all methods becomes more stable, with fewer misclassifications observed compared to the per-segment analysis. This comparison suggests that while per-segment analysis is useful for capturing short-term variations in HRV, it may introduce instability in classification performance due to fluctuating feature distributions. In contrast, the 15-minute combined approach offers better separability between classes by incorporating more complete physiological information, leading to more reliable classification outcomes.



(a) (b) (c)

(a) (b) (c)

Figure 6. Confusion Matrix of Classification Evaluation Results for Training Data Using Three Methods: (a) SVM, (b) DT, (c) KNN

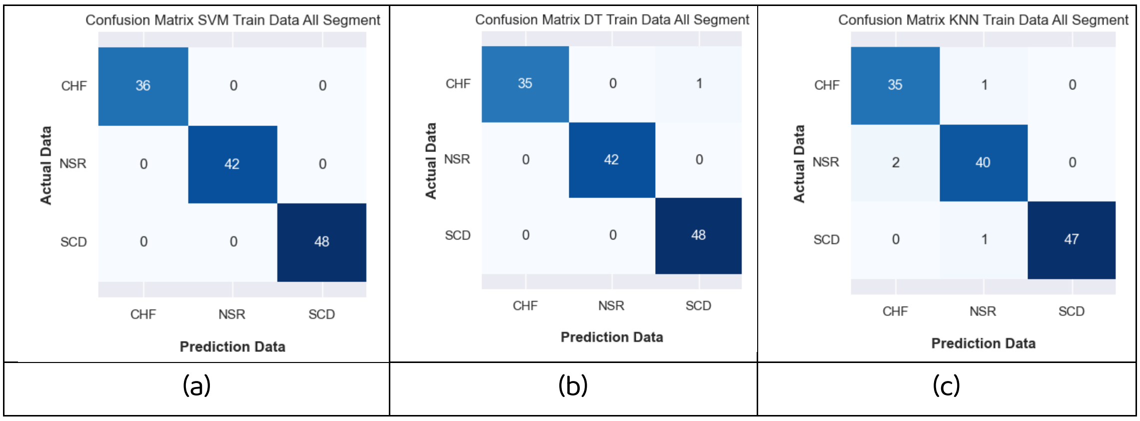

Confusion Matrix SVM Train Data Segment 1
Confusion Matrix DT Test Data Segment 1
Confusion Matrix KNN Test Data Segmen 1
Confusion Matrix SVM Test Data Segment 2
Confusion Matrix DT Test Data Segment 2
Confusion Matrix KNN Test Data Segment 2

(a) (b) (c)

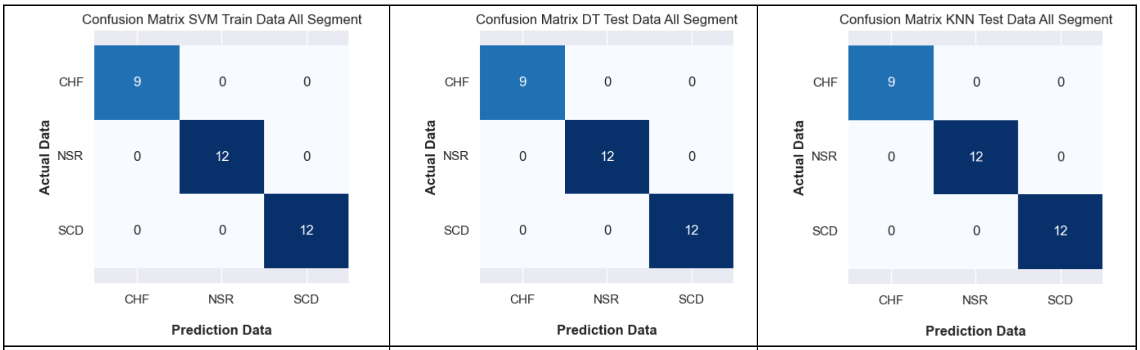
Confusion Matrix SVM Train Data Segment 3
Confusion Matrix DT Test Data Segment 3
Confusion Matrix KNN Test Data Segmen 3

(a) (b) (c)

Figure 7. Confusion Matrix of Classification Evaluation Results for Test Data Using Three Methods: (a) SVM, (b) DT, (c) KNN.


Confusion Matrix SVM Train Data All Segment
Confusion Matrix DT Train Data All Segment
Confusion Matrix KNN Train Data All Segment

(a) (b) (c)



(d) (e) (f)

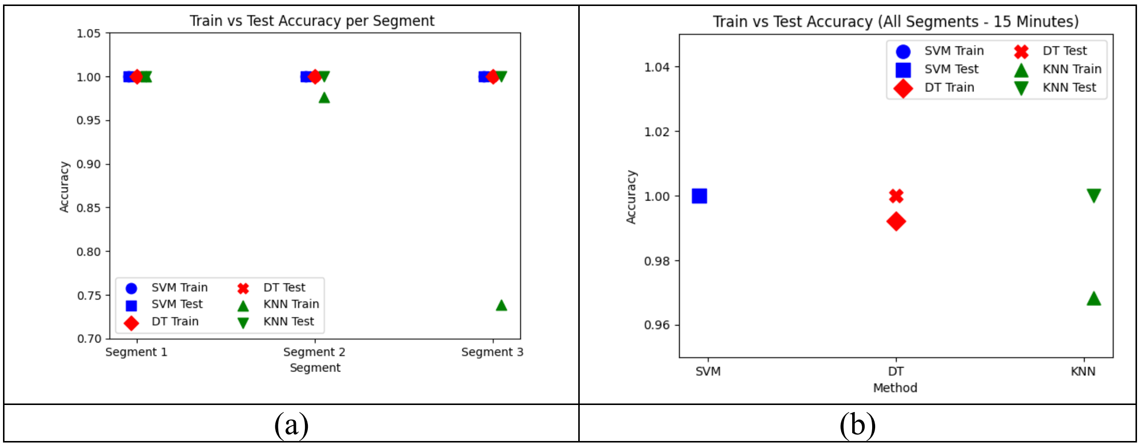
Figure 8. Confusion Matrix of Classification Performance for Training and Testing Data Using Three Methods: (a–c) Training Data (SVM, DT, KNN) and (d–f) Testing Data (SVM, DT, KNN)

3.4. Discussion

This study evaluated the performance of three classification methods, SVM, DT, and KNN, in analyzing ECG signals with a total duration of 15 minutes, consisting of three 5-minute segments. The results indicate that SVM and DT consistently achieved the highest performance across multiple evaluation scenarios. Although these results indicate strong classification capability, the repeated occurrence of 100% performance across multiple models and experimental settings requires careful interpretation. Given the relatively small dataset size, such results may reflect highly separable feature distributions, but they may also indicate potential overfitting or optimistic estimation. Even with subject-wise splitting, the possibility that the model captures subject-specific patterns rather than fully generalizable physiological characteristics cannot be completely ruled out.

However, such consistently optimal results should be interpreted with caution. As illustrated in Figure 9(a), which presents the comparison of training and testing accuracy across segments, SVM and DT show stable performance with minimal variation between segments. In contrast, KNN exhibits noticeable performance degradation, particularly in the third segment. The scatter visualization highlights that while SVM and DT maintain consistent accuracy across all segments, KNN experiences a significant drop in training performance in segment 3, indicating sensitivity to data complexity and feature overlap. The superior performance of SVM can be attributed to its ability to construct optimal hyperplanes that maximize class separation, particularly in high-dimensional feature spaces such as HRV data. Similarly, DT performs well due to its rule-based structure, which can effectively partition the feature space when class boundaries are clearly defined. However, the minimal gap between training and testing results observed in Figure 9 also raises concerns about potential overfitting or data leakage. If segments derived from the same subject are included in both training and testing sets, the model may learn subject-specific patterns rather than generalizable features, leading to overly optimistic performance.

Further insights can be observed in Figure 9(b), which presents the comparison between training and testing accuracy for the combined 15-minute signal. The results show that all methods achieve highly consistent performance, with only a slight decrease observed in KNN during training. Compared to the per-segment analysis, the combined 15-minute approach reduces performance variability and provides a more stable classification outcome. This suggests that aggregating multiple segments improves the separability of HRV features and reduces the impact of local fluctuations observed in individual segments. In contrast, KNN demonstrated a noticeable decline in performance, particularly in the third segment, where accuracy dropped significantly. This behavior can be explained by the nature of KNN, which relies on distance-based similarity. When feature distributions overlap as observed between CHF, SCD, and NSR in segment 3 KNN becomes more susceptible to misclassification, especially near class boundaries. The third segment may contain more complex or less separable HRV patterns, reducing the effectiveness of distance-based classification.



(a) (b)

Figure 9. Comparison of Training and Testing Accuracy Using Scatter Plots: (a) Per-Segment Analysis and (b) Combined 15-Minute ECG Signals for SVM, DT, and KNN.

Overall, the comparison between per-segment and combined 15-minute analysis reveals important insights. The per-segment approach captures short-term variations in HRV, which may reflect dynamic physiological changes but can also introduce variability and instability in classification performance. In contrast, the combined 15-minute approach provides a more comprehensive representation of the signal, resulting in more stable and consistent classification outcomes. These findings highlight the importance of selecting an appropriate temporal representation for HRV-based classification tasks.

Table 6 presents a comparative evaluation of three machine learning methods — SVM, DT and KNN used for classifying heart conditions based on HRV time-domain features extracted from ECG signals. Two studies are compared: a prior study by F. Panjaitan et al. (2023), which used a 5-minute ECG segment, and the current study (2026), which uses a longer 15-minute ECG segment. The comparison in Table 6 highlights notable differences between the prior study and the current work, particularly in dataset composition, the number of classes, and the evaluation design. The previous study used a shorter 5-minute ECG segment with a larger, more diverse set of subjects across multiple classes. In contrast, the current study employs a longer 15-minute signal but with a more limited distribution of subjects per class. This difference in dataset size and class composition may significantly influence the reported performance. Smaller datasets with well-separated feature distributions can lead to inflated performance metrics, including repeated 100% accuracy, precision, recall, and F1-score. In contrast, larger and more heterogeneous datasets generally present greater classification challenges and provide a more realistic estimation of model generalizability.

Furthermore, differences in evaluation design must be carefully considered. Although both studies report high performance, variations in data-splitting strategies, such as subject-wise versus record-wise partitioning, can substantially affect results. If data from the same subjects appear in both the training and testing sets, the model may capture subject-specific patterns rather than generalizable physiological characteristics, leading to overly optimistic performance estimates. Therefore, while the results of the current study appear competitive, direct comparison with prior work should be interpreted cautiously. A more rigorous evaluation, including larger datasets, balanced class distributions, and cross-dataset validation, is necessary to confirm the robustness and generalizability of the proposed approach.

Table 6. Comparison of results with state-of-the-art studies

Author/Year	Method	No. Of Subject	Signal Length Prediction Period	Acc (%)	Pre (%)	Rec (%)	F1-Score (%)
F. Panjaitan et al (2023)	SVM	18 NSR 51 CAD 15 CHF 20 SCD 11 VT	5 Min Interval	100	100	100	100
	DT			100	100	100	100
	KNN			96.78	100	98	98.95
Our Work (2026)	SVM	18 NSR	15 Min Interval	100	100	100	100
	DT	15 CHF		100	100	100	100
	KNN	20 SCD		100	100	100	100

Future work should include validation on larger, more diverse datasets and cross-dataset evaluation to ensure the robustness and generalizability of the proposed approach. Additionally, further investigation is needed to confirm whether the observed high performance reflects true physiological separability or is influenced by dataset-specific characteristics.

4. CONCLUSION

This study evaluated the performance of three classification methods, SVM, DT, and KNN, in distinguishing two heart disease conditions (SCD and CHF) and one normal condition (NSR) using 15-minute ECG data. The results indicate that SVM and DT consistently achieved higher performance compared to KNN across the evaluated scenarios. While both SVM and DT demonstrated strong and stable classification results, KNN showed slightly lower performance, particularly in handling more complex data distributions. The use of a 15-minute signal duration provided a more comprehensive representation of HRV features, which contributed to improved classification stability. However, the increased data complexity also appeared to affect the performance of KNN more noticeably than SVM and DT. These findings suggest that SVM and DT may be more suitable for this type of classification task under the given experimental conditions.

Despite these results, several limitations should be considered. The study was conducted using a limited set of classes and a dataset size, which may affect the generalizability of the findings. In addition, the evaluation design, particularly the data splitting strategy, may influence the reported performance and requires careful validation to avoid potential bias. Therefore, future research should focus on incorporating larger and more diverse datasets and evaluating additional classification methods to ensure more robust and generalizable results.

ACKNOWLEDGMENT

The authors gratefully acknowledge Universitas Satu for the support and facilities provided during this research

REFERENCES

[1] C. X. Wong *et al.*, "Epidemiology of sudden cardiac death: global and regional perspectives," *Heart Lung Circ.*, vol. 28, no. 1, pp. 6–14, 2019.

[2] F. Panjaitan, S. Nurmaini, and R. U. Partan, "Accurate prediction of sudden cardiac death based on heart rate variability analysis using convolutional neural network," *Medicina (B. Aires).*, vol. 59, no. 8, p. 1394, 2023.

[3] F. Z. Sudirman, W. Wisudawan, and J. Dase, "Karakteristik Kejadian Mati Mendadak Akibat Penyakit Kardiovaskular: Literature Review," *Jurnal Kesehatan Tambusai*, vol. 5, no. 4, pp. 11099–11109, 2024.

[4] D. Setiawan, A. Surtono, and S. W. Suciyati, "Ekstraksi Ciri Suara Jantung Menggunakan Metode Dekomposisi dan Korelasi Sinyal (Dekorlet) Berbasis Jaringan Syaraf Tiruan," *J. Teor. dan Apl. Fis*, vol. 3, 2015.

[5] S. Santoso *et al.*, "Penerapan Filter Digital untuk Menghilangkan Gangguan pada Sinyal Elektrokardiogram," *JREEC: Journal of Renewable Energy, Electronics and Control*, vol. 4, no. 2, pp. 36–42, 2024.

[6] M. I. Dewa, E. R. Widasari, and H. Fitriyah, "Analisis Perbandingan Performa Algoritma Pendeteksi Puncak R pada Realtime Akuisisi Sinyal Electrocardiography berbasis Shimmer," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, pp. 82–87, 2024.

[7] S. K. Nayak *et al.*, "A review of methods and applications for a heart rate variability analysis," *Algorithms*, vol. 16, no. 9, p. 433, 2023.

[8] B. Anwar, E. Faisal, V. Irvianita, R. Putranto, and H. Shatri, "Peran heart rate variability dan faktor inflamasi pada sindrom lelah kronik," *Indonesian Journal of Health Science*, vol. 4, no. 6, pp. 758–769, 2024.

[9] F. Panjaitan, S. Nurmaini, M. Akbar, A. H. Mirza, H. Syaputra, and T. B. Kurniawan, "Identification of classification method for sudden cardiac death: A review," in *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, IEEE, 2019, pp. 93–97.

[10] V. V. R. Maulina and H. Ohira, "Somatic Symptoms and Its Association with Anxiety and Interoception," *ANIMA Indonesian Psychological Journal*, vol. 39, no. 2, pp. E04–E04, 2024.

[11] S. Alinsaif, "Unraveling Arrhythmias with Graph-Based Analysis: A Survey of the MIT-BIH Database," *Computation*, vol. 12, no. 2, p. 21, 2024.

[12] A. Darmawahyuni, S. Nurmaini, M. Yuwandini, Muhammad Naufal Rachmatullah, F. Firdaus, and B. Tutuko, "Congestive heart failure waveform classification based on short time-step analysis with recurrent network," *Inform. Med. Unlocked*, vol. 21, 2020, doi: 10.1016/j.imu.2020.100441.

[13] M. G. Pradana, P. H. Saputro, and D. P. Wijaya, "Komparasi Metode Support Vector Machine Dan Naïve Bayes Dalam Klasifikasi Peluang Penyakit Serangan Jantung," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 5, no. 2, pp. 87–91, 2022.

[14] O. P. Barus, N. Phan, A. E. Widjaja, J. J. Pangaribuan, and R. Romindo, "Heart Disease Classification Using Decision Trees," in *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)*, IEEE, 2024, pp. 1–6.

[15] F. Panjaitan, S. Nurmaini, and R. U. Partan, "Hyperparameter Optimization in Machine Learning for Heart Disease Classification based on Heart Rate Variability Feature Analysis," in *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, IEEE, 2023, pp. 393–397.

[16] S. K. Nayak *et al.*, "A review of methods and applications for a heart rate variability analysis," *Algorithms*, vol. 16, no. 9, p. 433, 2023.

[17] "Sudden Cardiac Death Holter Database v1.0.0." Accessed: Mar. 17, 2023. [Online]. Available: https://physionet.org/content/sddb/1.0.0/

[18] "BIDMC Congestive Heart Failure Database v1.0.0." Accessed: Jun. 16, 2023. [Online]. Available: https://physionet.org/content/chfdb/1.0.0/

[19] "MIT-BIH Normal Sinus Rhythm Database v1.0.0." Accessed: Mar. 17, 2023. [Online]. Available: https://www.physionet.org/content/nsrdb/1.0.0/

[20] J. Naam, C. Suharinto, and S. Sumijan, "Digitalisasi Grafik Elektrokardiogram dengan Teknik Pixel Indexing," *Prosiding SISFOTEK*, vol. 1, no. 1, pp. 172–176, 2017.

[21] T. F. of the E. S. of C. the N. A. S. of P. Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[22] P. Shimpi, S. Shah, M. Shroff, and A. Godbole, "A machine learning approach for the classification of cardiac arrhythmia," in *2017 international conference on computing methodologies and communication (ICCMC)*, IEEE, 2017, pp. 603–607.

[23] F. Panjaitan, S. Nurmaini, and R. U. Partan, "Accurate Prediction of Sudden Cardiac Death Based on Heart Rate Variability Analysis Using Convolutional Neural Network," *Medicina (B. Aires)*, vol. 59, no. 8, p. 1394, 2023.

[24] F. Panjaitan, S. Nurmaini, and R. U. Partan, "Hyperparameter Optimization in Machine Learning for Heart Disease Classification based on Heart Rate Variability Feature Analysis," in *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, IEEE, 2023, pp. 393–397.

[25] A. M. Fuentes *et al.*, "Raman spectroscopy and convolutional neural networks for monitoring biochemical radiation response in breast tumour xenografts," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-28479-2.

[26] V. Artanti, M. Faisal, and F. Kurniawan, "Klasifikasi Cardiovascular Diseases Menggunakan Algoritma K-Nearest Neighbors (KNN).," *Techno. com*, vol. 23, no. 2, 2024.

[27] D. Tsiachris *et al.*, "Electrocardiographic characteristics, identification, and management of frequent premature ventricular contractions," *Diagnostics*, vol. 13, no. 19, p. 3094, 2023.

[28] M. Blondeel, T. Robyns, R. Willems, and B. Vandenberk, "Ventricular Depolarization Abnormalities and Their Role in Cardiac Risk Stratification—A Narrative Review," *Rev. Cardiovasc. Med.*, vol. 26, no. 1, p. 25921, 2025.