

Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks

ABDELRAHMAN M. SHAKER^{ID}, MANAL TANTAWI^{ID}, HOWIDA A. SHEDEED^{ID},
AND MOHAMED F. TOLBA^{ID}, (Senior Member, IEEE)

Department of Scientific Computing, Ain Shams University, Cairo 11566, Egypt

Corresponding author: Abdelrahman M. Shaker (abdelrahman.shaker@cis.asu.edu.eg)

ABSTRACT Electrocardiograms (ECGs) play a vital role in the clinical diagnosis of heart diseases. An ECG record of the heart signal over time can be used to discover numerous arrhythmias. Our work is based on 15 different classes from the MIT-BIH arrhythmia dataset. But the MIT-BIH dataset is strongly imbalanced, which impairs the accuracy of deep learning models. We propose a novel data-augmentation technique using generative adversarial networks (GANs) to restore the balance of the dataset. Two deep learning approaches—an end-to-end approach and a two-stage hierarchical approach—based on deep convolutional neural networks (CNNs) are used to eliminate hand-engineering features by combining feature extraction, feature reduction, and classification into a single learning method. Results show that augmenting the original imbalanced dataset with generated heartbeats by using the proposed techniques more effectively improves the performance of ECG classification than using the same techniques trained only with the original dataset. Furthermore, we demonstrate that augmenting the heartbeats using GANs outperforms other common data augmentation techniques. Our experiments with these techniques achieved overall accuracy above 98.0%, precision above 90.0%, specificity above 97.4%, and sensitivity above 97.7% after the dataset had been balanced using GANs, results that outperform several other ECG classification methods.

INDEX TERMS Class imbalance, convolution neural networks (CNNs), ECG classification, generative adversarial networks (GANs).

I. INTRODUCTION

An ECG is a standard tool for measuring the electrical activity of the heart and for diagnosing cardiac arrhythmias. Using an ECG involves placing electrodes on the surface of the body—such as the chest, neck, and arms—in order to detect electrical changes in the heart. An ECG record primarily consists of several distinctive wave forms, such as the P wave, the QRS complex, the T wave, and other wave forms. The P wave shows atrial contractions; the QRS complex shows ventricular contractions; the T wave shows the electrical activity produced as the ventricles are recharged for the next contraction [1]. Study of these complex waves and the cardiac activities they represent is vital for diagnosis of various arrhythmias [2]. It is difficult for a cardiologist to correctly analyze a large number of ECG records given their complexity and the amount of time required to analyze them [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang^{ID}.

Yet life-threatening arrhythmias need to be detected early and accurately [4].

Arrhythmias can be grouped into two main categories, life-threatening and non-life threatening. Life-threatening arrhythmias such as tachycardia and ventricular fibrillation cause heart attacks and sudden death [5], [6]. Non-life-threatening arrhythmias, which is our interest in this study, require attention in order to prevent deterioration of the heart muscle [3]. The category of the arrhythmia can be determined by recognizing the classes of consecutive heartbeats [7]. Manual beat-by-beat classification can be very time-consuming and too difficult in many scenarios. It is crucial to automate ECG analysis so that cardiac disorders can be discovered and treated as quickly as possible in clinical situations where speed in providing medical aid is essential.

Medical datasets like the MIT-BIH arrhythmia dataset are often very limited. They usually have data imbalance problem; they over-represent normal classes and common diseases and only sparsely represent rare diseases. Collecting medical

data is a complex and expensive process that requires the collaboration of cardiologists and researchers [8]. Imbalances in the datasets can make training the models—especially deep learning models—technically challenging, and the models will tend to be biased in favor of classes that contain large number of samples. The classification methods used in most studies tend not to perform well on minor classes because their aim is to optimize overall accuracy without giving appropriate consideration to the relative accuracy of each class [9]. Yet the cost of misclassifying minor classes in medical datasets is often much greater than the cost of misclassifying major classes, since the arrhythmias of high-risk patients usually fall in the minor classes of arrhythmia disease. The need for a good data augmentation technique for medical datasets is thus an urgent one.

One way to overcome the imbalance is to artificially create synthetic data by modifying the original training set using such classical data augmentation methods as translation, flipping, and rotation, which have become an essential step in computer vision tasks [10]. Although these modifications may result in marginal gains in diversity, they may also disrupt relevant orientation-related features, especially in such one-dimensional signals as those of an ECG.

The goal of generative models, the most promising models for data synthesis, is to learn the distribution of the training set and then to generate new samples from the learned distribution. In this paper, a specific kind of generative model called a Generative Adversarial Network (GAN), which has proved its effectiveness in synthesizing high-quality images in several domains [11], is used to generate synthetic heartbeats and thereby restore the balance of each class of the MIT-BIH dataset. The model generates realistic heartbeats that are very similar to actual heartbeats. Recently, GANs have been applied to balance a variety of medical imaging datasets, including generating MRI slices [12], retinal images [13], data for chest pathology [14], and data for bone lesion pathology [15]. To the best of our knowledge, this is the first attempt to apply the GANs to one-dimensional medical data.

In the last several years, deep learning has advanced rapidly. Its methods have shown promising results and remarkable success, providing state-of-the-art performance in such fields as bioinformatics [16], computer vision [17], and medical diagnosis [18]. In the literature, several approaches have been presented for automatic classification of heartbeats, including support vector machines (SVM) [19], [20], feed forward neural network (FNN) [21], back propagation neural network (BPNN) [22], general regression neural network (GRRN) [23], probabilistic neural network (PNN) [24], and recently different deep learning approaches have been utilized, including convolutional neural network (CNN) [25]–[28] and long short term memory (LSTM) [29], [30].

In this paper, we propose a novel data augmentation technique based on the combination of real and synthetic heartbeats using GAN to improve the classification of ECG heartbeats of 15 different classes from the MIT-BIH arrhythmia dataset. In addition, two approaches based on CNN are

proposed. The first approach (an end-to-end approach) classifies the heartbeats as one of the 15 classes in a direct way. The second approach (a two-stage hierarchical approach) classifies the heartbeats under one of the five main categories in the first stage, and each heartbeat is classified into one of the classes that belongs to that category in the second stage. To show the superiority of the GAN, we compared the results of the end-to-end approach after the dataset had been balanced using GAN and using other common data augmentation techniques. The proposed approaches are applied to lead I only from the MIT-BIH dataset to reduce the computational time. The contributions of this study are as follows: 1) the synthesis of high-quality heartbeats using GAN and 2) design of two deep-CNN approaches with superior performance compared with other studies.

The remainder of this paper is organized as follows: The related work is reviewed in section II. The proposed technique and methodologies are discussed in section III. The experimental results are presented in section IV. Finally, the conclusion and the future work are provided in section V.

II. RELATED WORK

The traditional approach to ECG classification is to develop an algorithm to extract the important features from the input signal and then choose an appropriate classifier to be used in the classification stage. The conventional approaches in most studies involve preprocessing, feature extraction, feature reduction, and classification. Many researchers in the literature have conducted studies on using such an approach for ECG classification [4], [31]–[44]. In these studies, the researchers considered different classes and utilized several methods in the feature extraction stage, such as independent component analysis (ICA), discrete wavelet transform (DWT), discrete cosine transform (DCT), principal component analysis (PCA), Gaussian mixture models (GMMs), higher order spectra (HOS), and the one-dimensional hexadecimal local pattern (1D-HLP). In the classification stage, they utilized several algorithms, including FNN, PNN, SVM, and BNN algorithms.

Although the studies described above achieved acceptable ECG classification performance, they have some disadvantages. For instance, the conventional approaches require developing a feature extractor and then reducing the extracted features to a set of optimal features that can be fed into an appropriate classifier. Computer-aided design (CAD) models developed using the above workflow show low performance when validated on a separate dataset and often suffer from overfitting [45]. Deep learning approaches have the capability of learning the most important features automatically from the input signals. Hence, the essential steps that are required in the traditional approaches, namely, feature extraction, feature reduction, and classification, can be developed with no need to be explicitly defined in the deep learning approaches. Recently, studies have applied several deep learning approaches for ECG classification.

Zhang *et al.* [27] proposed a CNN model consisting of six layers, comprising two convolutional layers, two pooling layers, and two fully connected layers. The model classifies five classes of the MIT-BIH dataset (Normal, Atrial Premature Contraction, Ventricular Premature Contraction, Right and Left Bundle Branch Blocks), and an overall accuracy of 97.50% is achieved. In [29], the authors explored the use of a DWT layer with bidirectional LSTM for ECG classification of five types of heartbeats obtained from the MIT-BIH arrhythmia dataset; they achieved an overall accuracy of 99.39%.

Acharya *et al.* [28] proposed a nine-layer CNN to classify five different categories of the MIT-BIH arrhythmia dataset. To overcome the imbalance in the number of heartbeats in the five (N, S, V, F, Q) categories, they generated synthetic heartbeats by varying the mean and standard deviation of the Z-score that was calculated from the original data. The number of heartbeats of the remaining types are increased to match the number of heartbeats of the N category. The researchers achieved an overall accuracy of 94.03% using the augmented data, and the overall accuracy was reduced to 89.07% when the model was trained only with the original data. The generated heartbeats are used in the training and testing phases. Tuncer *et al.* [31] proposed the use of DWT coupled with the 1D-HLP technique for automated arrhythmia detection. Ten-second segments of 17 ECG classes from the MIT-BIH dataset were considered and an overall accuracy of 95.0% was obtained using a 1 nearest neighborhood (1NN) classifier.

In [46], the researchers utilized 1,000 ECG fragments from the MIT-BIH arrhythmia dataset from lead 1, where each fragment was 10 seconds long. A deep one-dimensional (1D)-CNN consisting of 16 layers was proposed to classify 15 different classes, and an accuracy of 92.51% was achieved, whereas an accuracy of 91.33% was obtained when they considered 17 classes. Oh *et al.* [30] proposed a combination of CNN and LSTM for diagnosing five classes with variable length segments from the MIT-BIH arrhythmia dataset. The architecture consisted of six convolutional and pooling layers, followed by an LSTM layer and two fully connected layers. LSTM was used to extract the temporal information from the feature maps resulting from the convolutional layers. An accuracy of 98.10% was obtained.

Plawiak and Acharya [47] proposed a deep ensemble of classifiers for ECG classification based on deep learning approaches and genetic algorithms; the researchers used 10-second ECG segments from 29 people in the MIT-BIH arrhythmia dataset. Hence, they did not utilize the whole dataset, and only 744 segments from 29 out of 48 records are considered. The achieved overall accuracy for the utilized segments of the dataset was 95.00%. In [48], the authors proposed a convolutional autoencoder-LSTM system to automatically recognize five different types of arrhythmia heartbeats. They utilized the autoencoder to compress a large amount of ECG signals with a minimum loss and then classify the compressed signals using LSTM network. Finally, the

classification of the five classes from the MIT-BIH dataset achieved an accuracy of 99.21%.

Most studies have only considered overall accuracy. Yet, the overall accuracy as a parameter is not enough for measuring the robustness of the model because it is biased in favor of classes that contain large numbers of samples, while neglecting precision in the minor classes. Furthermore, the heartbeats of the minor classes need to be classified precisely because high-risk patients usually belong to these classes. In contrast, although [3], [43] considered the precision for the classes, they utilized data from leads 1 and 2, in addition to developing a fusion step to improve the results, which increased the computational time. Most studies also do not handle the imbalance problem in the MIT-BIH dataset, which negatively affects the achieved accuracy for classes with few heartbeats. In this study, we propose a novel approach to balance the classes of the utilized dataset using GAN, and we present two deep learning approaches to overcome the hand-crafted methods of feature extraction and reduction in the literature. The approaches classify 15 different classes from the MIT-BIH dataset using data only from lead 1, and multiple evaluation methods are considered.

III. METHODOLOGY

The proposed methods for preprocessing, data augmentation, and classification are discussed in this section. A detailed description of each method is introduced in the subsections below.

A. PREPROCESSING

The raw ECG signals are preprocessed to eliminate noise and improve the classification accuracy. The noise is eliminated by removing both high and low frequencies from the acquired ECG signal. A Butterworth bandpass filter with a range of 0.5–40 Hz is applied to extract the most valuable information from the ECG signal [49]. After that, ECG records are segmented into multiple heartbeats using the R-peak locations associated with the dataset; each heartbeat contains a P wave, QRS complex, and T wave.

Fixed segmentation methods are usually applied due to the difficulty of detecting the beginning and ending of each heartbeat [3], [4]. Yet, it is not always reliable because such an assumption cannot consider the variability of the heart rate. Hence, a dynamic segmentation strategy is used to overcome the heart rate variability, as proposed in [20]. To be invariant to the variability of the heart rate, the number of samples before and after each R peak are calculated according to the duration between the current and previous R peak (RR previous interval), as well as the duration between the current and next R peak (RR next interval). Thereafter, the number of samples in the larger interval is divided into three thirds; the first third is considered before the R peak, while the other two-thirds are considered after the R peak. Finally, the amplitudes of the heartbeats are normalized between 0 and 1; each heartbeat is resized to have 300 samples to unify the number

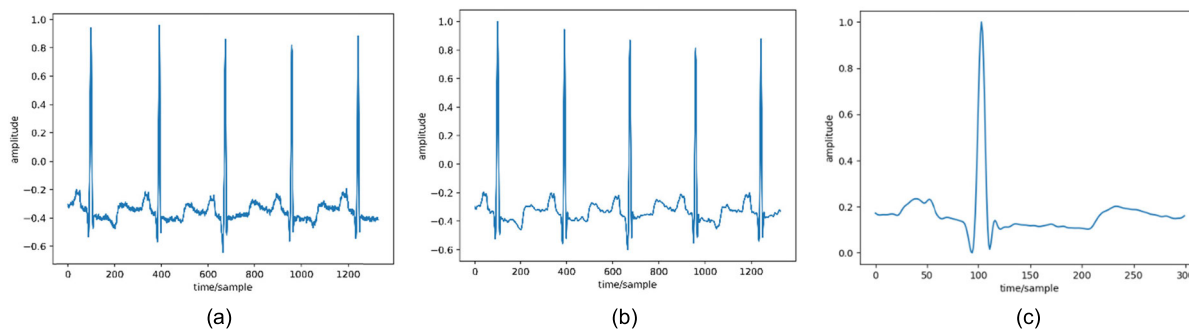


FIGURE 1. Preprocessing stage results (a) raw ECG record (b) filtered ECG record (c) segmented heartbeat.

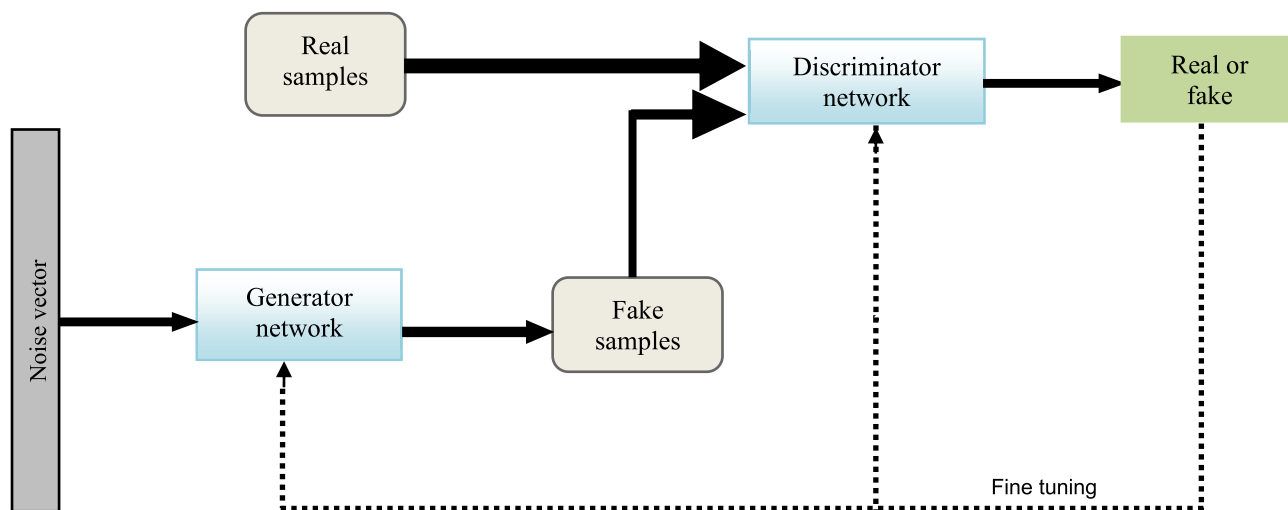


FIGURE 2. GAN training for the generator and the discriminator networks.

of samples for each heartbeat. The result of the preprocessing stage is shown in Figure 1.

B. DATA AUGMENTATION USING GAN

GAN consists of two neural networks—a generator and discriminator—and each network competes against the other. The generator network learns to map a noise vector to the distribution of the data it wants to generate; the goal of the generator is to produce data samples similar to the samples in the original dataset. In contrast, the discriminator network receives data samples from either the generated (fake) samples or the original (real) samples, and it is responsible for determining whether the received samples are real or fake. Figure 2 describes the training process of the GAN.

Two main problems arise when the GAN is trained with samples from all classes. The first problem is as follows: If the generator network in the GAN is trained to fool the discriminator network by generating realistic heartbeats, it will focus on the generation of the dominant classes to optimize the loss function of the network while collapsing away the other modes of the minor classes; this is known as the *mode collapse problem* [11]. However, such a problem may be partially tackled by using some advanced techniques, as proposed in [50]. The second problem is incurred if the GAN

is trained by using all heartbeats and somehow generates a diversity of fake samples; in such a case, the labels of the generated heartbeats cannot be determined precisely because some classes are highly similar. Hence, the GAN is trained using the heartbeats of each class independently to generate synthetic heartbeats to balance the training set for each class. The training of the GAN is terminated after the loss of the networks begins to saturate.

Synthetic heartbeats are generated after the segmentation stage. The generator network is trained on the segmented beats for each class, except the N class, because this is the dominant class. The number of samples in the other classes is increased to match the number of samples in the training set of the N class. Batch normalization [51] is used in the generator network to improve the performance and stability of the network and add diversity to the generated samples.

The generator network consists of four fully connected layers; it receives a vector of 100 random numbers sampled from standard uniform distribution as an input and outputs a heartbeat of a size of 300×1 . The discriminator network consists of five fully connected layers; it takes a heartbeat as an input of size 300×1 and outputs a decision on whether the heartbeat is real or fake. According to the decision, the parameters of the networks will be tuned to minimize the loss of the

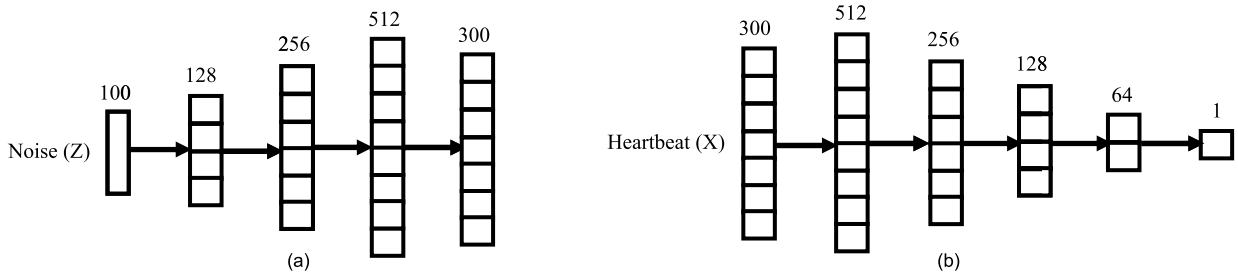


FIGURE 3. Proposed architecture of the GAN: (a) generator network, (b) discriminator network.

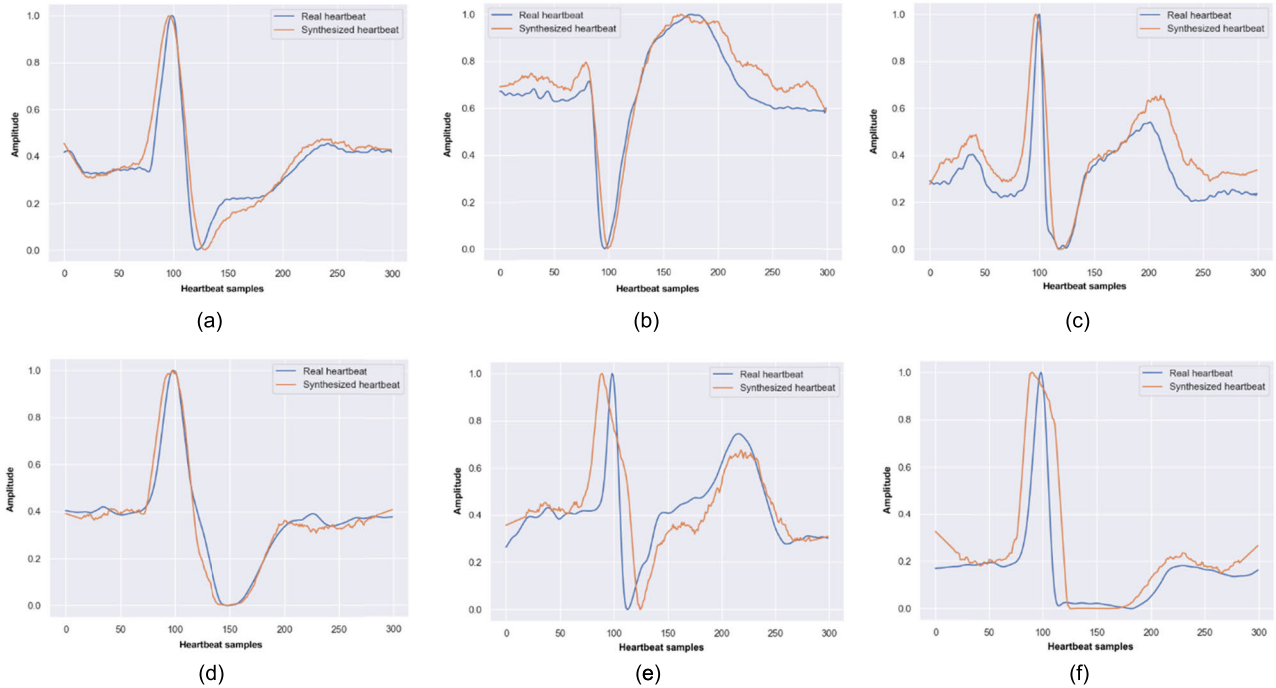


FIGURE 4. Samples of real and synthesized heartbeats using GANs: (a) Left Bundle Branch Block (LBBB) Class, (b) Premature Ventricular Contraction (PVC) Class, (c) Right Bundle Branch Block (RBBB) Class, (d) Ventricular Escape (VE) Class, (e) Aberrated Atrial Premature (AP) Class, (f) Atrial Escape (AE) Class.

networks according to (1) and (2) using Adam optimizer [52]:

$$D_{loss} = -\frac{1}{m} \sum_{i=1}^m \log(D(x^i)) + \log(1 - D(G(z^i))) \quad (1)$$

$$G_{loss} = \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^i))) \quad (2)$$

where m is the number of samples per minibatch, D is the discriminator network, G is the generator network, x is the real samples, and z is a noise vector. Figure 3 shows the proposed architecture of the generator and discriminator networks.

Two post-processing steps are applied after generating the heartbeats to improve the results. In the first step, Savitzky–Golay filter [53] is used to smooth the amplitude of the heartbeats for enhancing their quality. After that, the amplitude is normalized between 0 and 1 as the amplitude of the real heartbeats. In the second step, random sample consensus (RANSAC) is used for identifying and removing the outliers of the generated samples and ensuring that they come from

the same distribution. Figure 4 shows samples for real and synthetic heartbeats of different classes from the MIT-BIH dataset.

Although there are slight differences between the synthesized and original heartbeats in Figure 4, this is the intention and goal of the GAN. In this study, the aim is to generate heartbeats that have the main features of the original ones, not to generate identical versions of them. The same idea is adopted when using any data augmentation technique: The original samples are modified slightly to obtain diversity in the training set.

C. CLASSIFICATION STAGE

We propose two approaches based on deep CNNs to classify 15 arrhythmias from the MIT-BIH dataset that are distinct from other recent classification approaches; no significant feature extraction of ECG data is needed to achieve strong

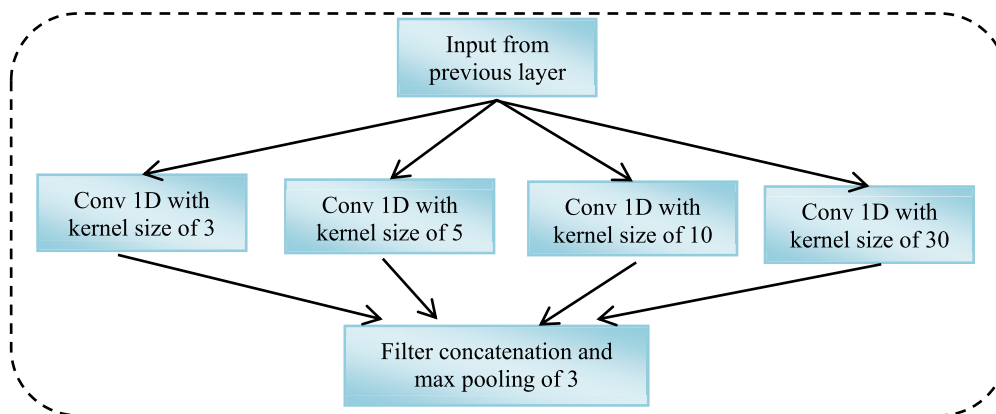


FIGURE 5. Components of the proposed inception module.

TABLE 1. Mapping the MIT-BIH classes to the five main categories.

Category	MIT-BIH Classes
N	NOR, LBBS, RBBB, AE, NE
S	APC, AP, BAP, NP
V	PVC, VE, VF
F	VFN
Q	FPN, UN

classification performance. The first approach is an end-to-end architecture that classifies the heartbeats in a direct way. The second approach is a two-stage hierarchical process that determines the category of the heartbeats in the first stage and classifies the class belonging to that category in the second stage. The details of each approach are discussed in the next subsections.

1) END-TO-END APPROACH

In this approach, the model takes the heartbeat as an input and classifies it as one of the 15 classes in an end-to-end way. The motivation behind choosing the proposed architecture, based on our analysis, is inspired from the inception network [54] for the following reason; In this study, the positions of the waves (P, QRS, and T wave) are not fixed and the length of the waves is not same for all heartbeats. So, the proposed approach calculates the features using multiple kernel sizes to guarantee that the considered wave features are invariant to the length of each wave. The proposed architecture consists of three inception modules followed by three fully connected layers, each inception module consists of multiple convolutional layers that operate on the same level; each layer has a number of filters with a specific kernel size, and padding is applied to unify the output sizes to be able to concatenate them. The outputs are concatenated and the size of the concatenated filters is reduced by applying a max-pooling operation. Figure 5 shows the components of the proposed inception module, and Figure 6 describes the proposed architecture of the first approach.

2) TWO-STAGE HIERARCHICAL APPROACH

According to ANSI/AAMI EC57: 1998 standard, the 15 classes of the MIT-BIH dataset are mapped into five main

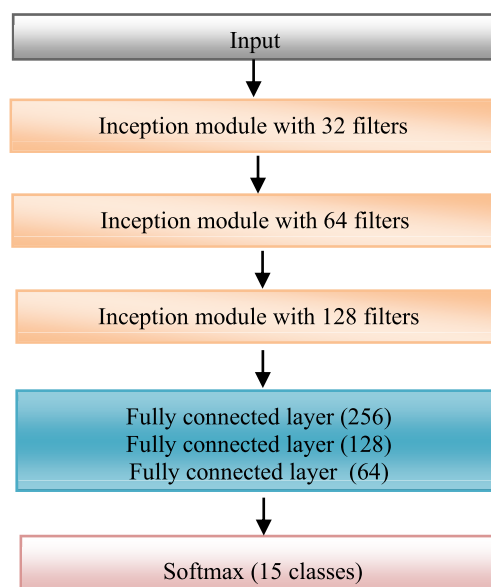


FIGURE 6. Proposed architecture of the end-to-end approach.

categories as shown in Table 1. The classification is done based on two stages in this approach. In the first stage, the heartbeats are classified into one of the five main categories. Subsequently, each heartbeat in the second stage is classified into one of the classes that belongs to that category. The proposed approach is shown in Figure 7.

The architecture of each CNN in Figure 7 is similar to the architecture in Figure 6 except for the number of fully connected layers and the number of neurons in each layer. Category F has only one class, so no classification network is needed for it in stage 2. Only the correctly classified heartbeats in stage 1 will be passed to the second stage.

The data augmentation process is slightly different in this approach. It is simple in the end-to-end approach because each class will contain 9,660 samples as the number of heartbeats in the Normal Class, but in this approach, the data augmentation is done across two stages. In the first stage, GAN is applied to the classes of categories S, V, F, and Q to match the number of heartbeats in category N; after augmentation, each

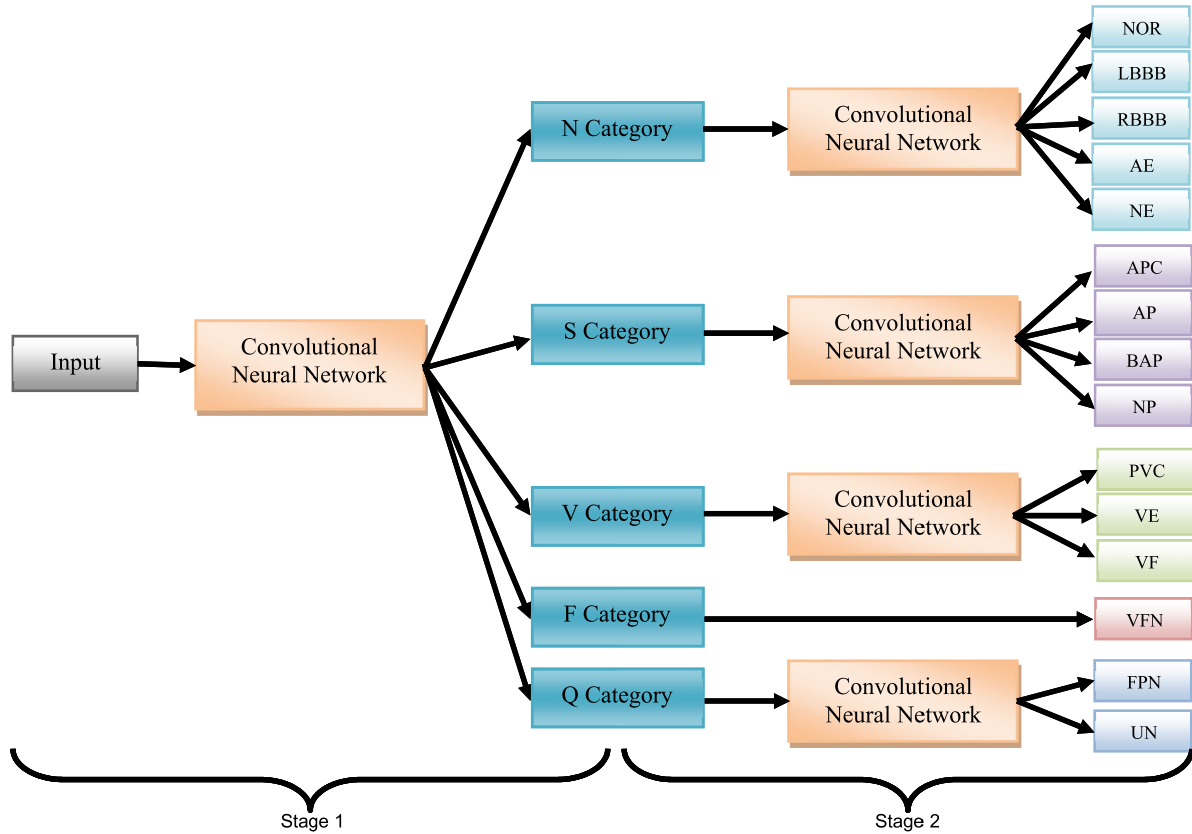


FIGURE 7. Proposed architecture of the two-stage hierarchical approach.

category will have a training set containing 15,904 heartbeats. In the second stage, each category is balanced using GAN separately based on the major class in each category. For instance, the major class in the first category is the Normal Class, with 9,660 heartbeats. Based on this, LBBB, RBBB, NE, and AE are balanced to have a training set containing 9,660 heartbeats for each class.

IV. EXPERIMENTAL RESULTS

A. DATASET

The MIT-BIH dataset [55] is the most popular dataset for arrhythmias, and it is used for arrhythmia detection in most studies. It contains 48 records of individuals of different genders and ages; each record is a 30-minute-long recording of heartbeat signals, with a sampling frequency of 360 Hz. The heartbeats and R-peak locations have been annotated by experts and associated with the dataset; these annotations and locations have been utilized as the ground truth in the training and evaluation phases. Only ECG data from lead 1 has been considered. According to ANSI/AAMI EC57: 1998 standard [56], only 44 records can be utilized because there are four paced records. Hence, 15 arrhythmias are considered in this study.

In this study, the beats of utilized records from lead 1 only were divided into training and testing sets. For comparison's

sake, the data division in [3] and [43] has been followed. The percentages of training and testing sets were not the same for all classes because the numbers of beats for the classes were not equally distributed. The training set consisted of 13% of the total beats from the Normal Class, which contains tens of thousands of beats; 40% of the total beats from the classes with large number of beats; and 50% of the total beats from the classes with a small number of beats. The division of the beats is described in Table 2.

B. EVALUATION METRICS

The following metrics have been utilized to evaluate the performance of the proposed methods: (1) precision, (2) sensitivity (recall), (3) specificity, and (4) overall accuracy. The metrics were calculated based on the normalized confusion matrices using the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. The equations of the metrics are as follows:

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

- Sensitivity = $\frac{TP}{TP+FN}$ (4)

- Specificity = $\frac{TN}{TN+FP}$ (5)

- Overall Acc = $\frac{TP+TN}{TP+FP+TN+FN}$ (6)

TABLE 2. Training and testing percentages utilized in the experiments.

Heartbeat type	Annotation	Number of total beats	Training ratio	Number of training beats
Normal beat (NOR)	N	74309	13%	9660
Left Bundle Branch block (LBBB)	L	8069	40%	3227
Right Bundle Branch block (RBBB)	R	7238	40%	2895
Atrial Premature Contraction (APC)	A	2544	40%	1017
Premature Ventricular Contraction (PVC)	V	6890	40%	2756
Aberrated Atrial Premature (AP)	a	150	50%	75
Ventricular Flutter Wave (VF)	!	472	50%	236
Fusion of Ventricular and Normal (VFN)	F	802	50%	401
Blocked Atrial Premature (BAP)	x	193	50%	96
Nodal (junctional) Escape (NE)	j	229	50%	114
Fusion of Paced and Normal (FPN)	f	982	50%	491
Ventricular Escape (VE)	E	106	50%	53
Nodal (junctional) Premature (NP)	J	83	50%	41
Atrial Escape (AE)	e	16	50%	8
Unclassifiable (UN)	Q	15	50%	7
Total	15	102098	20.6%	21077

C. RESULTS OF THE END-TO-END APPROACH

In this approach, the training set is selected randomly for each class according to the data division in Table 2, and the other beats are used in the testing set. There is no duplication between the training and testing sets. The generated heartbeats by the GAN were used to increase the training set for the classes (except the Normal Class). The final training set after data augmentation had 144,900 beats, with each class having 9,660 training beats. Finally, Adam optimizer [52] was utilized to tune the parameters, and the network weights were initialized with random values from standard normal distribution. The proposed approach was applied only to lead I from the MIT-BIH arrhythmias dataset, and 15 arrhythmia classes were considered.

The proposed model in Figure 6 was trained using the same techniques and hyperparameters as the original imbalanced dataset and the augmented dataset with GAN to observe the effect of balancing the dataset. The generated heartbeats were utilized only in the training phase, and the testing set contained unseen real heartbeats. The confusion matrix of this approach is shown in Table 3.

The precision, sensitivity, and specificity for each class before and after data augmentation are shown in Table 4. Although increasing the number of training samples decreased the precision slightly for some classes, such as APC, AP, and VF, the GAN had a great effect on the minor classes and significantly increased the precision for these classes, such as AE, UN, and NE. It is worth mentioning that the dangerous and rare diseases usually fall in the minor classes of arrhythmia disease. However, UN segments contain distortions in one or more of the three main waves (P, QRS, and T). So, UN segments can't be recognized as a specific heart disease.

Generating synthetic heartbeats using GAN and adding them to the training sets achieved better results. After data augmentation, the precision increased by 8.64%, achieving 90.0%, and the specificity increased by 0.76%, achieving 99.23%; the overall accuracy increased by 0.5%, achieving 98.3%. The effect of the GAN does not appear clearly on the overall accuracy because it is biased to the major classes, which already contain large numbers of samples and can be recognized easily; the effect has been shown on the precision of minor classes, which need to be balanced because they contain only tens of samples.

D. RESULTS OF THE BASELINE DATA AUGMENTATION TECHNIQUES IN THE END-TO-END APPROACH

To illustrate the efficiency of the GAN, we compared our results after data augmentation in the end-to-end approach against other common data augmentation techniques, such as random oversampling, the synthetic minority oversampling technique (SMOTE), and adaptive synthetic (ADASYN) sampling. Random oversampling randomly replicates the samples of the minor classes to match the number of samples in the dominant class. However, it increases the likelihood of overfitting. SMOTE [57] generates synthetic data based on the similarities of the feature space that exist in the samples of the minor classes. It randomly selects one of the neighbors of each sample in the minor classes and generates new samples by calculating linear interpolations for the samples. In contrast, He *et al.* [58] proposed another technique called ADASYN to generate synthetic samples based on the density distributions of the training data. Moreover, we trained and evaluated the proposed model with the original unbalanced dataset using the weighted loss strategy. Table 5 shows the results of the end-to-end approach using ten folds for each

TABLE 3. Confusion matrix of the end-to-end approach.

		Predicted Labels															Total
		N	L	R	A	V	a	!	F	X	j	f	E	J	e	Q	
Ground Truth Labels	N	63703	50	55	204	196	41	30	54	58	65	52	8	16	76	41	64649
	L	22	4781	0	11	24	0	2	1	1	0	0	0	0	0	0	4842
	R	27	0	4300	2	5	0	1	0	0	0	0	1	5	2	0	4343
	A	134	1	16	1341	10	3	1	0	2	3	2	0	6	8	0	1527
	V	68	4	8	9	4010	6	16	3	5	0	1	1	0	1	2	4134
	a	14	0	0	4	5	51	0	0	0	0	1	0	0	0	0	75
	!	9	1	0	3	2	2	215	0	0	0	0	0	0	0	4	236
	F	6	0	0	0	17	0	1	371	1	1	2	0	0	0	2	401
	x	5	0	5	0	0	0	0	0	86	1	0	0	0	0	0	97
	j	9	0	0	5	1	0	0	0	0	99	0	0	1	0	0	115
	f	14	0	0	1	1	0	0	0	0	3	468	0	0	2	2	491
	E	1	0	0	0	2	0	0	0	0	0	0	50	0	0	0	53
	J	1	0	0	0	1	0	0	0	0	1	0	0	38	1	0	42
	e	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	8
	Q	1	0	0	0	1	0	0	0	0	1	0	0	0	0	5	8
Total	64014	4837	4384	1580	4275	103	266	429	153	174	526	60	66	98	56	81021	

TABLE 4. Coefficients of classifying 15 ECG classes before and after data augmentation for the end-to-end approach. DS1 is the original set and DS2 is the balanced set.

Heartbeat type	Precision	Sensitivity	Specificity	Precision	Sensitivity	Specificity
	(DS1)	(DS1)	(DS1)	(DS2)	(DS2)	(DS2)
NOR	97.84%	98.73%	99.80%	98.54%	98.07%	99.89%
LBBB	99.38%	99.95%	99.94%	98.74%	99.93%	99.90%
RBBB	99.59%	99.86%	99.96%	99.00%	99.89%	99.92%
APC	92.80%	98.97%	99.36%	87.82%	99.70%	99.04%
PVC	95.77%	99.65%	99.62%	97.00%	99.65%	99.76%
AP	70.00%	99.76%	97.40%	68.00%	99.34%	97.56%
VF	94.07%	99.91%	99.47%	91.10%	99.36%	99.30%
VFN	95.76%	99.97%	99.62%	93.27%	99.27%	99.41%
BAP	80.41%	99.87%	98.31%	88.66%	99.92%	99.11%
NE	68.74%	99.96%	97.35%	86.09%	99.90%	98.91%
FPN	95.54%	99.94%	99.62%	95.32%	99.93%	99.63%
VE	96.22%	99.98%	99.66%	94.34%	99.87%	99.55%
NP	85.72%	98.88%	98.75%	90.48%	99.96%	99.25%
AE	37.50%	100.0%	94.98%	100.0%	99.89%	100.0%
UN	12.50%	99.98%	93.25%	62.50%	99.94%	97.17%
Average	81.46%	99.69%	98.47%	90.00%	99.77%	99.23%

technique compared with the results of the GAN. It demonstrates the efficiency of the GAN compared to the other common data augmentation techniques.

E. RESULTS OF THE TWO-STAGE HIERARCHICAL APPROACH

In this approach, the heartbeats are classified into one of the five main categories in the first stage, whereas in the second stage, the heartbeats that were correctly classified in the first

stage are classified into the classes belonging to that category; the overall accuracy is measured based on the misclassifications in the two stages.

The proposed model in Figure 7 was trained with the same techniques and hyperparameters using the original imbalanced dataset and the augmented dataset using GAN to see the effect of balancing the dataset in this approach. The confusion matrix for the first stage is shown in Table 6, whereas the results of the first stage before and after data augmentation are shown in Table 7. The effect of the GAN is

TABLE 5. Comparison of the classification results in the end-to-end approach using different techniques.

Technique	Precision	Sensitivity	Specificity	Overall accuracy
Weighted loss	86.12%	99.54%	98.80%	98.28%
Random oversampling	86.20%	99.76%	98.91%	97.79%
SMOTE	85.13%	99.78%	98.82%	97.98%
ADASYN	85.61%	99.73%	98.86%	97.42%
GAN	90.00%	99.77%	99.23%	98.30%

TABLE 6. Confusion matrix for the five main categories in the two-stage hierarchical approach (stage 1).

Ground Truth Labels	Predicted Labels					
	N	S	V	F	Q	Total
N	73085	517	165	82	108	73957
S	135	1587	15	0	4	1741
V	78	24	4179	79	63	4423
F	31	3	10	353	4	401
Q	10	3	4	0	482	499
Total	73339	2134	4373	514	661	81021

TABLE 7. Coefficients of classifying 15 ECG classes before and after data augmentation for the two-stage hierarchical approach (stage 1). DS1 is the original set and DS2 is the balanced set.

Category	Precision (DS1)	Sensitivity (DS1)	Specificity (DS1)	Precision (DS2)	Sensitivity (DS2)	Specificity (DS2)
N	98.89%	94.46%	99.68%	98.82%	96.29%	99.68%
S	87.07%	99.34%	96.53%	91.15%	99.30%	97.71%
V	95.99%	99.53%	98.83%	94.48%	99.74%	98.55%
F	76.56%	99.90%	93.94%	88.03%	99.80%	96.95%
Q	88.78%	99.96%	96.96%	96.59%	99.77%	99.09%
Average	89.54%	98.64%	97.19%	93.81%	98.98%	98.40%

clear from the precision of minor categories, such as F and Q categories because these categories had few samples before data augmentation. We also observed that the GAN has no effect on N and V categories because they are the dominant categories.

The confusion matrices for the categories—except category F because it contains only one class—in the second stage are shown in Figure 8, whereas the results of the second stage before and after data augmentation using the same techniques and hyperparameters are shown in Table 8. The achieved overall accuracy after data augmentation across the two stages is 98.0%, while the average precision, sensitivity, and specificity for the classes in each category are 93.95%, 97.71%, and 97.41%, respectively. In this approach, the GAN increased the precision significantly by 8.65% and increased the overall accuracy by 1.45%. The precision is slightly higher in this approach than it is in the end-to-end

approach. Moreover, the main category is also known, not only the disease class. In contrast, the sensitivity and specificity in this approach are reduced by nearly 2% compared with the end-to-end approach. It is also observed that the GAN has more effect in stage 2 than stage 1 because stage 2 contains the minor classes, whereas stage 1 contains the categories. For instance, the precision of AE class increased in stage 2 after data augmentation from 20% to 87.5%.

As mentioned in section 2, not all studies consider the precision or provide a confusion matrix for the considered classes. The comparison between our work and other studies that consider precision in addition to the overall accuracy is given in Table 9. Most studies consider only a few classes with utterly different beats, resulting in high overall accuracy and precision. However, the average precision decreases when more classes are considered. In contrast, in this study, the proposed approaches using only data from lead I achieved

TABLE 8. Coefficients of classifying 15 ECG classes before and after data augmentation for the two-stage hierarchical approach (stage 2). DS1 is the original set, and DS2 is the balanced set.

Heartbeat type	Precision (DS1)	Sensitivity (DS1)	Specificity (DS1)	Precision (DS2)	Sensitivity (DS2)	Specificity (DS2)
NOR	98.94%	99.64%	99.89%	99.73%	99.61%	99.93%
LBBB	99.58%	99.90%	99.86%	99.70%	99.89%	99.92%
RBBB	99.85%	99.87%	99.95%	99.63%	99.97%	99.90%
AE	20.00%	99.99%	82.81%	87.50%	99.98%	96.90%
NE	86.36%	99.91%	95.90%	91.74%	99.91%	97.91%
APC	99.04%	90.32%	99.63%	98.86%	91.71%	99.58%
AP	87.23%	99.66%	95.49%	84.75%	99.87%	94.95%
BAP	95.24%	99.37%	98.22%	92.94%	99.53%	97.53%
NP	76.00%	100.0%	92.14%	94.87%	99.41%	98.18%
PVC	99.61%	92.86%	99.80%	99.62%	93.45%	99.80%
VE	94.11%	99.98%	97.02%	96.15%	99.95%	98.04%
VF	92.17%	99.63%	96.12%	92.83%	99.67%	96.46%
VFN	82.04%	99.81%	95.49%	88.03%	99.79%	96.95%
FPN	99.38%	50.00%	99.59%	99.58%	83.33%	99.50%
UN	50.00%	99.79%	66.60%	83.33%	99.58%	85.66%
Average	85.30%	95.38%	94.57%	93.95%	97.71%	97.41%

TABLE 9. Comparison of our work with other results.

Study	Number of classes	Feature set	Classifier	Average precision	Overall accuracy
C. Ye et al. [3]	16	DWT + ICA	SVM	88.00%	99.00%
Martis [32]	5	DCT+PCA	PNN	98.69%	99.52%
Sahoo et al. [37]	4	DWT	SVM	96.86%	98.39%
Yazdaniyan et al. [39]	5	Wavelet	SVM	98.23%	96.67%
Elhaj et al. [40]	5	PCA+DWT+HOS	SVM	99.56%	98.91%
El-saadawy et al. [43]	15	DWT	SVM	93.19%	94.94%
Yang et al. [44]	5	PCA	SVM	86.34%	97.77%
Zhang et al. [27]	5	End-to-end	1D-CNN	97.76%	97.50%
Yildirim et al. [47]	5	End-to-end	LSTM	99.00%	99.25%
Acharya et al. [29]	5	End-to-end	1D-CNN	93.47%	94.03%
Proposed methods	15	End-to-end	Deep-CNN	90.00%	98.30%
	15	End-to-end	Two-stage deep-CNN	93.95%	98.00%

better results than other studies did, proving the robustness of the proposed approach. It is worth mentioning that other studies, such as [3] and [43], achieved their average precision by using data from leads 1 and 2, as well as developing a fusion step to make an accurate final decision, which dramatically increased the computational time.

GAN can generate heartbeats that are similar to real ones and significantly improve the results compared with the original data and the other data augmentation techniques. The advantages of the GAN can be summarized as follows:

- 1) It is an unsupervised method; GAN does not require the data to be labeled and can be trained using unlabeled data. However, as mentioned in this study, we trained the GAN on the classes independently to be able to precisely determine the label of each generated sample;
- 2) It can generate highly realistic heartbeats that are indistinguishable from real ones;
- 3) It has the ability to learn the distribution of the data, even if it is complicated; and
- 4) It can even be trained using a small number of samples.

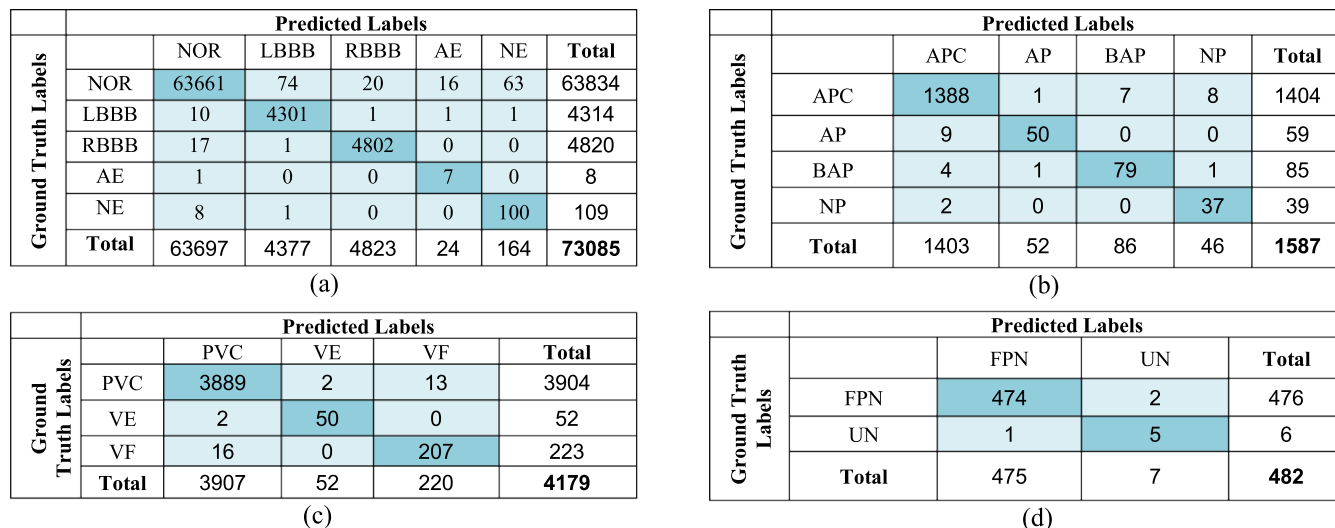


FIGURE 8. Confusion matrices for the classes in two-stage hierarchical approach (stage 2): (a) category N, (b) category S, (c) category V, (d) category Q.

However, GAN also has some limitations, which can be summarized as follows:

- 1) The generated heartbeats are not as smooth as real ones; a post-processing step using a smoothing filter needs to be applied to enhance the quality of the heartbeats; and
- 2) It occasionally generates distortion samples. To increase the precision of the results, outlier removal should be utilized to remove these outliers before using the generated samples.

V. CONCLUSION AND FUTURE WORK

The dynamic heartbeat segmentation technique was utilized because it is invariant for the heart rate variability after filtering the input signal to reduce noise. Thereafter, a novel data augmentation technique was proposed for ECG data using GAN to solve the imbalance problem in the MIT-BIH arrhythmia dataset. Two deep learning approaches were used to classify different heartbeats into 15 classes of the MIT-BIH dataset. The end-to-end approach classifies the heartbeats in a direct way, whereas the two-stage hierarchical approach recognizes the category in the first stage and determines the exact class that falls in that category in the second stage.

Adding synthetic heartbeats has impacted the minor classes and increased their precision significantly. An overall accuracy of 98.30% and precision of 90.0% are achieved by the first approach. The second approach has achieved an overall accuracy of 98.00% and precision of 93.95%, which means that the deep CNNs succeeded in learning the most important features automatically, without any handcrafted features. The results are superior and have been achieved using only the data of lead I, unlike other existing studies, which increase the computation significantly by utilizing data from two leads and adding a fusion step to increase the overall accuracy and average precision if considered. Moreover, we show that balancing the dataset by augmenting the

heartbeats using GAN achieved better results than augmenting using other common techniques.

The resources used in the experiments comprised a 1x Tesla K80 GPU with 2,496 CUDA cores, and 12 GB of GDDR5 VRAM. The training times were 47 minutes in the end-to-end approach and 61 minutes in the two-stage approach, while the testing times for classifying one heartbeat were 0.104 milliseconds in the end-to-end approach and 0.262 milliseconds in the two-stage approach. This means that classifying one 30-minute record of a patient, such as record number 100, will take only 0.235 seconds in the end-to-end approach and 0.590 seconds in the two-stage approach, which proves that both approaches are highly efficient and can be implemented in real-time monitoring systems.

The work in this study can be used in two clinical applications. The first usage is to deploy the models in real-time lightweight wearable devices, as proposed in [59], using an application program interface (API). The models can also be deployed in real-time monitoring using ECG devices in the hospitals. Our future work will develop different variants of the GANs, apply different classification architectures, utilize different sampling rates, and deploy the proposed models in real-time monitoring and classification systems.

REFERENCES

- [1] M. Tantawi, K. Revett, A.-B. Salem, and M. F. Tolba, "Electrocardiogram (ECG): A new burgeoning utility for biometric recognition," in *Bio-inspiring Cyber Security and Cloud Services: Trends and Innovations*, vol. 70, A. Hassanien, T. H. Kim, J. Kacprzyk, and A. Awad, Eds. Springer, 2014, pp. 349–382.
- [2] S. G. Artis, R. G. Mark, and G. B. Moody, "Detection of atrial fibrillation using artificial neural networks," in *Proc. Comput. Cardiol.*, Venice, Italy, Sep. 1991, pp. 23–26.
- [3] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012.
- [4] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 437–448, 2013.

- [5] V. X. Alfonso and J. Tompkins, "Detecting ventricular fibrillation," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 174–177, Jan. 2007.
- [6] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 179–185, 1999.
- [7] J. A. Kastor, *Arrhythmias*, 2nd ed. London, U.K.: W.B. Saunders, 1994.
- [8] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.
- [9] Y. Liu, X. Yu, J. X. Huang, and A. An, "Combining integrated sampling with SVM ensembles for learning from imbalanced datasets," *Inf. Process. Manage.*, vol. 47, no. 4, pp. 617–631, Jul. 2011.
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Jun. 2014, *arXiv:1406.2661*. [Online]. Available: <https://arxiv.org/abs/1406.2661>
- [12] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in *Proc. Int. Conf. Artif. Neural Netw.*, 2017, pp. 626–634, doi: [10.1007/978-3-319-68612-7_71](https://doi.org/10.1007/978-3-319-68612-7_71).
- [13] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," 2017, *arXiv:1709.01872*. [Online]. Available: <https://arxiv.org/abs/1709.01872>
- [14] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "Generalization of deep neural networks for chest pathology classification in X-rays using generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 990–994, doi: [10.1109/ICASSP.2018.8461430](https://doi.org/10.1109/ICASSP.2018.8461430).
- [15] A. Gupta, S. Venkatesh, S. Chopra, and C. Ledig, "Generative image translation for data augmentation of bone lesion pathology," in *Proc. MIDL*, 2019, pp. 1–11.
- [16] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings Bioinf.*, vol. 8, Jul. 2016, Art. no. bbw068.
- [17] B. Zebardast, A. Ghaffari, and M. Masdari, "A new generalized regression artificial neural networks approach for diagnosing heart disease," *Int. J. Innov. Appl. Stud.*, vol. 4, no. 4, p. 679, 2013.
- [18] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technol. Interaction*, vol. 2, no. 3, p. 47, Aug. 2018.
- [19] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.
- [20] H. El-Saadawy, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Electrocardiogram (ECG) classification based on dynamic beats segmentation," in *Proc. 10th Int. Conf. Informat. Syst. (INFOS)*, 2016, pp. 75–80, doi: [10.1145/2908446.2908452](https://doi.org/10.1145/2908446.2908452).
- [21] R. R. Perez, A. Marques, and F. Mohammadi, "The application of supervised learning through feed-forward neural networks for ECG signal classification," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, Vancouver, BC, Canada, May 2016, pp. 15–18.
- [22] M. K. Gautam and V. K. Giri, "A neural network approach and wavelet analysis for ECG classification," in *Proc. IEEE Int. Conf. Eng. Technol. (ICETECH)*, Coimbatore, India, Mar. 2016, pp. 17–18.
- [23] S. Sa'di, R. Hashemi, A. Abdollapour, K. Chalabi, and M. A. Salamat, "A novel probabilistic artificial neural networks approach for diagnosing heart disease," *Int. J. Found. Comput. Sci. Technol.*, vol. 5, no. 6, pp. 47–53, Nov. 2015.
- [24] A. M. Alqudah, A. Albadarneh, I. Abu-Qasmieh, and H. Alquran, "Developing of robust and high accurate ECG beat classification by combining Gaussian mixtures and wavelets features," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 1, pp. 149–157, Mar. 2019.
- [25] H. Alquran, A. M. Alqudah, I. Abu-Qasmieh, A. Al-Badarneh, and S. Almashaqbeh, "ECG classification using higher order spectral estimation and deep learning techniques," *Neural Netw. World*, vol. 29, no. 4, pp. 207–219, Sep. 2019.
- [26] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," *Procedia Comput. Sci.*, vol. 120, pp. 268–275, Jan. 2017, doi: [10.1016/j.procs.2017.11.238](https://doi.org/10.1016/j.procs.2017.11.238).
- [27] D. Li, J. Zhang, Q. Zhang, and X. Wei, "Classification of ECG signals based on 1D convolution neural network," in *Proc. IEEE 19th Int. Conf. E-Health Netw., Appl. Services (Healthcom)*, Oct. 2017, pp. 1–6, doi: [10.1109/healthcom.2017.8210784](https://doi.org/10.1109/healthcom.2017.8210784).
- [28] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017, doi: [10.1016/j.combiomed.2017.08.022](https://doi.org/10.1016/j.combiomed.2017.08.022).
- [29] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, May 2018, doi: [10.1016/j.combiomed.2018.03.016](https://doi.org/10.1016/j.combiomed.2018.03.016).
- [30] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Comput. Biol. Med.*, vol. 102, pp. 278–287, Nov. 2018, doi: [10.1016/j.combiomed.2018.06.002](https://doi.org/10.1016/j.combiomed.2018.06.002).
- [31] T. Tuncer, S. Dogan, P. Pławiak, and U. R. Acharya, "Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals," *Knowl.-Based Syst.*, vol. 186, Dec. 2019, Art. no. 104923.
- [32] R. J. Martis, U. R. Acharya, C. M. Lim, K. M. Mandana, A. K. Ray, and C. Chakraborty, "Application of higher order cumulant features for cardiac health diagnosis using ECG signals," *Int. J. Neural Syst.*, vol. 23, no. 4, Aug. 2013, Art. no. 1350014, doi: [10.1142/s0129065713500147](https://doi.org/10.1142/s0129065713500147).
- [33] R. J. Martis, U. R. Acharya, C. M. Lim, and J. S. Suri, "Characterization of ECG beats from cardiac arrhythmia using discrete cosine transform in PCA framework," *Knowl.-Based Syst.*, vol. 45, pp. 76–82, Jun. 2013, doi: [10.1016/j.knsys.2013.02.007](https://doi.org/10.1016/j.knsys.2013.02.007).
- [34] S. M. Mathews, L. F. Polania, and K. E. Barner, "Leveraging a discriminative dictionary learning algorithm for single-lead ECG classification," in *Proc. 41st Annu. Northeast Biomed. Eng. Conf. (NEBEC)*, Apr. 2015, pp. 1–2, doi: [10.1109/nebec.2015.7117118](https://doi.org/10.1109/nebec.2015.7117118).
- [35] A. Majumdar and R. Ward, "Robust greedy deep dictionary learning for ECG arrhythmia classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AL, USA, May 2017, pp. 4400–4407, doi: [10.1109/IJCNN.2017.7966413](https://doi.org/10.1109/IJCNN.2017.7966413).
- [36] S. Yu and K. Chou, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2841–2846, May 2008.
- [37] S. Sahoo, B. Kanungo, S. Behera, and S. Sabut, "Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities," *Measurement*, vol. 108, pp. 55–66, Oct. 2017.
- [38] R. J. Martis, U. R. Acharya, K. M. Mandana, A. K. Ray, and C. Chakraborty, "Application of principal component analysis to ECG signals for automated diagnosis of cardiac health," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11792–11800, Oct. 2012.
- [39] H. Yazdani, A. Nomani, and M. R. Yazdchi, "Autonomous detection of heartbeats and categorizing them by using support vector machines," in *Proc. 20th Iranian Conf. Biomed. Eng. (ICBME)*, Dec. 2013, pp. 223–227.
- [40] F. A. Elhaj, N. Salim, A. R. Harris, T. T. Swee, and T. Ahmed, "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals," *Comput. Methods Programs Biomed.*, vol. 127, pp. 52–63, Apr. 2016.
- [41] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector AC expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 582–589, Apr. 2004.
- [42] A. Khazaei, "Combining SVM and PSO for PVC detection," *Int. J. Adv. Eng. Sci.*, vol. 3, no. 4, pp. 1–5, Oct. 2013.
- [43] H. El-Saadawy, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Hybrid hierarchical method for electrocardiogram heartbeat classification," *IET Signal Process.*, vol. 12, no. 4, pp. 506–513, Jun. 2018, doi: [10.1049/iet-spr.2017.0108](https://doi.org/10.1049/iet-spr.2017.0108).
- [44] W. Yang, Y. Si, D. Wang, and B. Guo, "Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine," *Comput. Biol. Med.*, vol. 101, pp. 22–32, Oct. 2018.
- [45] S. Savalia and V. Emamian, "Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks," *Bioengineering*, vol. 5, no. 2, p. 35, May 2018, doi: [10.3390/bioengineering5020035](https://doi.org/10.3390/bioengineering5020035).
- [46] Ö. Yildirim, P. Pławiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Comput. Biol. Med.*, vol. 102, pp. 411–420, Nov. 2018, doi: [10.1016/j.combiomed.2018.09.009](https://doi.org/10.1016/j.combiomed.2018.09.009).

- [47] P. Pławiak and U. R. Acharya, "Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals," *Neural Comput. Appl.*, Jan. 2019, pp. 1–25, doi: [10.1007/s00521-018-03980-2](https://doi.org/10.1007/s00521-018-03980-2).
- [48] O. Yildirim, U. B. Baloglu, R.-S. Tan, E. J. Ciaccio, and U. R. Acharya, "A new approach for arrhythmia classification using deep coded features and LSTM networks," *Comput. Methods Programs Biomed.*, vol. 176, pp. 121–133, Jul. 2019, doi: [10.1016/j.cmpb.2019.05.004](https://doi.org/10.1016/j.cmpb.2019.05.004).
- [49] N. V. Thakor, J. G. Webster, and W. J. Tompkins, "Estimation of QRS complex power spectra for design of a QRS filter," *IEEE Trans. Biomed. Eng.*, vol. BME-31, no. 11, pp. 702–706, Nov. 1984, doi: [10.1109/tbme.1984.325393](https://doi.org/10.1109/tbme.1984.325393).
- [50] A. Srivastava, L. Valkov, C. Russell, M. U. Gutmann, and C. Sutton, "VEEGAN: Reducing mode collapse in GANs using implicit variational learning," 2017, *arXiv:1705.07761*. [Online]. Available: <http://arxiv.org/abs/1705.07761>
- [51] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. ICML*, 2015, pp. 1–11.
- [52] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR*, vol. abs/1412.6980, 2014.
- [53] R. Schafer, "What is a savitzky-golay filter? [Lecture notes]," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011.
- [54] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–12.
- [55] *MIT-BIH Arrhythmias Database*. Accessed: Oct. 12 2019. [Online]. Available: <http://www.physionet.org/physiobank/database/mitdb/>
- [56] *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, Standard ANSI/AAMI EC57:1998/(R)2008), Association for the Advancement of Medical Instrumentation, 1998.
- [57] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jul. 2018.
- [58] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1322–1328.
- [59] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016, doi: [10.1109/tbme.2015.2468589](https://doi.org/10.1109/tbme.2015.2468589).



ABDELRAHMAN M. SHAKER received the B.Sc. degree (Hons.) in computer science from Ain Shams University, in 2016. He is currently a Teaching Assistant with the Department of Scientific Computing, Faculty of Computer and Information Sciences, Ain Shams University. His research interests are deep learning, computer vision, and signal processing.



MANAL TANTAWI received the B.Sc. degree (Hons.) in scientific computing in 2003, and the M.Sc. and Ph.D. degrees in scientific computing from Ain Shams University, in 2008 and 2014, respectively. She is currently an Assistant Professor with the Department of Scientific Computing, Faculty of Computer and Information Sciences, Ain Shams University. Her research interests are biomedical signal processing and machine learning. She has published more than 20 publications in these areas.



HOWIDA A. SHEHEED received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering, computers and systems engineering, from the Faculty of Engineering, Ain Shams University, Cairo, Egypt. She is currently a Professor of scientific computing and the Head of the Scientific Computing Department and the Vice Dean for graduate studies and research from 2012 to 2019 with the Faculty of Computers and Information Sciences, Ain Shams University. She has supervised over 20 M.Sc. and Ph.D. students with the faculties of computers in Egyptian universities. Her research interests are biomedical signal processing, human-computer interaction, and bioinformatics. She has more than 50 publications in these areas. She shared in many computer sciences local and international conferences. She was a Reviewer in some distinguished international journals, such as *Computer Methods and Programs in Biomedicine* (Elsevier), the *International Journal of Medical Imaging and Health Informatics*, and the *International Journal of Intelligent Computing and Information Systems*, Egypt.



MOHAMED F. TOLBA (Senior Member, IEEE) has been a Professor of scientific computing with Ain Shams University, since 1984. He was the Vice President from Ain Shams University, from 2002 to 2006, and the Dean of the Faculty of Computers and Information Sciences, from 1996 to 2002. He has more than 220 publications in the fields of AI, image processing, pattern recognition, OCR, scientific computing, and simulation and modeling. He has supervised more than 90 M.Sc. and 50 Ph.D. degrees in Ain Shams University and other Egyptian Universities. He is currently a Consultant to different local and international organizations for IT. He was a member of the International Association for Science and Technology for Development (IASTED), Canada, from 1995 to 2007, and a member of the International Society for Computers and their Applications (ISCA), USA, from 1998 to 2007. He was a member of the Advisory Committee of Strengthening Science and Technology Researchers Project—STRP Ministry of Scientific Research, from 2006 to 2009, and a member of the committee for evaluation of Egyptian space program of the National Authority for Remote Sensing and Space Sciences—Ministry of Scientific Research. He has also been a member of the Association for Computing Machinery (ACM), USA, since 2000, and a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE), USA, since 2000. He has also been a member of the Software Engineering Competence Center (SECC), since 2004, a member of Information Technology Academic Collaboration (ITAC), since 2005, and a member of the eLearning Committee Board, since 2008. He is also the Honorary Chairman of many International conferences and the Chairman for several IT sector committees in Egypt.

...