

Overcoming Class Overlap and Imbalance in ECG Detection and Classification: A Deep Attention-Based Model on MIT-BIH

Asmaa Sami Mirdan*

College of Computer Science and Information Technology, Kirkuk University, Kirkuk, Iraq. Email: asmaa.sami50@uokirkuk.edu.iq

ARTICLE INFO

Article history:

Received: 12/04/2025

Revised form: 20/04/2025

Accepted : 21/04/2025

Available online: 30/06/2025

Keywords:

Machine Learning, Random Forest, Decision Tree, Support Vector Machine, ECG

ABSTRACT

ECG signals are essential for monitoring and diagnosing cardiovascular conditions, particularly arrhythmias, which can lead to severe complications if undetected. This study introduces an AI-based approach for arrhythmia classification using the MIT-BIH arrhythmia dataset, addressing challenges like class imbalance, class overlap, and intra-patient bias. To enhance data quality, the dataset was balanced using the Synthetic Minority Oversampling Technique (SMOTE) and augmented with Gaussian noise for minority class samples. A Conv1D-Attention network was employed during preprocessing to extract local ECG features and focus on key waveforms. Among the evaluated classifiers, decision tree, random forest, and support vector machine (SVM), the random forest achieved the highest accuracy of 91%. Although preprocessing reduced class imbalance and variance, a drop in performance was observed. This reflects a realistic evaluation scenario by preventing data leakage from similar ECG segments of the same patient in both training and test sets. Enforcing patient-independent segmentation compelled the model to generalize beyond individual patterns, a critical step for real-world applications. This study highlights the importance of rigorous evaluation protocols in biomedical machine learning. Combining data augmentation with attention-based feature extraction significantly enhances model generalizability, particularly in handling overlapping and imbalanced classes. This approach shows promise for developing reliable, patient-independent diagnostic tools for early arrhythmia detection in clinical settings.

MSC..

<https://doi.org/10.29304/jqcm.2025.17.22175>

1. Introduction

Irregular heartbeat occurring due to malfunctioning of the electrical impulses that regulate the heartbeats is one of the leading causes of death globally from cardiovascular disease (Ferreir, Kumar, Soni, Acharya, & Acharya, 2023). Very early and very accurate detection of arrhythmias is of utmost importance for timely medical intervention to prevent severe complications like sudden cardiac arrest (Nagpal, Pundkar, Singh, & Gadkari, 2024). Electrocardiogram (ECG) signals play a pivotal role in monitoring and diagnosing many cardiovascular conditions, therefore, accurate interpretation and classification of ECG signals are very important for making the right medical decision at the right time (Zeng, Shan, Yuan, & Du, 2024). Manual analysis of an ECG is time-consuming, prone to

*Corresponding author: Asmaa Sami Mirdan

Email addresses: asmaa.sami50@uokirkuk.edu.iq

Communicated by 'sub editor'

human error, and requires interpretation by cardiologists only (Maturi et al., 2025). With the great technological developments that have occurred in the modern era, machine learning models have emerged that have shown great potential in capturing patterns and relationships between big data, as these models learn from labeled data sets, thus the capabilities of these models can be leveraged to classify ECG signals into different types, as machine learning algorithms can recognize patterns and anomalies that may not be easily detected by traditional methods (Ansari, Mourad, Qaraqe, & Serpedin, 2023; Guo & Li, 2023). Traditional methods for ECG analysis, such as rule-based and statistical methods, often suffer from limitations in handling complex and high-dimensional data (Sengupta & Das, 2024).

To clear out some artifacts brought about by muscle movements during scanning or positioning of the electrodes, raw ECG signals are prone to noise and thus in need of filtering techniques (Dias, Probst, Silva, & Gamboa, 2024). Another hurdle is the quality of the dataset, which largely affects the performance of a model since inadequate or imbalanced data would lead to a biased model and subsequently biased predictions (Owusu, Quainoo, Mensah, & Appati, 2023). They may also suffer from overfitting problems, where the model learns some features unique to the training dataset and does not generalize well to new data (Gygi, Kleinstein, & Guan, 2023). These require integration of efficient feature selection and data augmentation techniques for improved model performance.

Following previous studies (M. A. Assaad & Shakah, 2024; M. Assaad, Boné, & Cardot, 2008; Baker, Mohammed, & Jihad, 2022; Baker, Taher, & Jihad, 2023; Hameed, 2023; H Khalid, 2024; Hind Khalid, 2024; Rashid et al., 2021; Umar, Rashid, Ahmed, Hassan, & Baker, 2024), this study investigates AI models' effectiveness in detecting and classifying ECG arrhythmias. In particular, the role of this research in the reliable development of automatic diagnostic systems would be to help earlier identification of arrhythmias to make better decisions, thus reducing the complications and risks associated with cardiovascular diseases.

Unlike many previous works that report high classification accuracy but do not address critical challenges such as patient-specific data leakage, segment-level class overlap, and limited generalizability, this study introduces a more robust evaluation pipeline by combining SMOTE-based balancing, Gaussian noise augmentation, and a 1D convolutional attention-based model architecture. The resulting model shows slightly lower performance than some existing methods, but reflects a more realistic and generalizable classification performance, offering meaningful contributions to the development of clinically applicable arrhythmia detection systems.

2. Related Works

Several prior studies have explored the detection of cardiovascular diseases, with a particular emphasis on arrhythmias. The study (Singh & Singh, 2019) proposed a learning-based approach that integrated three distinct machine learning algorithms alongside three filter-based feature selection techniques. The model identified the most relevant features from an arrhythmia dataset, with the highest performance achieved by a random forest classifier using a gain-percentage feature selection method and a selected subset of 30 features, yielding an accuracy of 85.58%. However, despite these promising results, there remains a notable gap in enhancing the generalizability and precision of arrhythmia detection models, particularly when dealing with imbalanced datasets or real-world noisy signals.

Numerous studies have employed deep learning frameworks to design innovative approaches for the automatic detection of myocardial infarction (MI) using electrocardiogram (ECG) signals, both in their raw form and after noise reduction. One notable method utilized a convolutional neural network (CNN) to distinguish between normal and MI-affected ECG pulses, achieving average accuracy of 93.53% (Acharya et al., 2017). Many existing approaches lack robustness when confronted with real-world ECG data that may be affected by noise, artifacts, or irregular patterns. This highlights the need for more adaptable and noise-tolerant deep learning models capable of generalizing effectively.

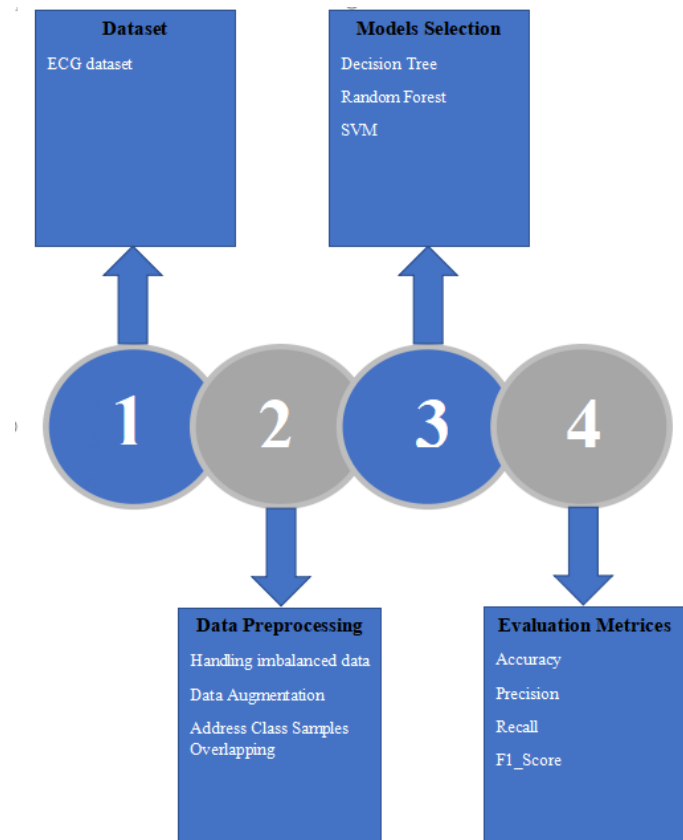
In (S.-F. Li, Huang, & Wu, 2023), a CNN was employed to analyze ECG data directly, without the need for signal transformation or manual feature extraction. The study aimed to classify individual heartbeat ECG images using CNNs combined with the Taguchi method for optimization. It considered all fifteen arrhythmia types, grouped into five classes, from the MIT-BIH arrhythmia dataset. The proposed model achieved a classification accuracy of 96.79%. The reliance on image-based ECG representations may limit the model's adaptability to raw time-series data often encountered in real-world clinical settings. This study did not address the problem of overlapping class samples in the MIT-BIH dataset.

In (Khan, Yu, Yuan, & Rehman, 2023), the proposed method leverages a CNN to directly extract features from input heartbeat signals, eliminating the need for manual preprocessing. To address the issue of class imbalance within the training dataset, the SMOTE was applied. This approach resulted in an average classification accuracy of 98.63%, with a specific class accuracy reaching 92.86%. A notable gap remains in evaluating the model's performance on real-world clinical data, where signal noise and patient variability may differ significantly from the training conditions.

Another study (Ramkumar, Babu, Priyanka, & Kumar, 2021) proposed a hybrid approach for classifying ECG arrhythmia signals from the MIT-BIH Arrhythmia database, utilizing SVM optimized through Particle Swarm Optimization (PSO) and a Genetic Algorithm (GA), in conjunction with Independent Component Analysis (ICA) for feature extraction. The optimization techniques enhanced the performance of the SVM classifier, resulting in a classification accuracy of 96%. This study did not address the problem of overlapping class samples in the MIT-BIH dataset.

3. Methodology

The given methodology employs machine learning algorithms in the production of a decision support system for



ECG signal classification. The proposed system, therefore, aids medical decision-making processes and accelerates the processing and interpretation of ECG data that would help save time for the doctors, particularly in emergencies. The same system assists the physicians in determining which cases require immediate attention. Figure 1 shows the overall framework of the proposed methodology.

Fig. 1- The general structure of the proposed methodology

3.1 Dataset Description

This research relies on a dataset of heartbeat signals obtained from Kaggle (Kaggle, 2023). This dataset consists of two sets of heartbeat signals derived from two popular datasets in heartbeat classification, the MIT-BIH arrhythmia

dataset and the PTB diagnostic ECG database. This dataset was used to explore the heartbeat. The signals correspond to ECG patterns of heartbeats of normal cases, cases affected by various heart rhythm disorders, and myocardial infarction. The dataset includes 5 categories (Oleiwi, AlShemmary, & Al-Augby, 2024):

- 0: "Normal".
- 1: "Atrial Premature".
- 2: "Premature Ventricular Contraction".
- 3: "Fusion of Ventricular and Normal".
- 4: "Fusion of Paced and Normal".

The dataset contains 87,554 data samples, where each sample corresponds to an ECG signal characterized by 188 features. Features 1-187 represent the heartbeat signal, while feature 188 denotes the class label for that signal. The number of samples per class is summarized in Table 1.

Table 1- The number of samples for each class within the data set

Category	Samples Number
Normal	57977
Fusion of Paced and Normal	5145
Premature Ventricular Contraction	4630
Atrial Premature	1778
Fusion of Ventricular and Normal	513

Figures 2, 3, 4, 5, and 6 show the ECG signal for each of the classes in the data set.

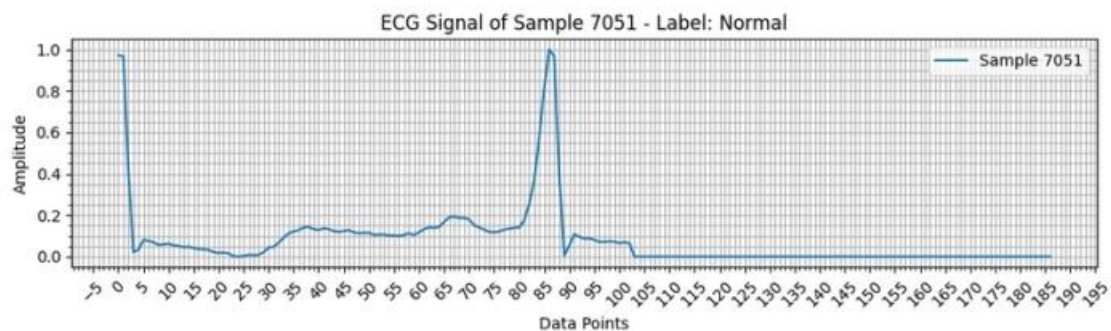


Fig. 2- ECG signal for 'Normal' label

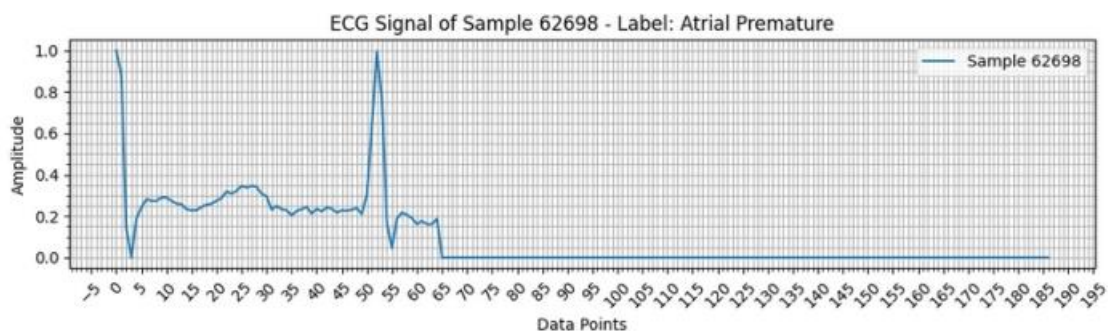


Fig. 3- ECG signal for 'Atrial Premature' label

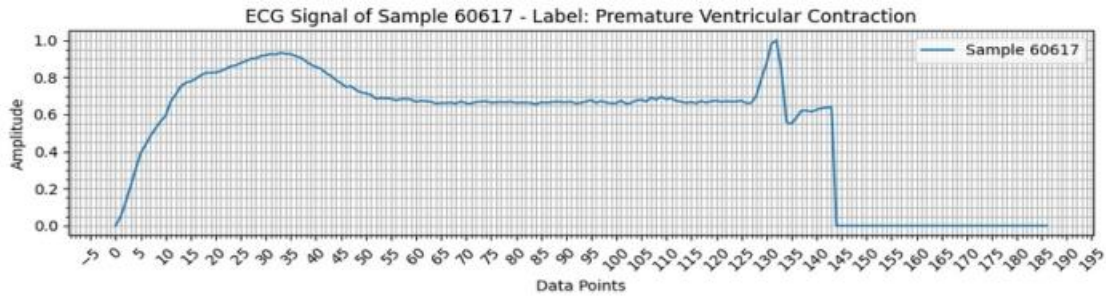


Fig. 4- ECG signal for `Premature Ventricular Contraction` label

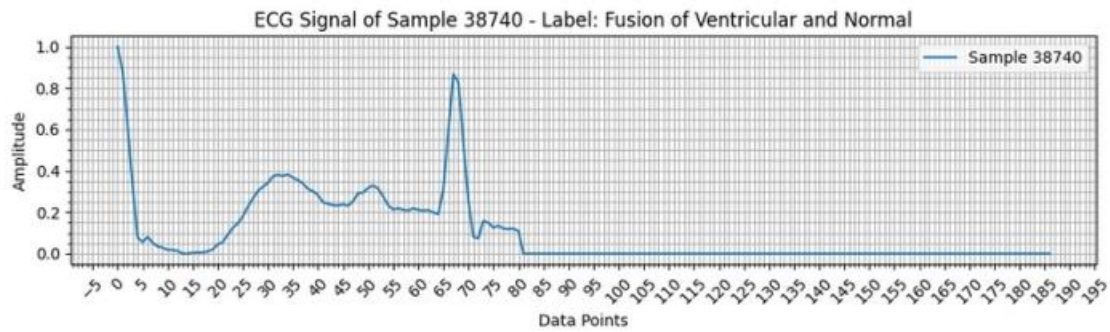


Fig. 5- ECG signal for `Fusion of Ventricular and Normal` label

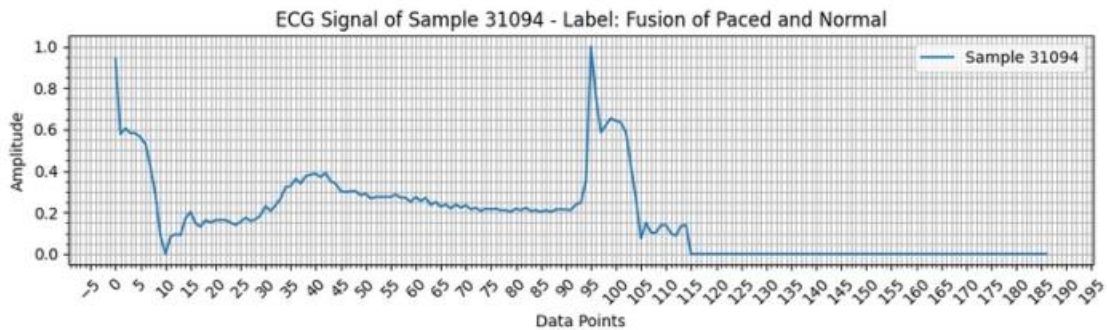


Fig. 6- ECG signal for `Fusion of Paced and Normal` label

3.2. Data Preprocessing

Data preprocessing is the crucial step that turns raw data into a clean, formatted version before it is fed into any machine learning model (Jansen, Aldous, Salminen, Almerexhi, & Jung, 2023). Intensive preprocessing is the foundation of maximizing accuracy and performance in any machine learning model. In practice, proper data preprocessing means that the quality of the input data is improved, further assisting the model during its performance in terms of faster convergence with better generalization capabilities (Tariq, Palade, Ma, & Altahhan, 2023). Besides, preprocessing prevents the learning process from becoming prejudiced or overfitted and thus allows different features to contribute uniformly during learning, thereby increasing accuracy and robustness (Habib & Okayli, 2024). Proper data preprocessing methods for ECG arrhythmia detection give the machine learning model advantages in detecting subtle patterns in ECG signals, which increase the reliability of diagnosis and promote timely medical intervention.

3.2.1. Handling Imbalanced Datasets:

In AI, the issue of imbalanced datasets is a significant challenge. A dataset is considered imbalanced when one class contains substantially more samples than the others (Balla, Habaebi, Elsheikh, Islam, & Suliman, 2023). This imbalance can lead to several challenges (Hasanin, Khoshgoftaar, Leevy, & Seliya, 2019), including:

- **Biased models:** In this case, the model is biased towards the majority class; in other words, the model prioritizes the majority class and learns poorly from the minority classes, resulting in poor model performance.
- **Unreliable performance metrics:** Performance metrics such as accuracy can be unreliable if the data is unbalanced, as they can give high results, but these results do not represent the true performance of the model. The reason for this is that most of the predictions for the majority class will be correct.
- **Overfitting:** The model may focus on memorizing the characteristics and patterns of the majority class, resulting in poor generalization of the model to the minority class.

Therefore, at this stage, the crucial step involves correcting the defect in the data set. Figure 7 shows the number of samples for each class.

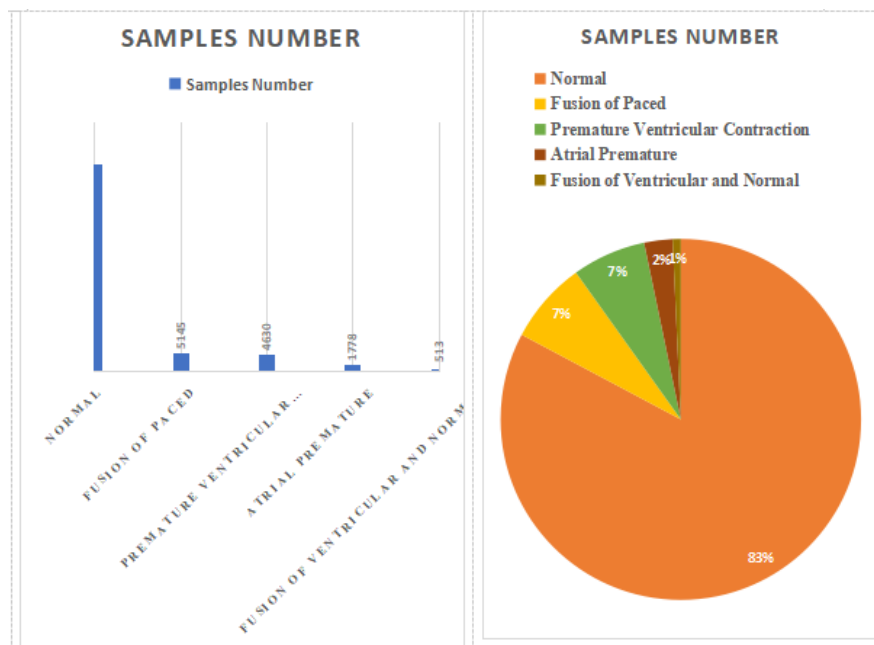


Fig. 7- Number of samples for each class

One way to solve this problem is to under sample the majority class, which means that a set of rows are excluded from the majority class so that the number of rows for both the majority and minority classes becomes equal, but doing this process leads to the loss of a lot of data that can be very important and useful in the process of training the model. Another option is to oversample the minority class, which means that a random iteration of the rows of the minority class is performed until the number of samples for the majority and minority classes becomes equal, but the problem with this method is that it leads to the problem of overfitting because the model learns from the same examples.

To solve these problems, this research will rely on the SMOTE, which is an oversampling technique that produces synthetic samples for the minority class (Wongvorachan, He, & Bulut, 2023). SMOTE is a preprocessing technique used to address imbalance in a dataset. Figure 8 shows the number of samples for each class after applying the oversampling process using the SMOTE method.

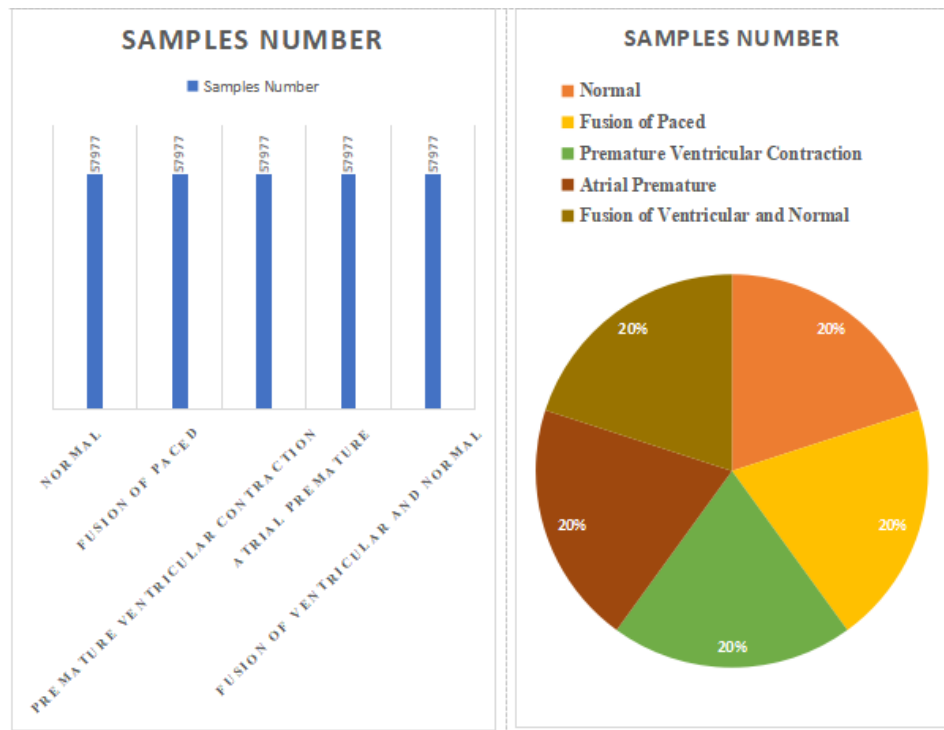


Fig. 8- Number of samples for each class after applying the SMOTE method

3.2.2. Data Augmentation:

Various sources can introduce noise into ECG signals, which can interfere with accurate heart rhythm analysis:

1. Power line interference: This is caused by electromagnetic interference from electrical devices, power lines, or poor grounding, resulting in periodic high-frequency noise appearing in the ECG (Mir & Singh, 2024).
2. Muscle distortions: These are caused by muscle contractions or patient movement, resulting in noise that affects the ECG signals (Atanasoski et al., 2024).
3. Baseline drift: Low-frequency deviation in the ECG signal caused by patient breathing, poor electrode contact, or movement (H. Li, Ditzler, Roveda, & Li, 2024).
4. Electrode contact noise: Poor electrode placement, dry electrodes, or loose connections can cause sudden changes in the signal amplitude, resulting in temporary spikes in the signal shape (Atanasov, 2023).
5. Motion distortions: Body movements such as twitching, coughing, or changing position can create irregular signal distortions that affect the interpretation of the ECG waveform (Khalili, GholamHosseini, Lowe, & Kuo, 2024).

Therefore, at this stage, noise will be introduced into the ECG signals, so that samples containing noise are added to the existing samples, so that the samples simulate the actual and real application of ECG signals, and so that the model is able to classify the signal even if there is noise in it.

Adding Gaussian noise is a popular data preprocessing technique used to improve the power and generalization of machine learning models, where Gaussian noise is introduced into the dataset to simulate real-world variations and boost the performance of the model (Dou et al., 2023). The model is highly prone to adapt based on memorizing specific patterns in the training data. Therefore, by adding Gaussian noise, the model could better cope with unseen or slightly modified data, thus accommodating changes in real-life ECG signals. This data augmentation would help enhance classification accuracy in arrhythmia detection while lowering the model's sensitivity to slight variations in the heart signal recordings. Thus, strengthens the development of robust and flexible machine learning models.

Mathematically, Gaussian noise is defined as a random variable with a probability density function (PDF) that follows a normal distribution, as described by the following equation (Camuto, Willetts, Simsekli, Roberts, & Holmes, 2020; Wodecki, Michalak, Wyłomańska, & Zimroz, 2021):

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

Where:

P(x): Probability density function.

x: The random variable

μ : The mean of the distribution.

σ : the standard deviation of the distribution.

Adding Gaussian noise to an ECG signal involves introducing random noise that follows a Gaussian distribution to the original signal. This technique serves as a form of data augmentation, enhancing the diversity of the training data by generating slightly altered versions of the original ECG signals. These variations simulate real-world conditions where signals are exposed to different noise sources (Lee, Zaheer, Astrid, & Lee, 2020). By doing so, the model is better equipped to generalize to unseen and real-world data, reducing the risk of overfitting. Figure 9 illustrates ECG signals before and after the addition of Gaussian noise.

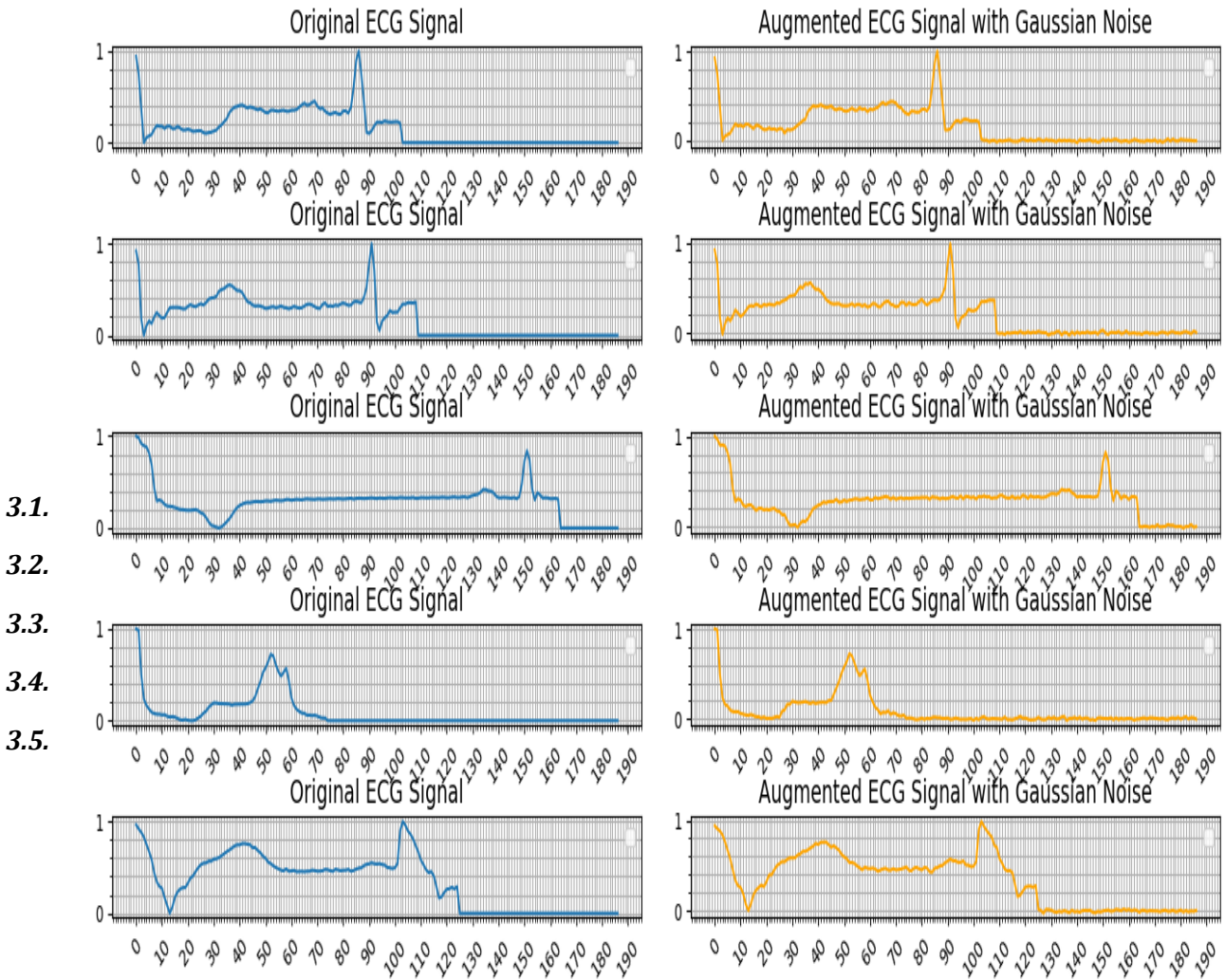


Fig. 9- ECG signals before and after adding Gaussian noise

3.2.3. Address Class Samples Overlapping

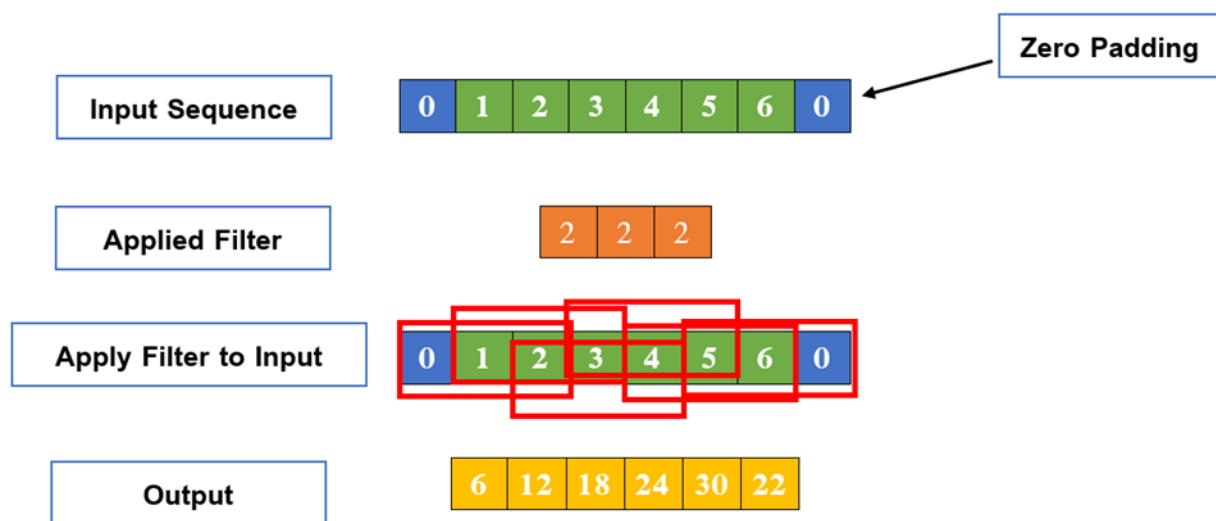
The dataset comprises ECG readings from 48 individuals, and each individual segmented their ECG into beats. The length of each beat varies depending on the window used for segmentation (Oleiwi, AlShemmary, & Al-augby, 2023). For this study, segments of length 187 were used, representing only a segment rather than a complete ECG signal for a single patient. Consequently, the training samples may include ECG segments (beats) from the same patient, which is known as the intra-patient scheme.

To address this problem, we will rely on a set of DL layers to extract subtle patterns within these signals. First, a one-dimensional convolution layer is used to capture local temporal features from the ECG pulse. This layer helps extract morphological patterns and is very useful in reducing subtle artifacts by creating more robust low-level representations (Lekkas, Vrochidou, & Papakostas, 2025).

The convolution layer is followed by the attention layer, which is considered the most important layer for handling interference between samples. It allows focusing on the most important parts of the ECG signal and weighting the informative parts more, thus reducing the influence of irrelevant overlapping regions (Huang et al., 2024). This helps reduce class confusion due to similar features or sample overlap.

This architecture addresses the problem of overlapping between class samples. The Conv1D layer extracts localized temporal patterns to identify unique features in ECG waveforms. The attention mechanism also allows focusing on informative regions within each pulse, mitigating the effect of overlapping segments. This selective focus helps distinguish between arrhythmias with overlapping morphologies.

The Convolution1D layer is designed to process sequential data by applying convolutional operations along a single dimension where a convolutional filter (or kernel) slides over the input sequence, performing element-wise multiplications with the input values within its receptive field (Muralidharan et al., 2021). These weighted values are then summed to produce a single output point. This process is repeated across the entire sequence, generating a transformed output sequence (Muralidharan et al., 2021). Figure 10 illustrates an example of a Convolution1D



operation.

Fig. 10- An example of a Convolution1D operation

Attention mechanisms enhance the models by dynamically focusing on the most relevant input elements by prioritizing essential information, thereby improving both prediction accuracy and computational efficiency. The architecture of an attention mechanism consists of three key components: the encoder, the attention module, and the decoder (Niu, Zhong, & Yu, 2021). Figure 11 illustrates the structure of the attention mechanism:

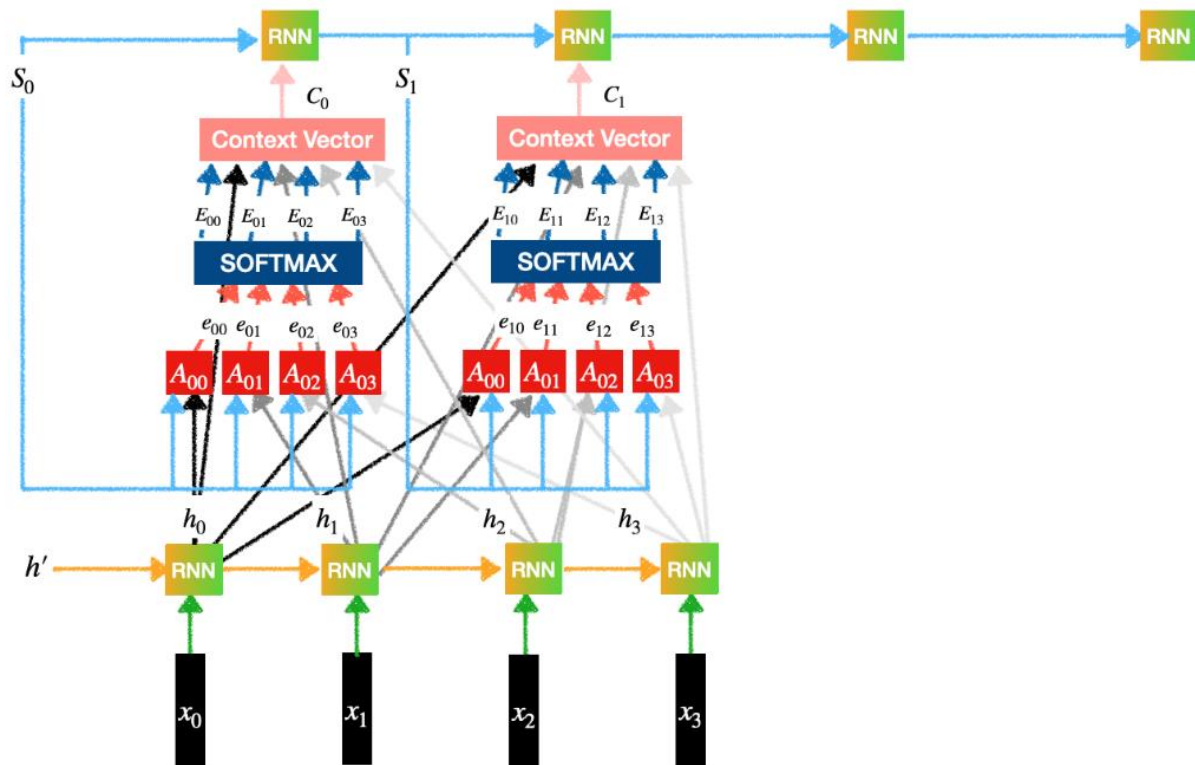


Fig. 11- The structure of the attention mechanism

The encoder processes the input data step-by-step, producing hidden states that capture both current and past information. The attention module, made up of a feed-forward network, softmax function, and context vector generation, helps the decoder predict the next output symbol using these context vectors and its current state. This iterative process continues until the full output is generated. By focusing on relevant parts of the input, the attention mechanism improves the model's ability to handle long sequences and generate accurate predictions.

3.2.4. Model Selection:

3.2.4.1. Decision Tree

A decision tree (Ahmed, Ahmed, & Jwmaa, 2023) algorithm is a predictive modeling technique used in machine learning and data analysis. It recursively partitions a dataset into subsets based on the most significant features, creating a tree-like structure. Each inner node represents a decision based on a feature, and each leaf node represents the expected score or classification. The feature that will split the most impurity will be chosen at each node in the algorithm, and the splitting procedure will continue until a given stopping criterion applies, for example, a set depth of the tree and a set minimum number of samples per leaf. Decision trees are susceptible to overfitting and usually pick up noise; thus, many strategies are put in place to counteract this, including pruning and specification of minimum sample sizes. Decision trees are used in different applications like classification and regression tasks mainly due to their simplicity, flexibility, and ease of implementation.

3.2.4.2. Random Forest:

The random forest (Schlenger, 2024) algorithm is a type of ensemble learning that builds a collection of decision trees, which work together to predict the target value. During the training phase, a 'forest' of decision trees is

created, with each tree generated from a random subset of the training data and a limited number of features. This randomization is achieved through bootstrapping, where subsets of the dataset are drawn with replacement. Additionally, at each split within a tree, a random selection of features is used to ensure no single feature dominates the model. For predictions, the algorithm aggregates the outputs of all individual trees, averaging their predictions for regression tasks or using majority voting for classification tasks. Random forests are widely used for applications such as classification, regression, and feature importance estimation due to their versatility and robust performance.

3.2.4.3. SVM Algorithm:

The SVM (Valkenborg, Rousseau, Geubbelmans, & Burzykowski, 2023) is a widely recognized and powerful algorithm commonly used for classification and regression tasks. In classification, SVM identifies the optimal decision boundary that separates different categories as distinctly as possible. The data points closest to this boundary are referred to as 'support vectors.' SVM is particularly effective for high-dimensional problems and incorporates kernel functions to handle non-linear decision boundaries. These kernel functions transform input features into a higher-dimensional space, with common types including polynomial functions, radial basis functions (RBF), and sigmoid functions. The core principle of SVM is to maximize the margin between the support vectors and the decision boundary, which enhances the model's robustness and generalization ability. SVM has been successfully applied across various domains, from complex tasks like image analysis to text classification, demonstrating its effectiveness in both linear and non-linear classification problems, especially with high-dimensional data.

3.2.5. Evaluation Metrics:

To evaluate the model, the dataset was split into 80% for training and 20% for testing. A confusion matrix (CM) is commonly used to assess the performance of models, particularly in supervised learning tasks (Haghighi, Jasemi, Hessabi, & Zolanvari, 2018; Markoulidakis, Kopsiaftis, Rallis, & Georgoulas, 2021). The structure of the CM is

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

illustrated in Figure 12.

Fig. 12- The CM structure

- TP (True Positive): The model correctly predicts a positive observation.
- TN (True Negative): The model correctly predicts a negative observation.
- FP (False Positive): The model incorrectly predicts a positive observation.
- FN (False Negative): The model incorrectly predicts a negative observation.

Given these parameters, they could calculate some of the significant performance metrics, which include recall, precision, and accuracy.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

The above equation can be discovered as "Measures the proportion of actual positives that are correctly identified by the model [38]

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

The above equation can be interpreted as: "It measures the proportion of positive predictions that are truly correct (Olewi, AlShemmary, & Al-Augby, 2023).

Accuracy represents the overall performance of the model by calculating the proportion of correct predictions (both positive and negative) to the total number of predictions (Roth et al., 2022). The equation is as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

The F1 Score is a metric that combines precision and recall, providing a balanced evaluation of a model's overall performance (Wodecki et al., 2021). The formula is as follows:

$$F1_Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

4. Results and Discussions

Figure 13 illustrates the CM of the decision tree (DT) algorithm before applying the preprocessing steps...

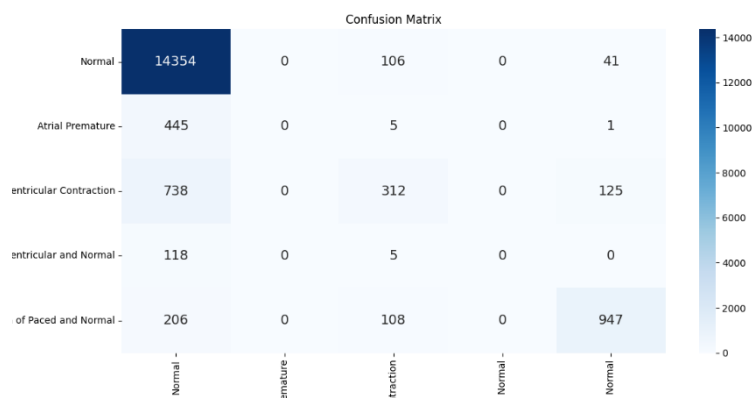


Fig. 13- The CM of the DT algorithm before the preprocessing steps

The CM provides a summary of the performance of the DT algorithm in classifying ECG signals. The diagonal elements represent correct predictions for each class. For instance, the top-left element (14,365) indicates 14,365 correct predictions for the Normal class. In contrast, the off-diagonal elements represent misclassifications. For example, the element in the second row, first column (448) shows that 448 instances of the Atrial Premature class were misclassified as the Normal class. Figures 14 and 15 display the CM for the random forest (RF) and SVM algorithms, respectively, before preprocessing. Table 2 presents the evaluation metrics for each algorithm before preprocessing.

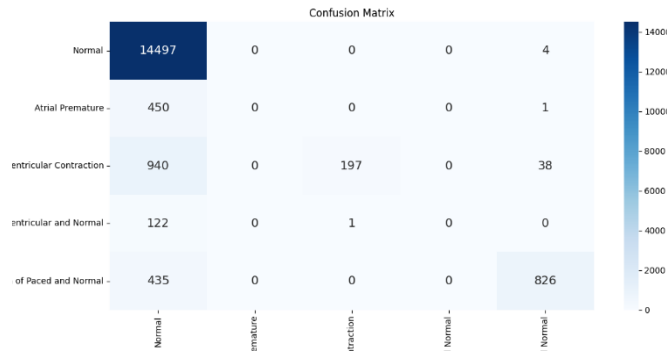


Fig. 14- The CM of the RF algorithm before the preprocessing steps

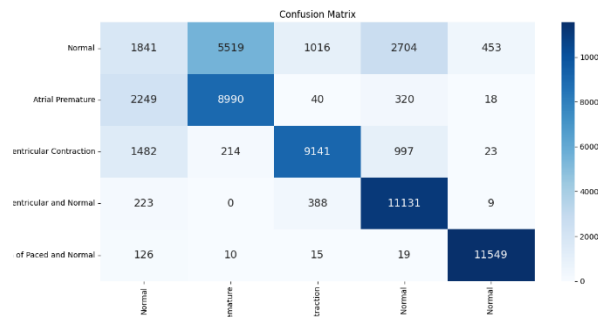


Fig. 15- The CM of the SVM algorithm before the preprocessing steps

Table 2. The evaluation metrics results of each algorithm before the preprocessing steps

Class	Precision	Recall	F1_Score	Accuracy
DT				
Normal	0.90	0.99	0.95	89.161≈89%
Atrial Premature	0.00	0.00	0.00	
Premature Ventricular Contraction	0.58	0.27	0.36	
Fusion of Ventricular and Normal	0.00	0.00	0.00	
Fusion of Paced and Normal	0.85	0.75	0.80	
RF				
Normal	0.88	1.00	0.94	88.63≈89%
Atrial Premature	0.00	0.00	0.00	
Premature Ventricular Contraction	0.99	0.17	0.29	
Fusion of Ventricular and Normal	0.00	0.00	0.00	
Fusion of Paced and Normal	0.95	0.66	0.78	
SVM				

Normal	0.31	0.16	0.21	72.938≈73%
Atrial Premature	0.61	0.77	0.68	
Premature Ventricular Contraction	0.86	0.77	0.81	
Fusion of Ventricular and Normal	0.73	0.95	0.83	
Fusion of Paced and Normal	0.96	0.99	0.97	

From Table 2, we observe the following:

1. Normal Class: All three models demonstrate strong performance in detecting normal ECG signals, achieving high precision, recall, and F1-scores. However, the DT and RF models slightly outperform SVM in recall.
2. Atrial Premature: Both the DT and RF models fail to detect this class, as indicated by precision, recall, and F1-scores of 0.00. In contrast, SVM achieves an F1-score of 0.68.
3. Premature Ventricular Contraction: The DT and RF models struggle to detect this class, while SVM performs significantly better with an F1-score of 0.81.
4. Fusion of Ventricular and Normal: The DT and RF models completely fail to detect this class, whereas SVM achieves strong performance with an F1-score of 0.83.
5. Fusion of Paced and Normal: All models perform well in detecting this class, but SVM achieves the highest performance with an F1-score of 0.97.

Figures 16, 17, and 18 illustrate the confusion matrices for the DT, RF, and SVM algorithms, respectively, after preprocessing. Table 3 presents the evaluation metrics for each algorithm following preprocessing.

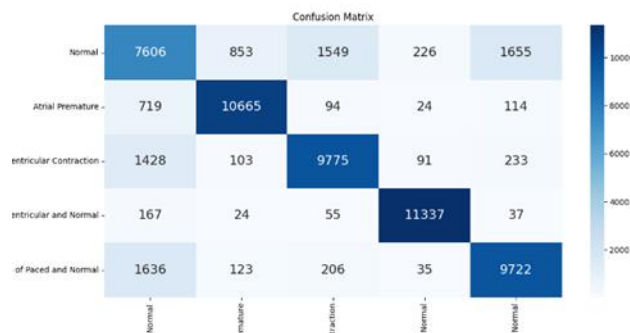


Fig. 16- The CM of the DT algorithm after the preprocessing steps

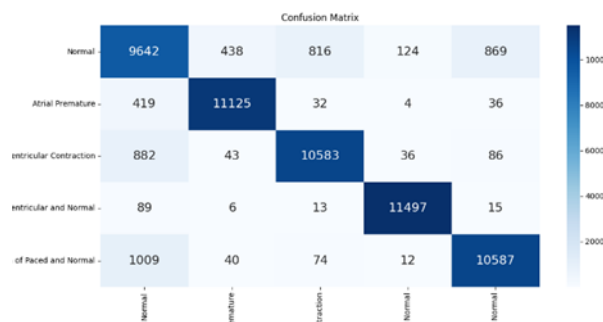
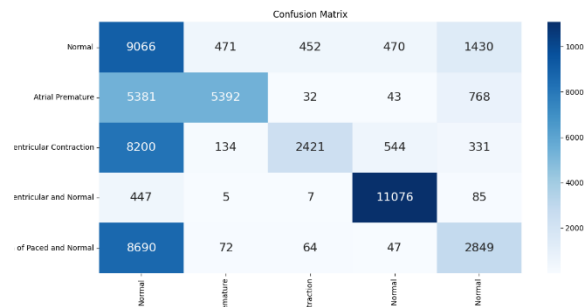


Fig. 17- The CM of the RF algorithm after the preprocessing st**Fig. 18- The CM of the SVM algorithm after the preprocessing steps****Table 3. The evaluation metrics results of each algorithm after the preprocessing steps**

Class	Precision	Recall	F1_Score	Accuracy
DT				
Normal	0.66	0.64	0.65	83.973≈84%
Atrial Premature	0.91	0.92	0.91	
Premature Ventricular Contraction	0.84	0.84	0.84	
Fusion of Ventricular and Normal	0.97	0.98	0.97	
Fusion of Paced and Normal	0.83	0.83	0.83	
RF				
Normal	0.80	0.81	0.81	91.376≈91%
Atrial Premature	0.95	0.96	0.96	
Premature Ventricular Contraction	0.92	0.91	0.91	
Fusion of Ventricular and Normal	0.98	0.99	0.99	
Fusion of Paced and Normal	0.91	0.90	0.91	
SVM				
Normal	0.29	0.76	0.42	52.677≈53%
Atrial Premature	0.89	0.46	0.61	
Premature Ventricular Contraction	0.81	0.21	0.33	
Fusion of Ventricular and Normal	0.91	0.95	0.93	
Fusion of Paced and Normal	0.52	0.24	0.33	

Comparing the results before and after preprocessing, we note that before preprocessing, the normal pulse class dominated the other classes because it was the majority class, which led to bias in the model's decisions. After preprocessing, we note that all classes exhibited effective precision and recall values, particularly those generated by the DT and RF. The RF achieved the highest overall accuracy (91%) and strong overall performance for all classes. The SVM algorithm struggled to achieve good metrics due to its sensitivity to noisy data.

The comparison of results before and after preprocessing highlights critical challenges related to class imbalance and patient data overlap in ECG signal classification. Before preprocessing, all classifiers, particularly the DT and RF,

exhibited strong performance in detecting the majority class (normal). However, they struggled to accurately identify minority classes due to the following reasons:

- Class imbalance: The majority classes dominate the model's learning.
- Patient overlap: Due to the method of obtaining data samples in windows of length 187, multiple samples of the same patient were scattered across the dataset. This resulted in similar segments of the same individual in both the training and test sets, leading to overfitting and overly optimistic performance metrics.

After applying SMOTE for class equalization and augmenting the data with Gaussian noise to enhance diversity, all models demonstrated improved generalization across all classes, with the RF achieving the best performance post-preprocessing.

It is crucial not to interpret the overall decline in performance metrics after preprocessing as a failure of the models. Instead, this decline reflects a more realistic and reliable generalization capability, achieved by minimizing data leakage between the training and test sets and improving class representation and diversity. These findings highlight the importance of careful data segmentation and preprocessing, particularly in healthcare applications, where patient-specific patterns can inadvertently lead to overfitting and overly optimistic results.

Table 4 provides a comparison between this study and previous as well as recent research.

Table 4. A comparison with previous and recent studies

Ref	Dataset	Preprocessing Steps	Performance
(Ren et al., 2025)	MIT-BIH	bi-directional long and short-term memory networks (BiLSTM) and autoencoder	96%
(Ketu & Mishra, 2022)	MIT-BIH	Balance the dataset using SMOTE	92%–99%
(Rai & Chatterjee, 2022)	MIT-BIH	Balance the dataset using SMOTE	99.89%
(Darmawahyuni et al., 2022)	MIT-BIH	Undesirable noise using DWT	98%
Proposed Model	MIT-BIH	Balance the dataset using SMOTE, Data augmentation, Address Class Samples Overlapping	91%

The proposed model achieves an accuracy of 91%, which is lower than the 99.89% reported in some previous studies. However, many of these studies overlooked the issue of class overlap within the dataset, leading to potential data leakage when patient segments appear in both the training and test sets. This study directly addresses this challenge by implementing a robust data pipeline that includes data balancing, the injection of Gaussian noise to enhance data diversity, and the application of a one-dimensional convolutional network with an attention mechanism to mitigate class overlap. This approach ensures a more realistic evaluation of the model. Consequently,

the high accuracy reported in previous works is likely misleading, as it stems from data leakage or unbalanced evaluation practices, resulting in biased models.

5. Conclusions and Future Works

Cardiovascular disease is a leading global cause of death, often due to arrhythmias caused by malfunctioning electrical impulses that regulate the heartbeat. Early and accurate arrhythmia detection is vital to prevent severe complications like sudden cardiac arrest. This study used machine learning models to classify ECG signals, with preprocessing steps such as SMOTE for data equalization and Gaussian noise augmentation, improving performance across classes. Addressing class overlap further enhanced model generalizability. Although preprocessing reduced class imbalance and increased data diversity, a decline in model performance was observed. However, this reflects a more realistic evaluation by preventing data leakage from similar ECG segments appearing in both training and test sets. Patient-independent segmentation forced the model to generalize beyond individual patterns, a key requirement for real-world applications. The observed accuracy drop highlights the importance of robust evaluation protocols, especially in biomedical machine learning. Among the evaluated classifiers, DT, RF, and SVM, the RF achieved the highest accuracy of 91%.

A limitation of this study is that the evaluation was conducted only on the MIT-BIH dataset. Although this is a widely used criterion, its generalizability to other datasets has not been validated. Furthermore, although Gaussian noise augmentation was used to simulate variability and improve robustness, this type of noise may not accurately represent common types of noise in clinical ECG data (e.g., baseline drift, power line interference, or muscle abnormalities).

Future work should explore advanced data augmentation techniques, such as wavelet transformations or synthetic ECG generation using GANs, and investigate deep learning methods for a broader evaluation of AI models. While the MIT-BIH dataset is widely used for benchmarking, it may not fully capture ECG variability across clinical populations. We plan to validate the proposed approach on additional datasets, such as PTB-XL or real clinical recordings, to assess its real-world accuracy. Likewise, in this study, we chose SMOTE as a widely adopted oversampling method to address class imbalance. However, alternative undersampling strategies such as Tomek Links, NearMiss, One-Sided Selection (OSS), and Edited Nearest Neighbors (ENN) should be explored. In our future work, we intend to combine and compare these undersampling methods to determine their impact on model performance and generalizability. This will help determine the most appropriate balancing strategy for the characteristics of ECG data and medical diagnostic needs. Finally, deploying these models in clinical settings will be crucial to ensure their reliability and effectiveness.

References

- [1] Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*, 415, 190–198.
- [2] Ahmed, S. R., Ahmed, A. K., & Jwmaa, S. J. (2023). Analyzing the employee turnover by using decision tree algorithm. 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 1–4. IEEE. <https://doi.org/10.1109/HORA58378.2023.10156709>
- [3] Ansari, Y., Mourad, O., Qaraqe, K., & Serpedin, E. (2023). Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. *Frontiers in Physiology*, 14, 1246746. <https://doi.org/10.3389/fphys.2023.1246746>
- [4] Assaad, M. A., & Shakah, G. H. (2024). Optimizing Health Pattern Recognition Particle Swarm Optimization Approach for Enhanced Neural Network Performance. *Cihan University-Erbil Scientific Journal*, 8(2), 76–83. <https://doi.org/10.24086/issn.2519-6979>
- [5] Assaad, M., Boné, R., & Cardot, H. (2008). A new boosting algorithm for improved time-series forecasting with recurrent neural networks. *Information Fusion*, 9(1), 41–55. <https://doi.org/10.1016/j.inffus.2006.10.009>
- [6] Atanasoski, V., Petrovic, J., Maneski, L. P., Miletic, M., Babic, M., Nikolic, A., ... Ivanovic, M. D. (2024). A Morphology-Preserving Algorithm for Denoising of EMG-Contaminated ECG Signals. *IEEE Open Journal of Engineering in Medicine and Biology*, 5, 296–305. <https://doi.org/10.1109/OJEMB.2024.3380352>
- [7] Atanasov, V. (2023). Application of Convolutional Neural Networks and Signal Processing Techniques to Identify and Filter Noise from ECG Signals. 2023 18th Conference on Electrical Machines, Drives and Power Systems (ELMA), 1–5. IEEE. <https://doi.org/10.1109/ELMA58392.2023.10202451>

- [8] Baker, M. R., Mohammed, E. Z., & Jihad, K. H. (2022). Prediction of colon cancer related tweets using deep learning models. *International Conference on Intelligent Systems Design and Applications*, 522–532. Springer. https://doi.org/10.1007/978-3-031-27440-4_50
- [9] Baker, M. R., Taher, Y. N., & Jihad, K. H. (2023). PREDICTION OF PEOPLE SENTIMENTS ON TWITTER USING MACHINE LEARNING CLASSIFIERS DURING RUSSIAN AGGRESSION IN UKRAINE. *Jordanian Journal of Computers & Information Technology*, 9(3), 189–206. <https://doi.org/10.5455/jcit.71-1676205770>
- [10] Balla, A., Habaebi, M. H., Elsheikh, E. A. A., Islam, M. R., & Suliman, F. M. (2023). The effect of dataset imbalance on the performance of SCADA intrusion detection systems. *Sensors*, 23(2), 758. <https://doi.org/10.3390/s23020758>
- [11] Camuto, A., Willetts, M., Simsekli, U., Roberts, S. J., & Holmes, C. C. (2020). Explicit regularisation in gaussian noise injections. *Advances in Neural Information Processing Systems*, 33, 16603–16614.
- [12] Darmawahyuni, A., Nurmaini, S., Rachmatullah, M. N., Tutuko, B., Sapitri, A. I., Firdaus, F., ... Predyansyah, A. (2022). Deep learning-based electrocardiogram rhythm and beat features for heart abnormality classification. *PeerJ Computer Science*, 8, e825.
- [13] Dias, M., Probst, P., Silva, L., & Gamboa, H. (2024). Cleaning ecg with deep learning: A denoiser tested in industrial settings. *SN Computer Science*, 5(6), 699. <https://doi.org/10.1007/s42979-024-03017-7>
- [14] Dou, B., Zhu, Z., Merkurjev, E., Ke, L., Chen, L., Jiang, J., ... Wei, G.-W. (2023). Machine learning methods for small data challenges in molecular science. *Chemical Reviews*, 123(13), 8736–8780. <https://doi.org/10.1021/acs.chemrev.3c00189>
- [15] Ferreir, J. L., Kumar, S., Soni, A., Acharya, N., & Acharya, S. (2023). Clinical management of cardiovascular diseases. In *Current Trends in the Diagnosis and Management of Metabolic Disorders* (pp. 254–278). CRC Press.
- [16] Guo, T., & Li, X. (2023). Machine learning for predicting phenotype from genotype and environment. *Current Opinion in Biotechnology*, 79, 102853. <https://doi.org/10.1016/j.copbio.2022.102853>
- [17] Gygi, J. P., Kleinstein, S. H., & Guan, L. (2023). Predictive overfitting in immunological applications: Pitfalls and solutions. *Human Vaccines & Immunotherapeutics*, 19(2), 2251830. <https://doi.org/10.1080/21645515.2023.2251830>
- [18] Habib, M., & Okayli, M. (2024). Evaluating the sensitivity of machine learning models to data preprocessing technique in concrete compressive strength estimation. *Arabian Journal for Science and Engineering*, 49(10), 13709–13727. <https://doi.org/10.1007/s13369-024-08776-2>
- [19] Haghighi, S., Jasemi, M., Hessabi, S., & Zolanvari, A. (2018). PyCM: Multiclass confusion matrix library in Python. *Journal of `Open-Source Software*, 3(25), 729.
- [20] Hameed, H. K. (2023). Deep learning Algorithm for predicting various retail store sales Using LSTM and ARIMA. *Dijlah Journal*, 6(4).
- [21] Hasanin, T., Khoshgoftaar, T. M., Leevy, J. L., & Seliya, N. (2019). Examining characteristics of predictive models with imbalanced big data. *Journal of Big Data*, 6, 1–21.
- [22] Huang, Y., Liu, W., Yin, Z., Hu, S., Wang, M., & Cai, W. (2024). ECG classification based on guided attention mechanism. *Computer Methods and Programs in Biomedicine*, 257, 108454.
- [23] Jansen, B. J., Aldous, K. K., Salminen, J., Almerikhi, H., & Jung, S. (2023). Understanding audiences, customers, and users via analytics: An introduction to the employment of web, social, and other types of digital people data. Springer.
- [24] Kaggle. (2023). Arrhythmia Database Modern. Retrieved from Kaggle website: <https://www.kaggle.com/datasets/protobioengineering/mit-bih/arrhythmia-database-modern-2023>
- [25] Ketu, S., & Mishra, P. K. (2022). Empirical analysis of machine learning algorithms on imbalance electrocardiogram-based arrhythmia dataset for heart disease detection. *Arabian Journal for Science and Engineering*, 47(2), 1447–1469.
- [26] Khalid, H. (2024). Modern techniques in detecting, identifying and classifying fruits according to the developed machine learning algorithm. *Journal of Applied Research and Technology*, 22(2), 219–229. <https://doi.org/10.22201/icat.24486736e.2024.22.2.2269>
- [27] Khalid, Hind. (2024). Efficient Image Annotation and Caption System Using Deep Convolutional Neural Networks. *BIO Web of Conferences*, 97, 103. EDP Sciences. <https://doi.org/10.1051/bioconf/20249700103>
- [28] Khalili, M., GholamHosseini, H., Lowe, A., & Kuo, M. M. Y. (2024). Motion artifacts in capacitive ECG monitoring systems: a review of existing models and reduction techniques. *Medical & Biological Engineering & Computing*, 62(12), 3599–3622. <https://doi.org/10.1007/s11517-024-03165-1>
- [29] Khan, F., Yu, X., Yuan, Z., & Rehman, A. U. (2023). ECG classification using 1-D convolutional deep residual neural network. *Plos One*, 18(4), e0284791.
- [30] Lee, J.-H., Zaheer, M. Z., Astrid, M., & Lee, S.-I. (2020). SmoothMix: a Simple Yet Effective Data Augmentation to Train Robust Classifiers. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3264–3274. <https://doi.org/10.1109/CVPRW50498.2020.00386>
- [31] Lekkas, G., Vrochidou, E., & Papakostas, G. A. (2025). Time–Frequency Transformations for Enhanced Biomedical Signal Classification with Convolutional Neural Networks. *BioMedInformatics*, 5(1), 7.
- [32] Li, H., Ditzler, G., Roveda, J., & Li, A. (2024). DeScod-ECG: Deep Score-Based Diffusion Model for ECG Baseline Wander and Noise Removal. *IEEE Journal of Biomedical and Health Informatics*, 28(9), 5081–5091. <https://doi.org/10.1109/JBHI.2023.3237712>
- [33] Li, S.-F., Huang, M.-L., & Wu, Y.-S. (2023). Combining the taguchi method and convolutional neural networks for arrhythmia classification by using ecg images with single heartbeats. *Mathematics*, 11(13), 2841.
- [34] Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-class confusion matrix reduction method and its application on net promoter score classification problem. *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference*, 412–419. <https://doi.org/10.3390/technologies9040081>
- [35] Maturi, B., Dulal, S., Sayana, S. B., Ibrahim, A., Ramakrishna, M., Chinta, V., ... Ravipati, H. (2025). Revolutionizing Cardiology: The Role of Artificial Intelligence in Echocardiography. *Journal of Clinical Medicine*, 14(2), 625. <https://doi.org/10.3390/jcm14020625>
- [36] Mir, H. Y., & Singh, O. (2024). Powerline interference reduction in ECG signals using variable notch filter designed via variational mode decomposition. *Analog Integrated Circuits and Signal Processing*, 118(2), 317–328. <https://doi.org/10.1007/s10470-023-02200-9>
- [37] Muralidharan, K., Ramesh, A., Rithvik, G., Prem, S., Reghunaath, A. A., & Gopinath, M. P. (2021). 1D Convolution approach to human activity recognition using sensor data and comparison with machine learning algorithms. *International Journal of Cognitive Computing in Engineering*, 2, 130–143.
- [38] Nagpal, A. K., Pundkar, A., Singh, A., & Gadkari, C. (2024). Cardiac arrhythmias and their management: An in-depth review of current practices and emerging therapies. *Cureus*, 16(8). <https://doi.org/10.7759/cureus.66549>
- [39] Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, 452, 48–62.
- [40] Oleiwi, Z. C., AlShemmary, E. N., & Al-augby, S. (2023). Efficient ECG Beats Classification Techniques for The Cardiac Arrhythmia Detection Based on Wavelet Transformation. *International Journal of Intelligent Engineering & Systems*, 16(2).
- [41] Oleiwi, Z. C., AlShemmary, E. N., & Al-Augby, S. (2023). Arrhythmia Detection Based on New Multi-Model Technique for ECG Inter-Patient Classification. *International Journal of Online & Biomedical Engineering*, 19(12).
- [42] Oleiwi, Z. C., AlShemmary, E. N., & Al-Augby, S. (2024). Developing Hybrid CNN-GRU Arrhythmia Prediction Models Using Fast Fourier Transform on Imbalanced ECG Datasets. *Mathematical Modelling of Engineering Problems*, 11(2).

- [43] Owusu, E., Quainoo, R., Mensah, S., & Appati, J. K. (2023). A deep learning approach for loan default prediction using imbalanced dataset. *International Journal of Intelligent Information Technologies (IJIT)*, 19(1), 1–16. <https://doi.org/10.4018/IJIT.318672>
- [44] Rai, H. M., & Chatterjee, K. (2022). Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data. *Applied Intelligence*, 52(5), 5366–5384.
- [45] Ramkumar, M., Babu, C. G., Priyanka, G. S., & Kumar, R. S. (2021). Ecg arrhythmia signals classification using particle swarm optimization-support vector machines optimized with independent component analysis. *IOP Conference Series: Materials Science and Engineering*, 1084(1), 12009. IOP Publishing.
- [46] Rashid, T. A., Shekho Toghranchi, C. I., Sindi, H., Alsadoon, A., Bačanin, N., Umar, S. U., ... Mohammadi, M. (2021). An improved BAT algorithm for solving job scheduling problems in hotels and restaurants. *Artificial Intelligence: Theory and Applications*, 155–171.
- [47] Ren, H., Sun, Q., Xiao, Z., Yu, M., Wang, S., Yuan, L., ... Yang, H. (2025). Heterogeneous feature fusion based machine learning strategy for ECG diagnosis. *Expert Systems with Applications*, 126714.
- [48] Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., & Gehler, P. (2022). Towards total recall in industrial anomaly detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14318–14328.
- [49] Schlenger, J. (2024). Random forest. In *Computer Science in Sport: Modeling, Simulation, Data Analysis and Visualization of Sports-Related Data* (pp. 201–207). Springer. https://doi.org/10.1007/978-3-662-68313-2_24
- [50] Sengupta, S., & Das, S. (2024). Analyzing Electrocardiogram Data-Statistical Insights Into Cardiac Health. In *Revolutionizing Healthcare Treatment With Sensor Technology* (pp. 126–154). IGI Global.
- [51] Singh, N., & Singh, P. (2019). Cardiac arrhythmia classification using machine learning techniques. *Engineering Vibration, Communication and Information Processing: ICoEVCi 2018, India*, 469–480. Springer.
- [52] Tariq, M., Palade, V., Ma, Y., & Altahhan, A. (2023). Diabetic retinopathy detection using transfer and reinforcement learning with effective image preprocessing and data augmentation techniques. In *Fusion of Machine Learning Paradigms: Theory and Applications* (pp. 33–61). Springer.
- [53] Umar, S. U., Rashid, T. A., Ahmed, A. M., Hassan, B. A., & Baker, M. R. (2024). Modified Bat Algorithm: a newly proposed approach for solving complex and real-world problems. *Soft Computing*, 28(13), 7983–7998. <https://doi.org/10.1007/s00500-024-09761-5>
- [54] Valkenborg, D., Rousseau, A.-J., Geubbelmans, M., & Burzykowski, T. (2023). Support vector machines. *American Journal of Orthodontics and Dentofacial Orthopedics*, 164(5), 754–757. <https://doi.org/10.1016/j.ajodo.2023.08.003>
- [55] Wodecki, J., Michalak, A., Wylomańska, A., & Zimroz, R. (2021). Influence of non-Gaussian noise on the effectiveness of cyclostationary analysis—Simulations and real data analysis. *Measurement*, 171, 108814. <https://doi.org/10.1016/j.measurement.2020.108814>
- [56] Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information*, 14(1), 54. <https://doi.org/10.3390/info14010054>
- [57] Zeng, W., Shan, L., Yuan, C., & Du, S. (2024). Advancing cardiac diagnostics: Exceptional accuracy in abnormal ECG signal classification with cascading deep learning and explainability analysis. *Applied Soft Computing*, 165, 112056. <https://doi.org/10.1016/j.asoc.2024.112056>.