ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

Estuscience – Se, 2024, 25 [3] pp. 427-441, DOI: 10.18038/estubtda.1466349

RESEARCH ARTICLE

MULTI-LEVEL CLASSIFICATION BASED ON DEEP LEARNING FOR ACCURATE RISK STRATIFICATION OF ARRHYTHMIAS

Evin ŞAHİN SADIK[1,*],

[1] Kütahya Dumlupınar University, Electrical and Electronics Enginnering Department, Kütahya, Turkey
evin.sahin@dpu.edu.tr - 0000-0002-2212-4210

Abstract

Arrhythmias, also known as irregular heartbeats, are important health problems that must be accurately identified to diagnose and treat cardiovascular disease. Within the scope of this study, a network for classifying arrhythmias, which are important in the diagnosis and treatment of cardiovascular diseases, was proposed by using one-dimensional convolutional neural network (1D CNN), one of the deep learning techniques. With the proposed 1D-CNN architecture, arrhythmia types and normal rhythm ECGs were subjected to a more detailed examination from general to specific according to urgency situations. In the classifications made, first of all, a binary classification was made and an evaluation was made as whether there was a life risk or not. In triple, quadruple and six-fold classification, the detection of arrhythmia status is detailed. More complex classifications have helped to define different types of arrhythmias in more detail. This study proposes a deep learning network for automatic identification and classification of arrhythmias and shows that different arrhythmia conditions can be diagnosed with a single network model by applying the proposed network structure to multi-class arrhythmia disorders.

Keywords

Cardiac Arrhythmias,
CNN,
Classification,
ECG

Time Scale of Article

Received :07 April 2024
Accepted : 20 August 2024
Online date : 30 September 2024

1. INTRODUCTION

The heart is the most vital organ for humans, and diseases that affect the heart and blood vessels are called cardiovascular diseases [1]. The term cardiovascular disease refers to a wide range of conditions that affect the heart and blood vessels [2]. According to the World Health Organization, 17.9 million deaths occurred from cardiovascular diseases in 2016. This number is expected to increase to 23.6 million in 2030 [3]. The heart's continuous electrical activity is essential for assessing its health and function. A medical procedure called an electrocardiogram (ECG) is used to observe and record the electrical activity of the heart. ECG is very important for diagnosing and following up heart diseases as it quickly evaluates the electrical rhythm and structure of the heart non-invasively.

Additionally, using ECG, diseases such as arrhythmia [4], coronary artery [5], myocardial infarction [6], pericarditis [7], cardiomyopathy [8], hypertrophic cardiomyopathy [9] and pulmonary embolism [10] can be diagnosed. Arrhythmia, being among these ailments, manifests in various forms, including AF, atrial flutter, ventricular tachycardia, bradycardia, supraventricular tachycardia (SVT), ventricular fibrillation, and premature atrial or ventricular contractions. Occasionally, arrhythmias cannot be diagnosed immediately using ECG alone [11]. Additionally, 20% of atrial fibrillation cases occur without any symptoms [12]. In such cases, doctors resort to additional methods such as Diagnostic Holter monitor, event monitor, stress test, electrophysiological test or echocardiogram [13]. When arrhythmias are not diagnosed and treated early, they can lead to serious health problems such as heart failure, stroke, cardiac arrest or cause symptoms that reduce the quality of life such as dizziness, fainting,

*Corresponding Author: evin.sahin@dpu.edu.tr

shortness of breath, chest pain and depression [14]. Early diagnosis and appropriate treatment can reduce the risk of complications from arrhythmia and significantly improve a patient's quality of life [15].

Nowadays, detecting arrhythmia from ECG signals using artificial intelligence techniques is a Machine Learning (ML) and Deep Learning (DL) techniques are employed to detect arrhythmias. A review of these research studies, focusing on work from the past twelve years, reveals that out of 40 studies, 29 (or 72.5%) utilized DL methods, nine (22.5%) applied ML methods, and two studies integrated both approaches for arrhythmia prediction [16]. Traditional methods for diagnosing arrhythmias with computer assistance rely on established medical rules but exploring artificial intelligence approaches can provide doctors with highly precise tools for diagnosing arrhythmias. Artificial intelligence algorithms could be embedded in smart ECG devices to help more people screen for arrhythmias early [17]. In classifications made using ML algorithms, the feature extraction part usually varies significantly in the success of the study, so different approaches are used for feature extraction [18]. One of the most essential features extracted is peak detection [19], [20]. Some of the other methods used are empirical mode decomposition, Pan Tompkins Algorithm, Hilbert and Wavelet Transform [21]–[23].

Literature reviews on the efficacy of deep learning (DL) algorithms versus traditional ML in diagnosing arrhythmias from electrocardiograms (ECGs) indicate that DL algorithms demonstrate enhanced performance in detecting and classifying ECG arrhythmias, surpassing the capabilities of conventional ML techniques. Deep learning methods offer superior performance with automatic feature selections without requiring manual feature selection and extraction [24]. DL algorithms are known to be effective in detecting conditions such as AF and Premature Ventricular Contraction (PVC), which rely on single heartbeats and require the identification of patterns between multiple beats [25]. Studies have been conducted in the literature using different DL methods to detect arrhythmia from ECGs [26], [27]. Studies have shown that artificial intelligence and DL techniques are increasingly used in ECG-based arrhythmia detection and that the use of DL algorithms in the healthcare sector is developing. Increasing success in the methods significantly reduces the pressure on doctors to analyze ECGs, resulting in significant advances in arrhythmia diagnosis.

Evaluation of existing methods does not show any significant difference in arrhythmia classification based on ECG signals. Most existing methods use multilayered and complex deep learning models. This leads to overfitting on limited and unbalanced data sets. Additionally, models trained on specific datasets have difficulty generalizing across different patient populations and varying ECG signal characteristics. The high computational requirements and long processing time of deep models limit timely decision making in clinical environments. Traditional methods often rely on manual feature extraction. Since it is not known which features perform best on the model, it can be time-consuming and may not effectively capture all relevant signal features. This study aims to address these gaps and proposes a simplified CNN architecture with two convolutional layers. The architecture is restricted to two convolutional layers, achieving a balance between model complexity and performance and reducing the risk of overfitting. When the simpler model structure is combined with regularization techniques such as batch normalization and dropout, the model's generalization ability across different datasets is enhanced. The reduced computational burden of the model makes it more suitable for real-time applications in clinical settings. By leveraging the automatic feature extraction capabilities of CNN, our approach eliminates manual intervention and is able to capture signal features more comprehensively.

Within the scope of this study, the data were subjected to four different classifications. In the first classification, arrhythmia types were classified as life-threatening arrhythmias and less dangerous conditions. These two categories differ markedly in terms of potential health risks and the need for immediate intervention. In addition, the data were examined in three and four classes to examine it more thoroughly. Finally, the performance metrics of the model were examined by classification for six different situations.

The proposed study aims to offer a layered approach according to the severity or urgency of arrhythmias and develop an auxiliary artificial intelligence system to help manage situations requiring rapid intervention more effectively.

It is aimed to prevent the development of a new model for each problem by applying the same network structure to different classification types. In this part of the study, a controversial analysis of the literature was made. In the remaining sections, information about the dataset in the material and method is given, the proposed method is mentioned, arrhythmia classification is mentioned and evaluation metrics are explained. The classification results were examined with graphs and tables in the evaluation results section. In the Conclusion section, a general evaluation of the study was made.

2. MATERIALS AND METHODS

In this study, a series of processing steps were applied to classify ECG signals. First, the raw ECG signal was divided into 2-second segments. The segmented data transformed into the frequency domain using the Fast Fourier Transform (FFT) method. The frequency spectra obtained from FFT were processed with a 1D Convolutional Neural Network (1D CNN), a DL based classification model. Figure 1 gives the implementation steps of the proposed model. Starting from the segmentation of the ECG signal, FFT application, processing with the CNN model and finally the classification results steps are visualized in the figure. This process shows how the model distinguishes different types of rhythms and how results are obtained for each classification group.

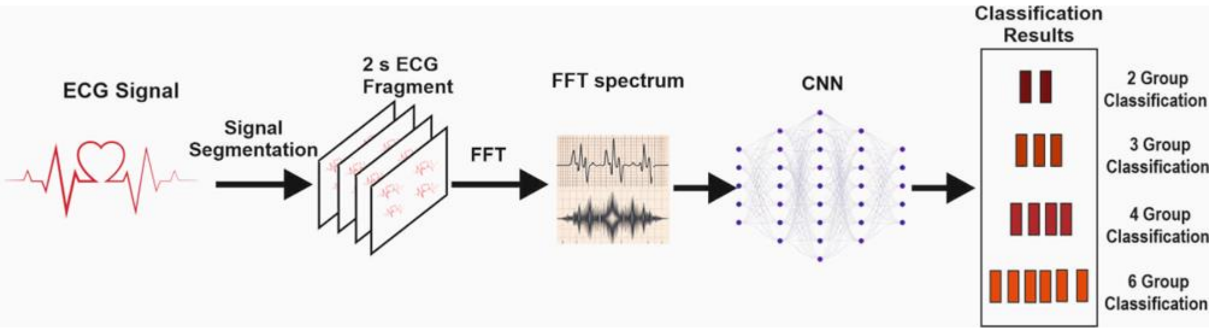

Figure 1. Application steps of the proposed model.

The classification process was performed in different scenarios based on clinical significance. The visual showing the classification scenarios and the data in these groups is given in Figure 2. The first of these classification scenarios is to divide the data into two groups: life-threatening arrhythmias and less dangerous or normal rhythms. Secondly, the data were divided into three groups. These are ventricular arrhythmias (more dangerous), potentially dangerous arrhythmias, supraventricular and normal rhythms. In another classification, the data is divided into four groups. In this classification, very dangerous ventricular arrhythmias, threatening ventricular arrhythmias, potentially dangerous arrhythmias, low-risk ventricular arrhythmias, and supraventricular and normal rhythms are discussed separately. In the final classification scenario, six groups were classified. Here, the entire ECG data set is divided into six separate groups and each group is classified within itself. Metrics like recall, specificity, F1 score, accuracy and the ROC curve were utilized to assess the performance of each classification scenario. These metrics were used to measure the potential effectiveness and reliability of our model in clinical practice. The model used in our study is CNN. In this model, the hyperparameters of the model were adjusted for each classification scenario and the model was classified into four different scenarios to provide the highest validation set performance.

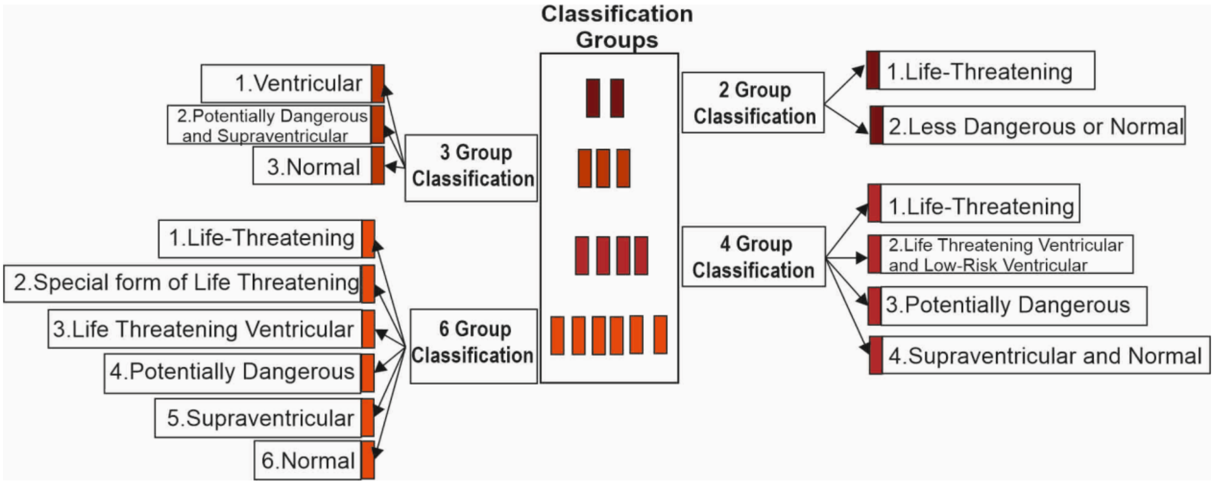

Figure 2. Diagram showing classification groups.

The figure shows how the classification scenarios are separated in our study. Binary classification aims to distinguish between life-threatening arrhythmias and less dangerous or normal rhythms. This distinction is critical to prioritizing emergency medical attention in clinical settings. Binary classification ensures that the model can accurately distinguish between critical categories. This allows patients requiring urgent treatment to be quickly identified.

The three-class classification scenario, in contrast to the two-class classification, categorizes ECG signals into normal rhythms, potentially dangerous and supraventricular arrhythmias, and ventricular arrhythmias. It helps prioritize patients according to the severity of their condition. Unlike binary classification, triple classification also classifies potentially dangerous arrhythmias.

Four-class and six-class classification scenarios aim to provide more detailed information for diagnosis and treatment by simultaneously classifying different arrhythmia types, capturing the entire spectrum of arrhythmia diversity.

2.1. Dataset

The first version of the arrhythmia dataset used in the study was the MIT-BIH Malignant Ventricular Ectopia Database (MVED) [28]. This database collected 48 half-hourly 2-channel ECG recordings from 47 people. The data in the MVED dataset was divided into 2-second segments and labeled according to arrhythmia types by expert cardiologists. It became the High-Risk Labeled ECG Fragment dataset in PhysioNet [29], [30].

The data contains ECG signals consisting of 6 different classes, one class containing normal and the other arrhythmia types. There are a total of 1016 ECG records in the data set presented in PhysioNet. Each recording is presented in 4 different ways. The first of these is recorded as 2-second fragments, while the second, the version used in this study, is the full spectrum of the 2-second signal containing the range 0-180 Hz in 0.5 Hz steps. The third signal is the smoothed spectrum signal called 15_2, from which 15 features are extracted in 1 Hz steps in the 0-15 Hz range. The fourth and last recording method is the signal containing 10 features extracted in 1.5 Hz steps from 0-15 Hz. Since the process was carried out with a deep learning model within the scope of the study, the entire spectrum of the 2-second signal was used. Since there was an unbalanced distribution between classes in the data set, the SMOTE algorithm was used to balance the data set.

2.2. Arrhythmia Classification

In this study, (1) Dangerous Arrhythmia, (2) The Early Form of Life-Threatening Arrhythmia, (3) Life-Threatening Ventricular Arrhythmia, (4) Potentially Dangerous Ventricular Arrhythmia, (5) Supraventricular Arrhythmia and (6) Normal Rhythm classification has been made. It is as given in Figure 3.

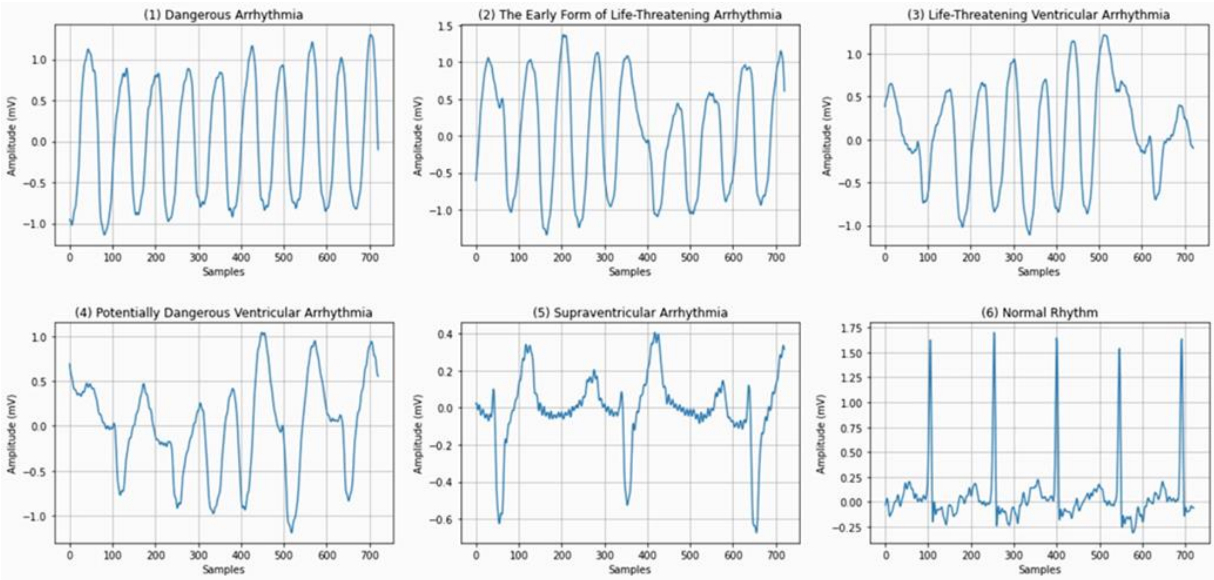

Figure 3. Types of ECG waves used for classification

The heart rhythm of healthy people is a regular ECG, known as normal sinus rhythm. Dangerous Arrhythmias are Arrhythmias that are Dangerous and Require Urgent Intervention. These arrhythmias have two subsections: Ventricular Flutter (VFL) and Ventricular Fibrillation (VF). In VFL, the lower chambers of the heart (ventricles) beat too fast and irregularly, resulting in the heart's inability to pump blood effectively. VF is an emergency condition in which the ventricles contract irregularly and ineffectively, requiring cardiopulmonary resuscitation (CPR) and defibrillation. (2) The Early Form of Life-Threatening Arrhythmia (TdP) is a special type of ventricular tachycardia. It is associated with prolongation of the QT interval and is potentially dangerous. (3) Life-Threatening Ventricular Arrhythmia (VT), Ventricular tachycardia, is the ventricles beating in an abnormal and rapid rhythm. Fast heartbeats can lead to the inability to pump blood effectively. (4) Potentially Dangerous Ventricular Arrhythmia, these rhythms are divided into Ventricular Bigeminy (B), High Grade Ventricular Ectopic Activity (HGEA) and Ventricular Escape Rhythm (VER). B: It is the occurrence of a premature ventricular contraction (PVC) after each normal beat. This creates a regular pattern. HGEA is a high degree of abnormal ventricular activity, such as frequent ventricular extrasystoles or consecutive PVCs. VER, A condition in which the ventricles beat in an abnormally slow rhythm. It usually occurs when the sinus node or AV node cannot produce a rhythm fast enough. (5) Supraventricular Arrhythmia is divided into these rhythms as Atrial Fibrillation (AFIB), Supraventricular Tachyarrhythmia (SVTA), Sinus Bradycardia (SBR), First Degree Heart Block (BI) and Nodal Rhythm (NOD). AFIB is irregular and rapid beating of the atria (upper chambers of the heart). It may increase the risk of blood clots and stroke. SVTA is a fast heart rhythm originating from the atrium or AV node. SBR is a slower-than-normal heart rate originating from the sinus node. BI is when each beat is delayed in passing through the AV node, but no beats are missed. NOD is the rhythm originating from the AV node and is usually slow. (6) These are divided into two in Normal Rhythm. Normal Sinus Rhythm (N) and Normal Rhythm with Extrasystole (Ne). N are heartbeats originating from the sinus node and occurring at normal intervals. The Ne is the normal rhythm but with occasional extra beats (extrasystoles). Table 1 shows the classes in the ECG Database and the number of records in the classes.

Table 1. Classes in the ECG Database and the number of records in the classes.

Type	Record
Class 1 (VFL, VF)	337
Class 2 (VTTdP)	72
Class 3 (VTHR)	169
Class 4 (VTLR, B, HGEA, VER)	132
Class 5 (AFIB, SVTA SBR, BI, NOD)	106
Class 6 (BBB, N, Ne)	200
Total	1016

2.3. Proposed Model

In our study, the CNN architecture used to classify ECG data is shown in detail in Figure 4. In the first step, ECG signals were given as input to the CNN. The convolution layer contains two consecutive convolution layers to extract features from ECG data. Each convolution layer performs feature extraction on the data using 5x5 filters. The purpose of this process is to detect important signal patterns in the ECG data. Regularization Layers, following convolution layers, a dropout layer is implemented to prevent overlearning and increase model generalization ability. The dropout rate was set to 0.5, meaning half of the neurons were randomly disabled at each training step. After the pooling layer, the size of the feature maps was reduced using a max pooling layer. Using a 2x2 window, it selects the maximum values from each feature map, thus reducing the data size. The outputs of the flattening and fully connection pooling layer were flattened and then transferred to a fully connected layer containing 200 neurons. This layer takes the flattened features and creates the final feature vector for classification.

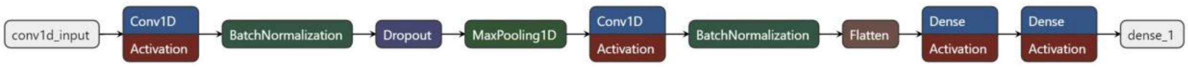

Figure 4. Structure of the proposed 1D-CNN Network.

The model was trained and evaluated for different classification scenarios. Each classification scenario corresponds to ECG signal classes that represent specific medical conditions. Model performance was evaluated using various metrics such as precision, recall, accuracy F1 score and Matthews Correlation Coefficient (MCC).

For the task at hand, which involves classifying ECG signals, we found that a simpler architecture with fewer layers was sufficient to achieve high accuracy. Adding more convolutional layers can lead to increased computational complexity and longer training times without significantly improving performance for such data. With limited training data, deeper networks are more prone to overfitting. While designing the CNN architecture chosen within the scope of the study, this architecture was preferred because a good balance between performance and calculation was not achieved in multi-layer networks. Using only two convolution layers, we aimed to balance model complexity and generalization ability. In our study, the hyperparameters used in the CNN architecture to classify arrhythmias from ECG data are learning rate, batch size, epoch size, dropout rate and optimizer. 0.001 was chosen as the learning rate. This value was found by performing a grid search to find the optimal value that effectively minimizes the loss function. 32 was used as a batch size. By balancing the batch size used, batch size, training speed and model performance, it was ensured that the gradient descent steps were neither too large nor too small. The model was trained with 150 epochs. This number is determined by where the training and validation losses meet. Thus, the data in the model was learned sufficiently and did not show overfitting. Additionally, a 50% dropout rate was applied to prevent overfitting. This value was chosen based on standard practices. RMSprop optimizer was used as the optimizer. This optimizer was chosen for its efficient handling of sparse gradients and adaptive learning rate features, which helped in faster convergence. These hyperparameters were selected by a combination of grid search and manual tuning, considering the model's computational efficiency and generalization ability. In classifying ECG signals, we found that a simpler architecture with fewer layers was sufficient to achieve high accuracy. Adding more convolution layers can increase computational complexity and training times without significantly improving performance for this type of data.

Since ECG signals are time series, the 1D-CNN method was chosen to process these signals in this study. Compared to other machine learning methods, 1D-CNN automatically extracts the features for classification. The model provides high accuracy with its simplicity and high performance, and it prevents overfitting. Thus, it stands out as a suitable network model for clinical studies.

2.4. Performance Evaluation Metrics

In our study, seven basic metrics were used to evaluate the performance of the CNN model used in classifying ECG signals. These are, respectively, confusion matrix [31], accuracy, precision, recall, F1 Score, MCC, and ROC Curve. Accuracy is calculated is given in Equation (1). Here, TP represents the number of true positives, that is, those that are positive and predicted as positive by the classifier, and FP represents the number of positives that are positive but not predicted as positive by the classifier. FN refers to the number of false negatives, that is, it shows the number of those that were not predicted as negative by the classifier when they were negative. It shows the situations that are negative in the classification given by FP but are predicted as positive by the classifier. The precision metric measures how many of the cases classified as positive are actually positive. This indicates the accuracy of the positive predictions produced by our model and is calculated by the formula given in Equation (2). Recall indicates how much of all true positive cases are correctly classified. It is calculated as given in Equation (3). F1 score is the harmonic mean of precision and recall and combines both metrics in a balanced manner. How it is calculated is given in Equation (4). It evaluates how well the model performs overall. MCC is a metric that measures the quality of a classification performance by considering all four confusion matrix values. Both binary and multiple classifications can be used. How it is calculated is given in Equation (5).

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

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

$$\text{Re}\,call = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1 Score=2}\times\frac{\text{Precision+Recall}}{\text{Precision}\times\text{Recall}} \tag{4}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{5}$$

3. FINDINGS AND DISCUSSION

Four different classifications were made within the scope of the study. With the help of CNN network, arrhythmias were classified according to different conditions. Figure 5 shows the training process and losses of classification models with different numbers of groupings. Figure 5 a) shows the training and loss graph for binary classification. While the accuracy of the model increased over time, the loss value decreased. It is seen that the model showed a steady improvement during the training process and reached a high accuracy rate without overfitting at the end of approximately 150 epochs. Figure 5 b) shows the training and loss graph for classifying ECG signals belonging to 3 different groups with the CNN network. In this graph, accuracy increased and loss decreased as the number of epochs increased. However, fluctuations in the loss graph may indicate that the model is starting to overfit the training data at specific points. Training and loss for 4 Group Classification in 5 c) the accuracy rate fluctuates slightly but generally increases. The loss rate is in a downward trend but decreasing more slowly. Given in Figure 5 d) Training and Loss for 6 Group Classification. This model has a lower accuracy rate and a higher loss value than others. This suggests that the model faces a more complex classification task and may need more training epochs or model tuning. Each graph tracks both accuracy and loss values throughout training to measure the model's performance. Such graphs are an important tool in understanding the training process of the model and adjusting the hyperparameters.

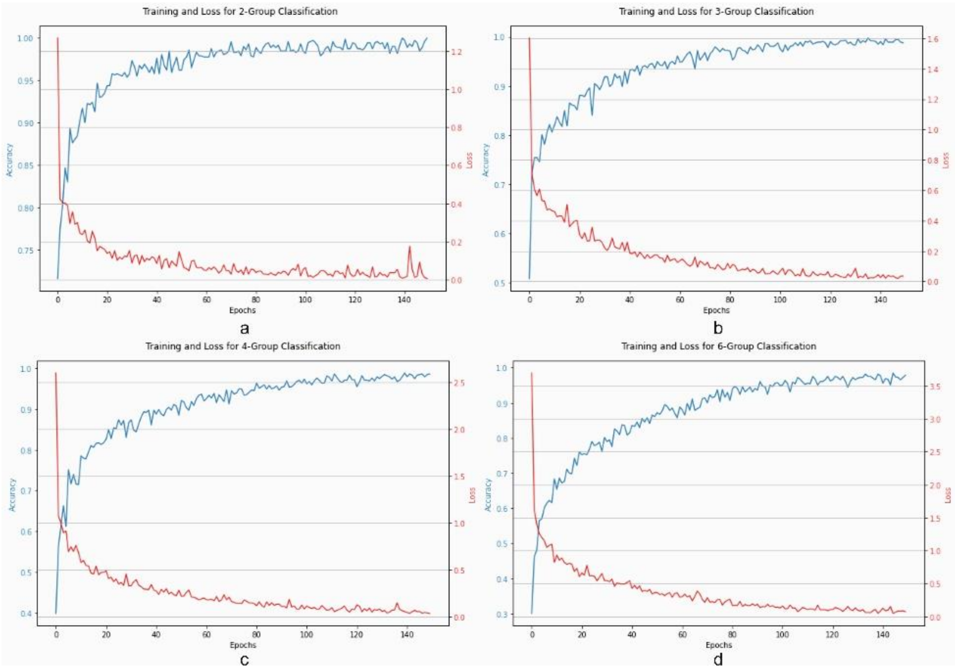

Figure 5. Training and loss graphs a) For 2-class classification b) For 3-class classification c) For 4-class classification d) For 6-class classification.

Separate confusion matrices were calculated for four different situations with different class numbers and are given in Figure 6. Figure 6 shows the confusion matrices that visualize the performance of the model for classification tasks with 2 classes in (a), 3 classes in (b), 4 classes in (c) and 6 classes in (d). Each matrix shows how well the model predicts for each class. In figure 6 (a), the predictions for Class 0 and Class 1 are largely correct. For Class 0, there are 161 correct predictions and 8 false negatives, and for Class 1, there are 173 correct predictions and 5 false positives. This indicates high accuracy and recall rates. In figure 6 (b), Class 0 showed high accuracy with 162 correct and 4 incorrect predictions. Class 1 demonstrated very good performance with 183 correct predictions and minimal confusion (1 false positive and 1 false negative). Class 2 also performed well with 151 correct predictions, but some confusion is seen with 12 false negatives and 4 false positives. In figure 6 (c), the model performs relatively well for Class 0 and Class 3, with 87 and 84 correct predictions, respectively. However, it experiences more confusion (false positives and false negatives) for Class 1 and Class 2. Specifically, Class 0 has 23 false positives, and Class 1 has 4 false negatives. In figure 6 (d), high accuracy is observed in Class 0 and Class 5, with 87 and 81 correct predictions, respectively. However, significant confusion exists among other classes. For example, Class 0 has 16 false negatives, and Class 1 has 10 false

negatives, while Class 4 has 7 false negatives and 5 false positives. The number of false positives and false negatives, especially in Classes 2, 3, and 4, is notable.
It has been observed that there is a general decrease in the performance of the model as the number of classes in the confusion matrices increases. This shows that the difficulty of the classification task increases as the number of classes increases and the model finds it difficult to manage the complexity. High rates of false positives and false negatives indicate that the model may confuse certain classes with other classes, which may indicate that the boundaries between classes are unclear or have insufficient representation in the training data set. Since the data distribution in this dataset is unbalanced, it is thought that it does not show high success in multiple classification.

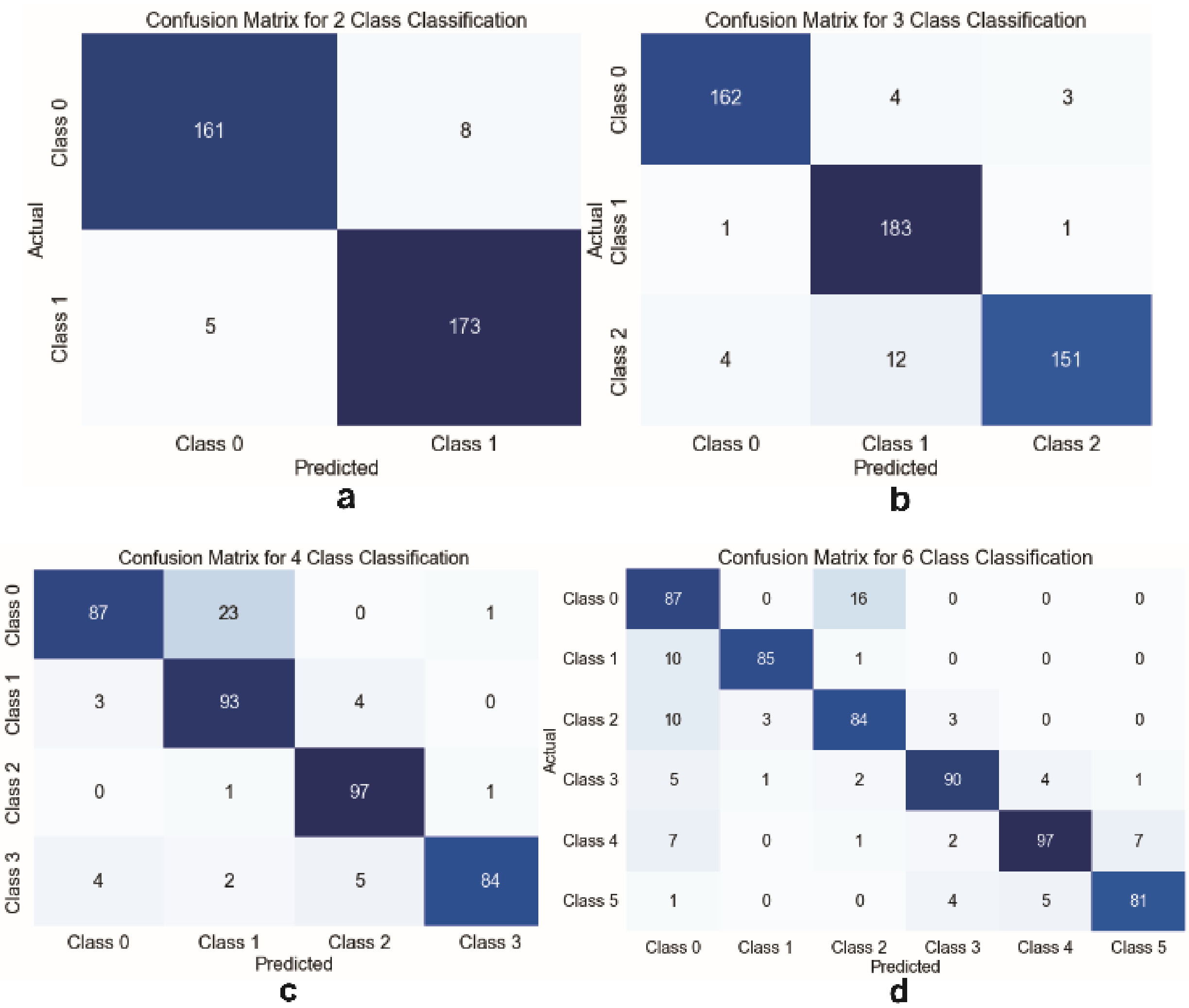

Figure 6. Confusion matrices for four different classification results: a) For 2-class classification b) For 3-clas classification c) For 4-class classification d) For 6-class classification.

We also assess our model's performance using the ROC curve, which illustrates the correlation between the true positive rate (TPR) and the false positive rate (FPR) across various possible threshold settings of the model. The area under the curve (AUC) numerically expresses how well the model performs. In Figure 7, separate ROC curves are drawn for the 4 different classification cases performed in this study. When the graphs are examined, Figure 7 shows the ROC curve of the model for a classification task. The AUC value is 0.99 and the model showed very high performance. Figure 7 b) contains the ROC curves for a three-class task. There is a separate curve for each class, and each has high AUC values, meaning that the model appears to discriminate all classes well. Figure 7c shows AUC values for four classes and shows high-performance curves for each class. Figure 7 d) shows ROC curves and AUC values for six different classes. This task is more complex and models generally perform less well with more classes, but the model shown here has quite high AUC values.

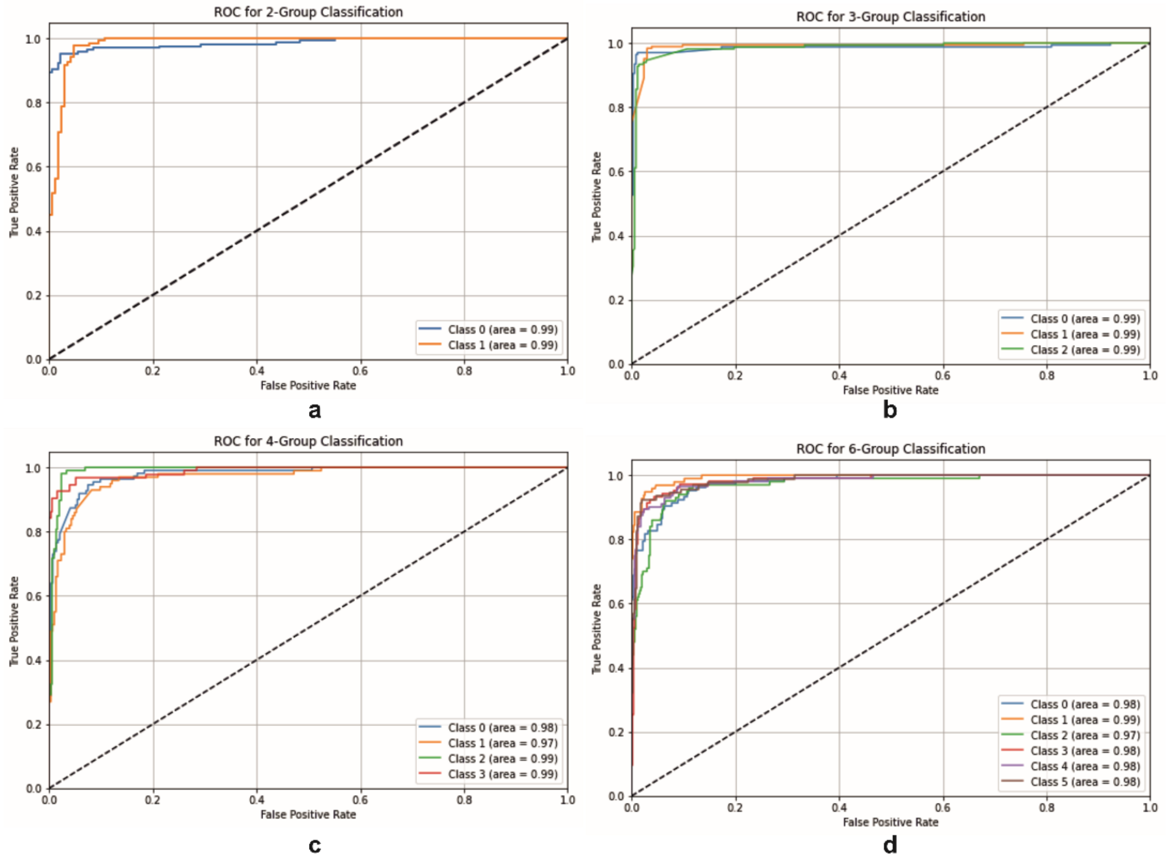

Figure 7. ROC curves for different classifications. a) For 2-class classification b) For 3-class classification c)For 4-class classification and d) 6-class classification.

When the metrics of each classification results are evaluated, the results of the two-class classification are given in Table 2. When the results are evaluated, it clearly shows that the model has high performance. Precision, Recall, F1-score, AUC, Accuracy and MCC values are all quite high, showing that the model works successfully for both classes. These results show that the model detects true positives well and the false positive rate is low. Overall, the performance of the model is near perfect and can be used safely in classification tasks.

Table 2. Classification metrics for two-class classification

Classes	Precision	Recall	F1-score	AUC	Accuracy	MCC
Class0	0.97	0.95	0.96	0.99	0.96	0.93
Class1	0.96	0.97	0.96	0.99		

Both classes in the binary classification exhibit equally high F1-scores, demonstrating the model's balanced and effective performance in distinguishing between the two classes. Three-class classification results are given in Table 3. When the table is examined, it is clearly seen that the model has high performance. Precision, Recall, F1-score, AUC, Accuracy and MCC values are all quite high, showing that the model works successfully for all three classes. Especially high recall and F1-score values for Class 1 show that the model is quite effective in this class. Overall, the performance of the model is close to perfect and it can be used safely in classification tasks.

Table 3. Classification metrics for three-class classification.

Classes	Precision	Recall	F1-score	AUC	Accuracy	MCC
Class0	0.97	0.96	0.96	0.99		
Class1	0.92	0.99	0.95	0.99	0.95	0.93
Class2	0.97	0.90	0.94	0.99		

The high F1-scores across all classes in the three-class classification indicate that the model is well-calibrated and performs consistently across different classes, with minimal performance degradation. The four-class classification results are given in Table 4. The overall performance of the model is quite high. Precision, Recall, F1-score, AUC, Accuracy and MCC values are high, showing that the model works successfully for all four classes. However, some metrics for Class 0 and Class 1 appear to be slightly lower than other classes. Especially the recall value for Class 0 being 0.78 shows that there are more false negatives in this class. These results show that the overall performance of the model is quite good, but it may need improvement in some classes.

Table 4. Classification metrics for four-class classification

Classes	Precision	Recall	F1-score	AUC	Accuracy	MCC
Class0	0.93	0.78	0.85	0.98		
Class1	0.78	0.93	0.85	0.97	0.89	0.86
Class2	0.92	0.98	0.95	0.99		
Class3	0.98	0.88	0.93	0.99		

The six-class classification results are given in table 5. When the model results are examined, it can be seen that it generally performs well. This model, where most of the Precision, Recall, F1-score, AUC, Accuracy and MCC values are high, shows that it works successfully in all six classes. It can be seen that some metrics for Class 0 and Class 2 are slightly lower than other classes. Especially the low precision and F1-score values for Class 0 indicate that there are more false positives in this class. These results show that the overall functioning of the model is good, but it may need improvement in some classes.

Table 5. Classification metrics for six-class classification

Classes	Precision	Recall	F1-score	AUC	Accuracy	MCC
Class0	0.72	0.84	0.78	0.98		
Class1	0.96	0.89	0.92	0.99		
Class2	0.81	0.84	0.82	0.97	0.87	0.84
Class3	0.91	0.87	0.89	0.98		
Class4	0.92	0.87	0.89	0.98		
Class5	0.91	0.87	0.89	0.98		

A detailed evaluation of the average key metrics for each classification group is provided in Table 6. It is clearly seen that the performance of the model in classification tasks decreases as the number of classes increases. In two- and three-class classifications, the model exhibits near-perfect performance with high accuracy, precision, sensitivity, F1-Score and MCC values. However, these metrics drop slightly in four- and six-class classifications. This shows that as the number of classes increases, the model's ability to distinguish classes becomes more difficult and the complexity increases. Nevertheless, the overall performance of the model across all classification groups is satisfactory, indicating that the model is reliable and consistent. These results show that the model performs reasonably even on more complex classification tasks.

Table 6. Comparative table showing average metrics for all classification groups.

	Accuracy	Precision	Recall	F1-Score	MCC
For 2 Class	0.96	0.96	0.96	0.96	0.93
For 3 Class	0.95	0.95	0.95	0.95	0.93
For 4 Class	0.89	0.90	0.89	0.89	0.86
For 6 Class	0.87	0.87	0.86	0.87	0.84

When the results obtained in the study were compared with other studies in the literature, the results in Table 7 were obtained. When the table is examined, studies compare the results obtained using different databases and methods for classifying ECG signals. The results show that the proposed methodology provides high accuracy rates compared to existing methods. The high accuracy rates obtained especially in two-class and three-class classification reveal that this approach is a potentially effective classification method.

Table 7. Comparison table with other existing methods.

Study	Database	Class	Method	Accuracy(%)
Acharya et al. (2013) [37]	MIT-BIH arrhythmia database, MIT-BIH malignant ventricular arrhythmia database, Creighton University ventricular tachyarrhythmia database	2	CNN	93.18
Mathews et al. (2018**)** [32]	MIT-BIH Arrhythmia Database	2	RBM & DBN	93.78
Yao et al.,(2018) [34]	China Physiological Signal Challenge	9	TI-CNN	77.3
Tripaty et al., (2018) [36]	Creighton university ventricular tachy-arrhythmia database (CUDB) and MIT-BIH malignant ventricular arrhythmia database	2	LS-SVM	89.81
Ebrahimzadeh et al. (2019) [38]	MIT-BIH Sudden Cardiac Death Holter and Sudden Cardiac Death Holter	2	Mixture of Experts (ME)	82.85
Yao et al., (2020) [35]	China Physiological Signal Challenge	9	ATI-CNN	81.2
Prabhakararao et al. (2021) [33]	PhysioNet/CinC-2017	3	DMSCE	88
	PTBXL-2020	5		85.65
Popadina et al. (2023) [39]	MIT-BIH Malignant Ventricular Ectopy Database (MVED)	2	1D CNN	95
This Study	MIT-BIH Malignant Ventricular Ectopy Database (MVED)	6	1D CNN	87
		4	1D CNN	89
		3	1D CNN	95
		2	1D CNN	96

4. CONCLUSIONS AND RECOMMENDATIONS

Within the scope of this study, we focused on proposing a CNN architecture and classifying different arrhythmia types according to their risk level. The effectiveness of the CNN architecture in the study in classifying cardiac arrhythmias was examined in detail. Our results show that CNN-based models provide significant accuracy in identifying and classifying cardiac arrhythmias. Accuracy rates of 96% in binary classification, 95% in triple classification, 89% in quadruple classification and 87% in six-fold classification were achieved. These results show that overall accuracy decreases as the complexity of the model increases, but high levels of accuracy are still maintained.

The high performance at the binary classification level reveals that the CNN can effectively distinguish basic arrhythmia types. On the other hand, more detailed levels of classification allow specific types of arrhythmias to be recognized in more detail. Study results highlight the potential of CNN-based DL models in automatic identification and classification of cardiac arrhythmias and encourage the use of these technologies in the field of cardiology.

In conclusion, this study demonstrates the feasibility and effectiveness of deep learning techniques in cardiac arrhythmia classification. These approaches may play an important role in the diagnosis and treatment of cardiovascular diseases in the future and may support clinical decision-making processes. However, additional studies on larger and more diverse data sets will help further develop these models and find broader applications in clinical settings.

CONFLICT OF INTEREST

The author stated that there are no conflicts of interest regarding the publication of this article.

CRediT AUTHOR STATEMENT

Evin Şahin Sadık: Writing - original draft, Writing – Review & Editing, Visualization, Conceptualization, Methodology, Software.

REFERENCES

[1] Tada H, Fujino N, Nomura A, et al. Personalized medicine for cardiovascular diseases. J Hum Genet. 2021;66(1):67-74.

[2] Gaidai O, Cao Y, Loginov S. Global cardiovascular diseases death rate prediction. Curr Probl Cardiol. Published online 2023:101622.

[3] Zheng Q, Tang Q, Wang ZL, Li Z. Self-powered cardiovascular electronic devices and systems. Nat Rev Cardiol. 2021;18(1):7-21.

[4] Zheng Q, Tang Q, Wang ZL, Li Z. Self-powered cardiovascular electronic devices and systems. Nat Rev Cardiol. 2021;18(1):7-21.

[5] Buja LM, Willerson JT. The role of coronary artery lesions in ischemic heart disease: insights from recent clinicopathologic, coronary arteriographic, and experimental studies. Hum Pathol. 1987;18(5):451-461.

[6] Reed GW, Rossi JE, Cannon CP. Acute myocardial infarction. Lancet. 2017;389(10065):197-210.

[7] LeWinter MM. Acute pericarditis. N Engl J Med. 2014;371(25):2410-2416.

[8] Wexler R, Elton T, Pleister A, Feldman D. Cardiomyopathy: an overview. Am Fam Physician. 2009;79(9):778.

[9] Maron BJ, Maron MS. Hypertrophic cardiomyopathy. Lancet. 2013;381(9862):242-255.

[10] Agnelli G, Becattini C. Acute pulmonary embolism. N Engl J Med. 2010;363(3):266-274.

[11] Gawałko M, Balsam P, Lodziński P, et al. Cardiac arrhythmias in autoimmune diseases. Circ J. 2020;84(5):685-694.

[12] Yusupova NI, Bogdanov MR. Assessment of the Risk of Sudden Cardiac Death Using Machine Learning Methods. Pattern Recognit Image Anal. 2023;33(3):536-543.

[13] Galli A, Ambrosini F, Lombardi F. Holter monitoring and loop recorders: from research to clinical practice. Arrhythmia Electrophysiol Rev. 2016;5(2):136.

[14] Mikhaylov AY, Yumashev A V, Kolpak E. Quality of life, anxiety and depressive disorders in patients with extrasystolic arrhythmia. Arch Med Sci AMS. 2022;18(2):328.

[15] Nie L-Y, Wang X-D, Zhang T, Xue J. Cardiac complications in systemic sclerosis: early diagnosis and treatment. Chin Med J (Engl). 2019;132(23):2865-2871.

[16] Boulif A, Ananou B, Ouladsine M, Delliaux S. A Literature Review: ECG-Based Models for Arrhythmia Diagnosis Using Artificial Intelligence Techniques. Bioinform Biol Insights. 2023;17:11779322221149600.

[17] Liu J, Li Z, Jin Y, et al. A review of arrhythmia detection based on electrocardiogram with artificial intelligence. Expert Rev Med Devices. 2022;19(7):549-560.

[18] Aziz S, Ahmed S, Alouini M-S. ECG-based machine-learning algorithms for heartbeat classification. Sci Rep. 2021;11(1):18738.

[19] Sahoo S, Dash M, Behera S, Sabut S. Machine learning approach to detect cardiac arrhythmias in ECG signals: A survey. Irbm. 2020;41(4):185-194.

[20] Mohanty M, Sahoo S, Biswal P, Sabut S. Efficient classification of ventricular arrhythmias using feature selection and C4. 5 classifier. Biomed Signal Process Control. 2018;44:200-208.

[21] Rodríguez R, Mexicano A, Bila J, Ponce R, Cervantes S, Martinez A. Hilbert transform and neural networks for identification and modeling of ECG complex. In: Third International Conference on Innovative Computing Technology (INTECH 2013). IEEE; 2013:327-332.

[22] Martis RJ, Krishnan MMR, Chakraborty C, et al. Automated screening of arrhythmia using wavelet based machine learning techniques. J Med Syst. 2012;36:677-688.

[23] Devi RL, Kalaivani V. Machine learning and IoT-based cardiac arrhythmia diagnosis using statistical and dynamic features of ECG. J Supercomput. 2020;76(9):6533-6544.

[24] Şahin Sadık E, Saraoğlu HM, Canbaz Kabay S, Keskinkılıç C. Deep Learning-Based Approach for Classification Of Mental Tasks From Electroencephalogram Signals. Avicenna J Neuro Psycho Physiol. 2023;10(1):15-21.

[25] Ansari Y, Mourad O, Qaraqe K, Serpedin E. Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. Front Physiol. 2023;14.

[26] Essa E, Xie X. An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification. ieee access. 2021;9:103452-103464.

[27] Parvaneh S, Rubin J, Babaeizadeh S, Xu-Wilson M. Cardiac arrhythmia detection using deep learning: A review. J Electrocardiol. 2019;57:S70-S74.

[28] Greenwald SD. The development and analysis of a ventricular fibrillation detector. Published online 1986.

[29] Manilo LA, Nemirko AP, Evdakova EG, Tatarinova AA. ECG Database for Evaluating the Efficiency of Recognizing Dangerous Arrhythmias. In: 2021 IEEE Ural-Siberian Conference on Computational Technologies in Cognitive Science, Genomics and Biomedicine (CSGB). IEEE; 2021:120-123.

[30] Nemirko A, Manilo L, Tatarinova A, Alekseev B, Evdakova E. ECG fragment database for the exploration of dangerous arrhythmia. Published online 2022.

[31] Heydarian M, Doyle TE, Samavi R. MLCM: Multi-label confusion matrix. IEEE Access. 2022;10:19083-19095.

[32] Mathews SM, Kambhamettu C, Barner KE. A novel application of deep learning for single-lead ECG classification. Comput Biol Med. 2018;99:53-62.

[33] Prabhakararao E, Dandapat S. Multi-scale convolutional neural network ensemble for multi-class arrhythmia classification. IEEE J Biomed Heal Informatics. 2021;26(8):3802-3812.

[34] Yao Q, Fan X, Cai Y, Wang R, Yin L, Li Y. Time-incremental convolutional neural network for arrhythmia detection in varied-length electrocardiogram. In: 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech. IEEE; 2018:754-761.

[35] Yao Q, Wang R, Fan X, Liu J, Li Y. Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network. Inf Fusion. 2020;53:174-182.

[36] Tripathy RK, Zamora-Mendez A, De la O Serna JA, Paternina MRA, Arrieta JG, Naik GR. Detection of life threatening ventricular arrhythmia using digital taylor fourier transform. Front Physiol. 2018;9:722.

[37] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput Biol Med. 2018;100:270-278.

[38] Ebrahimzadeh E, Foroutan A, Shams M, et al. An optimal strategy for prediction of sudden cardiac death through a pioneering feature-selection approach from HRV signal. Comput Methods Programs Biomed. 2019;169:19-36.

[39] Popadina AO, Nemirko AP. Detection of Dangerous Arrhythmias Using a 1D Convolutional Neural Network Trained on the Full Spectrum of a Short ECG Signal. In: 2023 Seminar on Digital Medical and Environmental Systems and Tools (DMEST). IEEE; 2023:109-112.