Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 1158-1168 YEAR: 2024
DOI: 10.17798/bitlisfen.1523524

Classification of ECG Signals Using GAN, SMOTE, and VAE Data Augmentation Methods: Synthetic vs. Real

Turgut ÖZSEVEN[1*]

[1]Tokat Gaziosmanpaşa University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Tokat, Türkiye
(ORCID: 0000-0002-6325-461X)

Keywords: Synthetic data, Data augmentation, GAN, SMOTE, VAE, Heartbeat classification, Deep learning.

Abstract
Classification is separating data into predefined categories by obtaining descriptive features. In the classification process, machine and deep learning algorithms assume that the class samples are evenly distributed. In particular, the dataset size used in deep learning is significant for classification success. However, obtaining balanced data distribution in real-life problems is very difficult. This negatively affects class-based accuracy. Various methods are used in the literature to overcome the unbalanced data problem. This study investigated the effects of GAN, SMOTE, and VAE methods on ECG data. For this purpose, the heartbeat signals in the MIT-BIH dataset were used. To test the performance of the methods, a performance comparison was made using real and synthetic data, and finally, the model trained with synthetic data was tested with real data. According to the results, 96.5% accuracy was obtained with the real data. The highest classification accuracy of 100.0% was obtained in VAE when using only synthetic data. In training with synthetic data and test results with real data, the highest classification success was 86.4% with SMOTE. When synthetic and real data sets are used together, the highest success rate is 98.6% with VAE. In addition, the accuracy of all classes is evenly distributed after data augmentation.

1. Introduction

An electrocardiogram (ECG) is the recording and graphing of electrical activities in the heart to study the heart's working. It is a physiological signal that contains important pathological information for detecting heart diseases. Therefore, the amount of ECG data is often very large. Manually reviewing this data by experts is both time-consuming and subjective. Computer-aided diagnosis methods are used to overcome this problem [1]. Traditional computer-aided diagnosis methods consist of two stages: feature extraction (discrete wavelet transform, time-domain features, frequency-domain features, etc.) and classification [2]. Recently, feature extraction has not been used as a separate process with deep learning models. Deep learning models perform the feature extraction process within the model. Convolutional neural network (CNN), one of the deep learning models, has shown high performance in many studies, such as biosignal classification, image recognition, and arrhythmia classification [2].

Traditional and deep learning-based computer-aided diagnosis systems assume a balanced distribution of class samples. In particular, the dataset size used in deep learning is extremely important for classification success. However, obtaining balanced data distribution in real-life problems is very difficult. This negatively affects class-based accuracy. Unbalanced data increases classification accuracy for classes with more samples [3]. Because some arrhythmias in ECG data are rare, unbalanced data is much more prominent [1].

Undersampling and oversampling methods are used to overcome the unbalanced data problem. In

[*]Corresponding author: turgut.ozseven@gop.edu.tr
Received: 28.07.2024, Accepted: 25.09.2024

the undersampling method, data is discarded from the class with more examples, and the number of examples is equalized to the class with a smaller number of examples. Oversampling, on the other hand, is to balance the number of samples for the class with a large number of samples by reusing the samples in the class with a low number of samples [4]. The most basic method used for oversampling in image-based studies is to change the dimensional and morphological structure of the image. However, both methods corrupt the original structure of the data. Undersampling leads to data waste. In oversampling, it causes an overfitting problem because the same data is reused. To overcome these problems, deep learning-based synthetic data production methods are mostly used. The most widely used method is the Generative Adversarial Network (GAN). Also, a few studies used the Synthetic Minority Oversampling Technique (SMOTE) for oversampling [5] - [9]. Another deep learning-based method, Variational Autoencoders (VAE), has been used in limited studies [10].

GAN contains two neural networks, a generator and a discriminator. The generator generates the candidate image, and the discriminator evaluates whether the image produced is real or fake. Both networks are competitively trained in parallel. The discriminator compares the real image with the image produced by the generator. The comparison result is sent to the generator, which will update itself according to the result. This cycle will continue until the generator produces images that are more like the image held by the discriminator. This process can be followed with several iterations, and the loss curve [3]. Studies on synthetic data generation and data balancing in recent years mostly use GAN [11] - [15]. Studies show that the data produced with GAN can be used for training purposes [11-13]. Tran et al. (2024) produced synthetic data with GAN by converting ECG signals into a two-dimensional format. High success was achieved with the proposed method, and a new benchmark was set for ECG synthesis [14]. In the study using Bi-LSTM and CBAM based GAN for data augmentation, 99.46% classification success was achieved on the MIT-BIH-AR dataset [15]. Hyland et al. (2017) used recurrent conditional GANs (RCGAN) for the synthetic generation of time series data and concluded that they could be used with a small drop in performance when tested with real data [16]. In the study to overcome the data imbalance problem in the MIT-BIH database, the GAN method increased the classification accuracy by about 9% [3]. Synthetic data generation is not only used with the physiological structure of the signal. Wulan et al. (2020) used GAN with signal structure, spectrogram, and wavelet transform [17]. In another study, a SynSigGAN model was proposed for producing ECG, EEG, EMG, and PPG synthetic data [18]. While the GAN can be designed separately for each class in the synthetic ECG data production process, there are also class-based models [19]. Zhou et al. (2021) proposed the ACE-GAN model, which provides class-based data augmentation, and achieved 99.0% success in the MIT-BIH dataset [2]. The GAN method was also used in hybrid form with other methods. Zhu et al. (2019) used BiLSTM for the generator and CNN for the discriminator in the GAN model [20]. Salazar et al. (2021) suggested that GAN and vector Markov Random Field based GANSO model be used in datasets with a very small number of samples [6]. The proposed model was compared with SMOTE, and higher accuracy was obtained in GANSO. In the study in which the data augmentation method called ProEGAN-MS was proposed, the results were compared with the GAN, and it was determined that the proposed model showed higher success [1]. In another study, VAE and GAN were used together, and 98.5% accuracy was obtained in the MIT-BIH dataset [10]. In the study where the transformer and GAN were used together, the TCGAN was proposed, and a success of 94.7% was achieved in the 1D-CNN and MIT-BIH datasets [21]. While creating GAN models, 1D-CNN, 2D-CNN, or BiLSTM models are used. In the study in which GAN-1D-CNN and GAN-BiLSTM models were used, data augmentation was performed with GAN and classification with 1D-CNN and BiLSTM [22]. According to the results of the study, GAN-BILSTM achieved higher success. GAN increased the models' classification success by approximately 11% [22]. In another study, the ResNet CNN model was used with BiLSTM, and the classification success was 99.4% in the MIT-BIH dataset [23]. GAN is used not only for data augmentation/production but also for noise reduction in real data. Global discriminator, local discriminator, and GAN were used in the study to generate noisy ECG signals due to scanning and digitizing ECG recordings [24].

SMOTE was first proposed in 2002, and the interpolation of the original samples with their neighbors was used to generate synthetic samples [25]. For the classification of ECG signals, features were obtained with wavelet packet decomposition and 1D-CNN, and selective ensemble learning framework and SMOTE were used. Classification success was 96.3% in the MIT-BIH dataset [9]. In another study using MIT-BIH and SMOTE, the classification success was 98.3% [8]. In another study called ArrhyNet, classification success was obtained at 92.7% in the MIT-BIH dataset using wavelet transform, SMOTE, and CNN [7]. For ECG classification, the hybrid structure is used not only for data augmentation but also for classification. The results of the study, in which CNN and LSTM were used together for classification and SMOTE and Tomek were used for data augmentation, showed that data augmentation increased the classification success by 20% [5].

As explained above, GAN is the most common method for synthetic data generation in recent years. The SMOTE method is an older method than the GAN method. However, both SMOTE and VAE have been used in limited studies [5-9]. Some studies did not detail the results before data augmentation while giving the classification accuracy [1, 2, 6, 21-23]. This poses a problem when comparing the performance of synthetic data. In addition, data production methods are not often compared in many places. This poses a challenge for researchers to formulate a study plan. In this study, heartbeat data in the MIT-BIH database were obtained, and synthetic data were produced using these data with GAN, SMOTE, and VAE methods. Then, the performance of the real and synthetic data in ECG classification was compared with ten different analyses. Therefore, the contributions of this research article are as follows: 1) The efficiency of synthetic data in ECG classification was analyzed. 2) The classification success performances of three different data generation methods were compared. 3) The performance of the systems trained only with synthetic data in classifying real data was examined. 4) ECG classification success was evaluated when the hybrid data set was used by increasing the data. 5) A guide was created for researchers working in ECG classification on synthetic data generation.

The remainder of this article is organized as follows: Section II contains the dataset and methods used in the study. Section III contains the results of the analyses performed. Section IV contains the conclusion and discussion.

2. Material and Method

2.1. Data Descriptions

The MIT-BIH [26] database, created in cooperation with MIT and Beth Israel Hospital, was used. This database contains 48 ECG recordings of 47 participants, 30 minutes long. Annotation of the database was done by two or more cardiologists. Each recording contains two channels, MLII and V1, V2, V4, and V5, and we used the data of the MLII channel. Since 90% of the MIT-BIH database contains a normal heartbeat, it has a highly unbalanced structure [27]. Forty-two ECG recordings containing beats belonging to four classes (F: ventricular fusion beat, N: normal beat, S: Supraventricular premature beat, V: ventricular premature beat) were used in the analyses. The annotations in the MIT-BIH database and the heartbeat distribution used according to the AAMI standards are given in Table 1. Since heartbeat creates different graphs for the same diagnosis in different patients, the records used in the dataset were chosen as participant-dependent.

Table 1. Heartbeat distribution used

AAMI	MIT-BIH	Beat Count	Record Numbers
F	F	749	205, 208, 213, 219, 223
N	N, L, R, e, j	10440	100, 101, 103, 105, 106, 108, 109, 111, 112, 113, 115, 116, 117, 118, 119, 121, 122, 123, 124, 200, 201, 202, 205, 208, 209, 210, 212, 213, 215, 220, 221, 222, 223, 230, 234
S	A, a, J, S	2612	100, 101, 103, 108, 112, 113, 116, 117, 118, 121, 124, 200, 201, 202, 203, 205, 208, 209, 210, 213, 215, 220, 222, 223, 228, 232, 233, 234
V	V, E	4428	105, 108, 109, 111, 114, 116, 118, 119, 121, 123, 201, 202, 207, 208, 209, 213, 214, 215, 219, 221, 223, 228, 230, 233, 234

2.2 Convolutional Neural Network (CNN)

CNN is a multi-layered perceptron that carries out the feature extraction and classification process with the help of more than one layer. Therefore, the hardware and data it needs is more than that of traditional methods. CNN obtains features from the data with a certain number of filters and filter sizes in each convolutional layer, giving a linear output. The activation function is used after the convolutional layer to use the model on nonlinear problems. Another layer used in CNN architecture is the dropout layer. This layer prevents overfitting by removing some neurons from the model. The layer used for classification in CNN models is the dense layer [28]. Within the scope of the study, a simple CNN model was created to perform the analysis. The developed model is given in Figure 1.

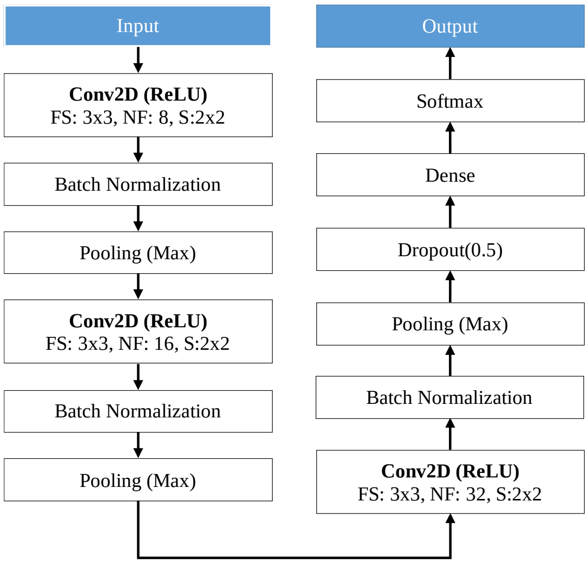


Figure 1. The structure of the 2D-CNN model used in the study (FS: FilterSize, NF: NumFilters, S:Stride)

The parameters used in the training phase of the created CNN model are given in Table 2.

Table 2. Parameters used for the CNN model

Parameters	Value
Data selection for Training, Validation, and Testing	Random permutation
The part of the data reserved for training, validation, and testing	70%, 15%, %15
Optimizer	Adam
Learning Rate	0.001
Epochs	20
Mini Batch Size	64

2.3. Generative Adversarial Network (GAN)

The GAN approach was first proposed by Goodfellow et al. in 2014 [29]. It is used for data generation in many fields, such as image, video, natural language processing, audio analysis, and time series. GAN includes the generator (G) and discriminator (D) model. The generator is used to generate data. It generates synthetic data *G(z)* using random noise *z*. On the other hand, the discriminator tries to determine whether the sample is synthetic by taking samples from real and synthetic data. The GAN training process is defined as a game between two competing networks. The discriminator learns to distinguish between real and fake samples. The Generator learns to deceive the discriminator. The input parameter is *x* representing the data, and the output *D(x)* is the probability that *x* is real data. The formula for GAN is given in Equation 1 [4]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim P_{data}(x)}[logD(x)] + \mathbb{E}_{z\sim P_z(z)}[\log(1 - D(G(z)))] \qquad (1)$$

where *D(G(z))* is the probability that network *D* decides that the data produced by *G* is real. The purpose of *G* is to maximize *D(G(z))*. The stronger *D*'s ability, the bigger *D(x)* should be and the smaller *D(G(x))* should be. Meanwhile, *V(D, G)* will get bigger. P_{data} represents the distribution of actual data. P_z represents the distribution of noise, usually a Gaussian distribution, from which we can form a synthetic image. $\mathbb{E}_x$ and $\mathbb{E}_z$ represent the expected log probability from different outputs of both real and generated images.

The model used for synthetic data generation with GAN is given in Figure 2.

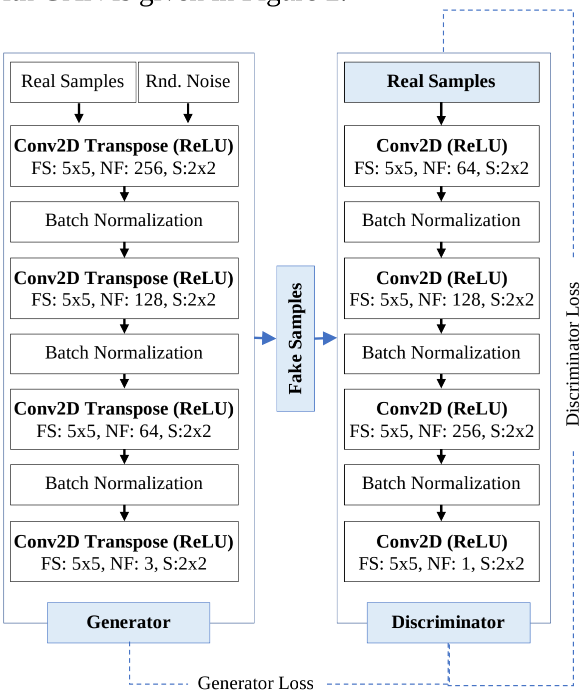


Figure 2. The structure of the GAN model used (epochs: 500, minibatchsize: 128, learningrate: 0.0002)

Samples of synthetic images obtained with the GAN model are given in Figure 3.

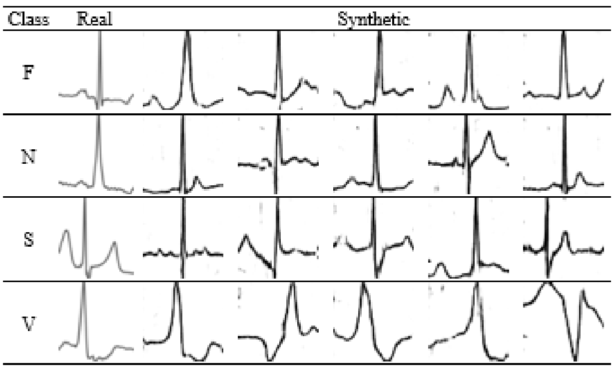


Figure 3. Synthetic images obtained with GAN.

2.4. Synthetic minority Over-sampling Technique (SMOTE)

The SMOTE method, first introduced in 2002, is used for oversampling for classes with a small number of samples and undersampling for classes with a large number of samples [25]. Unlike random sampling methods, it produces synthetic samples based on the k nearest neighbors of the samples examined. The details of the algorithm are given below [30]:

Suppose we have a dataset $D \in R^{nxm}$ with m number of features and n number of samples and K categories. Calculate the Euclidean distance between each sample in the minority class $T(T \in K)$ and other classes. A sampling rate is then adjusted based on the sample imbalance rate to determine the sample multiplier N. For each minority sample $X_i(i = 1, \ldots, num(T))$, several samples are randomly selected among k neighbors $X_i^j(j = 1, \ldots, k)$. For each randomly selected nearest neighbor X_i^j, the new sample is synthesized with sample X_i according to the formula:

$$X_{synthetic} = X_i + \left(X_i - X_i^j\right).rand() \tag{2}$$

where *rand()* is a uniformly distributed non-random number. Samples of synthetic images obtained with SMOTE are given in Figure 4.

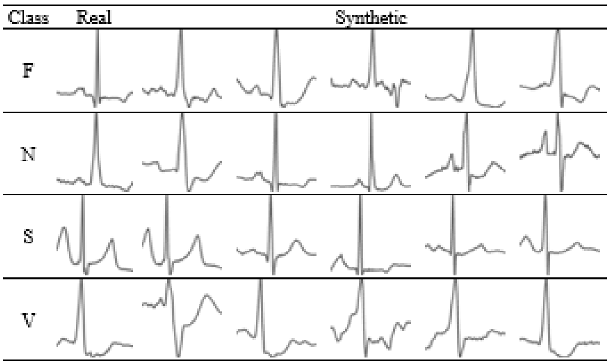


Figure 4. Synthetic images obtained with SMOTE.

2.5. Variational Autoencoder (VAE)

The main purpose of the autoencoder is to compress data with very low losses. It consists of two parts: an encoder and a decoder. The encoder compresses the data, and the decoder decompresses the compressed data. Compression and low-level data representations with the encoder are expressed as latent vectors. In standard autoencoders, each input (x) is encoded by passing through the encoder and converted into a latent vector. This vector is called code. Then, the code is passed through the decoder and restored to its original state (x'). In VAE, conversely, by adding the variation term to the autoencoder, the encoder output is obtained as mean (z_μ) and variance (z_σ) values. The code (z) is obtained by sampling from these values. The decoder part is the same in the autoencoder and VAE [10], [31].

$$z_\mu, z_\sigma = E(x; Q_e), z = z_\mu + \epsilon z_\sigma = q(z|x) \tag{3}$$

As given in Equation 3, the z code is derived from the real values of the data, $q(z|x)$. The reconstructed data x' is obtained from the distribution expressed by $p(x|z)$ (Equation 4).

$$x' = G(z; Q_d) = p(x|z) \tag{4}$$

For synthetic data generation with VAE over ECG signals, a 2-layer CNN architecture is used in the encoder and decoder. For the encoder, the filter size in the layers is 3x3, and the number of filters is 8 and 16. For the decoder, the filter size is 3x3 and the filter numbers are 16 and 8. Learning rate 0.0001, epoch 100, and mini batch size 32 were used in the model's training process. Samples of synthetic images obtained with VAE are given in Figure 5.

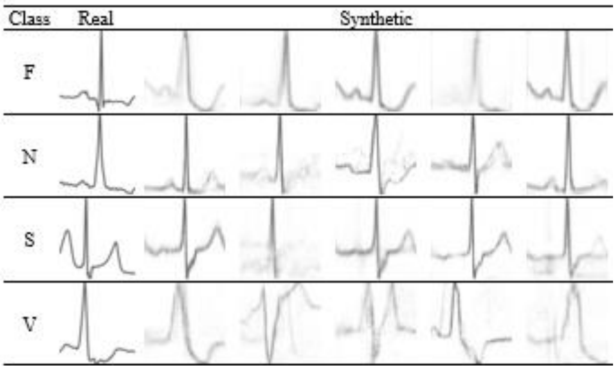


Figure 5. Synthetic images obtained with VAE.

3. Experimental Results

Our study performed ten different analyses for ECG classification with real and synthetic data. The first of these analyses was made with real data, and the other was performed with synthetic and real data. Three analyses were performed for each of the three data generation methods. While performing the analyses, 70%, 15%, and 15% of the dataset were used for training, validation, and testing, respectively. Specificity, accuracy, recall, and f1-score were used to evaluate its performance.

3.1. Classification Results with Real Data

Heartbeat images obtained from the MIT-BIH dataset were used to analyze real data. The dataset has an extremely unbalanced structure in its current form. No action was taken to solve the unbalanced problem in the data set. The dataset contains 749 samples for class F, 10440 for class N, 2612 for class S, and 4428 for class V. Training and validation curves obtained after training with the real dataset are given in Figure 6.

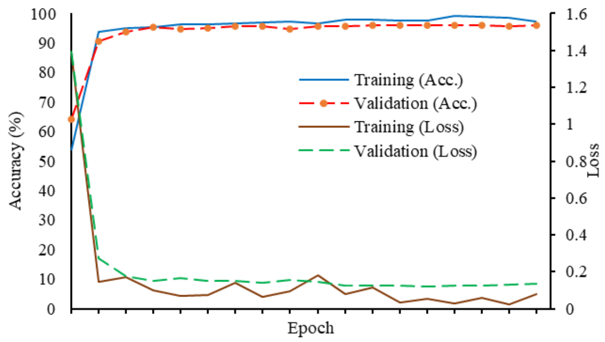


Figure 6. Training and validation curves with the real dataset.

When the graph obtained with the real data set is examined, the model's training has been carried out successfully, even though the data set is unbalanced. In addition, since there is no significant difference between training and validation, the overfitting problem has not been encountered. To analyze the results obtained with the real data in more detail, classification was made with the test dataset, and the details are given in Table 3.

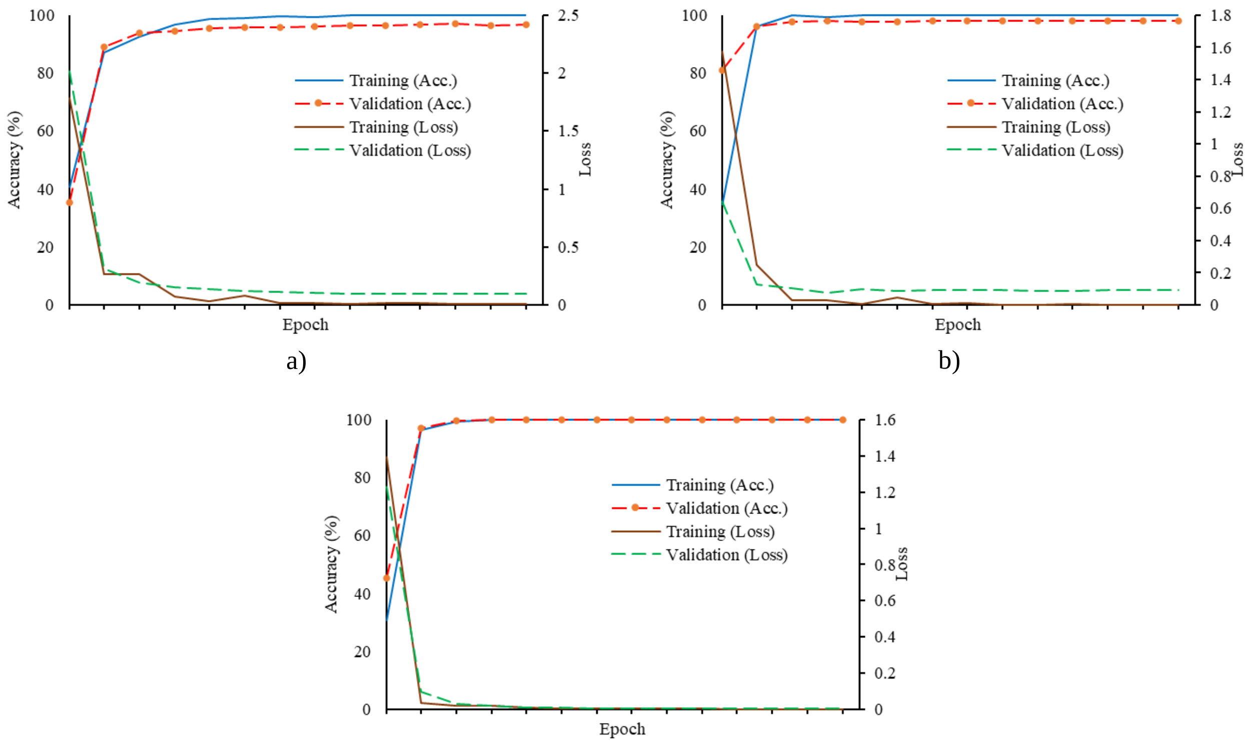
Table 3. ECG classification metrics with real dataset

Class	Acc. (%)	Specificity	Recall	f1-score	Overall
F	74.11	0.9850	0.9088	0.9283	96.53 %
N	98.40				
S	93.11				
V	97.89				

When the results in Table 3 are examined, the overall classification accuracy with the test data set is 96.53%. When class-based classification accuracies were examined, the lowest success was obtained in the F class with the lowest number of samples. In addition, class-based classification accuracies vary in proportion to the number of class samples.

3.2. Classification Results with Synthetic Data

GAN, SMOTE, and VAE methods were used to generate synthetic data. With these methods, 1000 synthetic samples were produced for each class. The training and validation curves obtained after training the CNN model with the synthetic dataset are given in Figure 7.



a) b) c)

Figure 7. Training and validation curves with synthetic data a) GAN, b) SMOTE, c) VAE

There is no significant difference between training and validation values in all three data generation methods. After a certain iteration, it continued horizontally in three graphs. However, the VAE method has higher training and lower loss values than the other methods. To test the validity of these results obtained after the training process, classification was made with the test dataset, and the details are given in Table 4.

Table 4. ECG classification metrics with a synthetic dataset

Method	Accuracy (%)				Specificity	Recall	f1-score	Overall Acc. (%)
	F	N	S	V				
GAN	100.00	95.33	94.67	96.67	0.9889	0.9667	0.9666	96.67
SMOTE	98.00	98.67	96.67	98.00	0.9928	0.9783	0.9784	97.83
VAE	100.00	100.00	100.00	100.00	1	1	1	100.00

When the results given in Table 4 are examined, it is seen that the synthetic data are classified with high accuracy. Also, class-based accuracy rates are evenly distributed. However, the fact that all metrics of the VAE model are 100% raises the question of whether the model has an overfitting problem. Although the training, validation, and test data show that there is no overfitting problem, the use of real data and synthetic data in the next analysis will further clarify the result.

3.3. Results of Training with Synthetic Data and Testing with Real Data

In this analysis, GAN, SMOTE, and VAE models trained in the previous analysis were tested with real ECG data to examine the effectiveness of synthetic data. Random 500 samples from each class of real ECG data were used for the test. The results obtained are given in Table 5.

Table 5. Results of the test were obtained using synthetic data and training real data

Method	Accuracy (%)				Specificity	Recall	f1-score	Overall Acc. (%)
	F	N	S	V				
GAN	85.25	82.02	75.15	96.97	0.9495	0.8485	0.8476	84.85
SMOTE	67.88	89.72	90.30	97.58	0.9546	0.8637	0.8616	86.37
VAE	97.37	94.34	68.69	82.63	0.9525	0.8576	0.8550	85.76

According to the results given in Table 5, the three methods provided similar overall accuracy rates. However, when class-based accuracies were examined, there were differences according to the methods. The highest accuracy in the F and N classes was obtained with the VAE. In S and V classes, the SMOTE method provided higher accuracy than other methods.

3.4. Classification Results with Hybrid Data

This analysis created a dataset of 2000 samples for each class using 1000 randomly selected samples from the real and synthetic datasets. However, since the real dataset in the F class is 749, synthetic data is used more in this class. This data set was randomly divided into 70% training, 15% validation, and 15% testing. Then, the training process was carried out using the CNN model in other analyses. The training and validation curves obtained after training the CNN model are given in Figure 8.

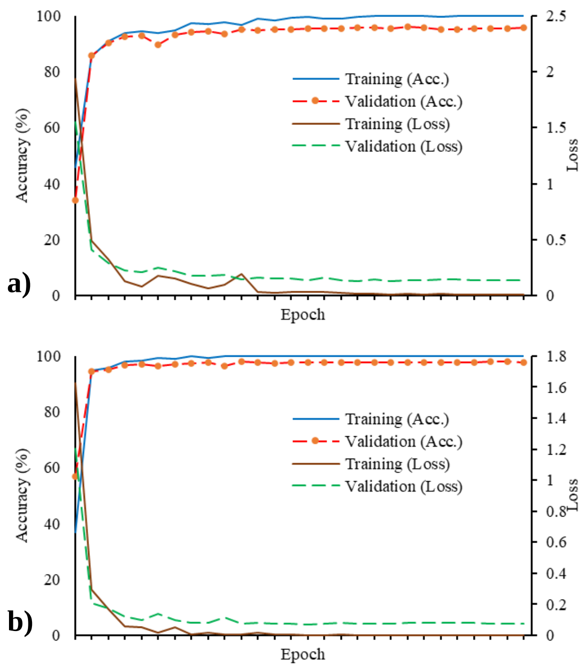


Figure 8. Training and validation curves with synthetic and real data **a)** GAN+Real, **b)** SMOTE+Real

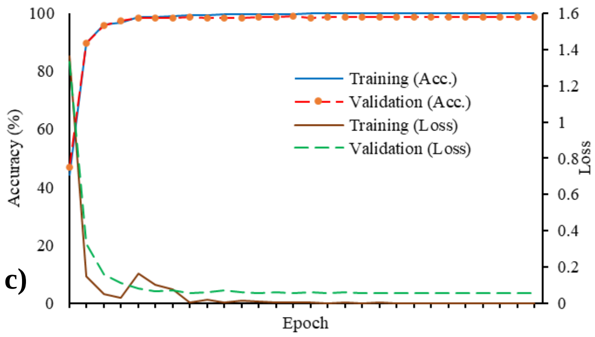


Figure 8. (Continuous) Training and validation curves with synthetic and real data **c)** VAE+Real

There is no significant difference between training and validation values in all three trainings. After a certain iteration, it continued horizontally in three graphs. However, the VAE method has higher training and lower loss values than the other methods. To test the validity of these results obtained after the training process, classification was made with the test dataset, and the details are given in Table 6.

Table 6. ECG classification metrics with synthetic and real datasets

Method	Accuracy (%)				Specificity	Recall	f1-score	Overall Acc. (%)
	F	N	S	V				
GAN+Real	96.00	95.00	96.00	97.33	0.9869	0.9608	0.9609	96.08
SMOTE+Real	98.33	99.67	95.67	97.33	0.9925	0.9775	0.9775	97.75
VAE+Real	98.33	99.67	97.67	98.67	0.9953	0.9858	0.9858	98.58

When the results in Table 6 are examined, the combined use of synthetic and real datasets increased the overall classification accuracy and contributed to the balanced and high class-based accuracies. The highest accuracy was obtained with the VAE method if analyzed on a method basis.

4. Discussion and Conclusion

ECG contains a visual representation of electrical signals related to heart rhythm. The amount of data is very large, as these signals can be one day or longer depending on the duration of the symptoms. Although the amount of data is very large, some arrhythmias are rare so that the data can have an extremely unbalanced structure. A similar situation exists in the MIT-BIH dataset used in this study. When the sample distribution of the 4-classes used is examined, there are 749 examples in the F class and 10440 in the N class.

This study analyzed the effectiveness of GAN, SMOTE, and VAE methods on ECG data to solve the unbalanced data problem and generate a limited number of data synthetically. For this purpose, real datasets, only synthetic datasets, synthetic data, and training and testing with real data and analyses were made using both datasets. The accuracy of these analyses on the test data set is summarized in Table 7.

Table 7. Summary of test accuracies obtained as a result of this study

Data	Accuracy (%)				Overall (%)
	F	N	S	V	
Real	74.11	98.40	93.11	97.89	96.53
GAN	100.00	95.33	94.67	96.67	96.67
SMOTE	98.00	98.67	96.67	98.00	97.83
VAE	100.00	100.00	100.00	100.00	100.00
Training GAN Test Real	85.25	82.02	75.15	96.97	84.85
Training SMOTE Test Real	67.88	89.72	90.30	97.58	86.37
Training VAE Test Real	97.37	94.34	68.69	82.63	85.76
GAN+Real	96.00	95.00	96.00	97.33	96.08
SMOTE+Real	98.33	99.67	95.67	97.33	97.75
VAE+Real	98.33	99.67	97.67	98.67	98.58

With the real dataset, the overall accuracy is 96.53%. However, in class-based accuracies, the accuracy of the F class is low. Also, class-based accuracies are proportional to the number of samples it contains. Class-based accuracies were balanced in the analysis results made with only synthetic data, and VAE achieved 100% classification success. The fact that all classes have 100% success in the VAE method creates the question of whether there may be an overfitting problem. However, the highest classification success was also obtained with VAE in the hybrid dataset. Although training ECG signals with only synthetic data and testing with real data is lower than other methods, it has achieved a classification success of over 85%. The use of data augmentation increased the overall classification success by 2.05%.

The comparison of our results with the results of studies using the MIT-BIH database and data augmentation in the literature is given in Table 8.

Table 8. Comparison of our analysis results with the literature

Model	Data Aug.	Acc. (%)	References
CNN	SMOTE	92.7	[7]
CNN	TCGAN	94.7	[21]
CNN	**GAN**	**96.08**	**This study**
CNN	**SMOTE**	**97.75**	**This study**
CNN	SMOTE	98.0	[3]
CNN	SMOTE	98.3	[8]
CNN	GAN	98.3	[3]
CNN	VAE-ACGAN	98.5	[10]
CNN	**VAE**	**98.6**	**This study**
CNN	GAN	99.0	[2]
CNN + (ensemble) CNN-LSTM	SMOTE	99.0	[5]
ResNet-BiLSTM	GAN	99.4	[23]

The accuracy rates obtained in this study are similar to those of the studies in the literature. The most important difference from the literature studies is that data augmentation was used, and the classification success of synthetic data was also investigated. Synthetic data generation is usable for representing ECG signals if the results are generally evaluated. Models can only be trained with synthetic data and used with real data. In addition, how much the real data used in the production of synthetic data represents, the problem directly affects the results. Although the general success of the data generation methods used is close, there are differences in class-based success. Therefore, using hybrid data augmentation methods will provide higher classification accuracies. Another process that affects the production of synthetic data is the architecture of the methods used. The layers used for GAN and VAE and the parameters of these layers directly affect the result. Neighborhood parameters for SMOTE are similarly effective in achieving the result. As a result, GAN, SMOTE, and VAE methods can be used successfully for synthetic data generation and positively affect classification success.

The main purpose of this study is not to achieve high classification accuracy. For this reason, the CNN model was used at the basic level. In future studies, higher success can be achieved by developing the CNN model, which is used for synthetic data production. In addition, the variability of class-based achievements according to the model indicates that hybrid data generation methods will show higher success.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] H. Yang, J. Liu, L. Zhang, Y. Li, and H. Zhang, ‘Proegan-ms: A progressive growing generative adversarial networks for electrocardiogram generation’, *IEEE Access*, vol. 9, pp. 52089–52100, 2021.

[2] Z. Zhou, X. Zhai, and C. Tin, ‘Fully automatic electrocardiogram classification system based on generative adversarial network with auxiliary classifier’, *Expert Systems with Applications*, vol. 174, p. 114809, 2021.

[3] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, ‘Generalization of convolutional neural networks for ECG classification using generative adversarial networks’, *IEEE Access*, vol. 8, pp. 35592–35605, 2020.

[4] W. Li, Y. M. Tang, K. M. Yu, and S. To, ‘SLC-GAN: An automated myocardial infarction detection model based on generative adversarial networks and convolutional neural networks with single-lead electrocardiogram synthesis’, *Information Sciences*, vol. 589, pp. 738–750, 2022.

[5] H. M. Rai, K. Chatterjee, and S. Dashkevych, ‘The prediction of cardiac abnormality and enhancement in minority class accuracy from imbalanced ECG signals using modified deep neural network models’, *Computers in Biology and Medicine*, vol. 150, p. 106142, 2022.

[6] A. Salazar, L. Vergara, and G. Safont, ‘Generative Adversarial Networks and Markov Random Fields for oversampling very small training sets’, *Expert Systems with Applications*, vol. 163, p. 113819, 2021.

[7] S. S. Aphale, E. John, and T. Banerjee, ‘ArrhyNet: a high accuracy arrhythmia classification convolutional neural network’, in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, 2021, pp. 453–457.

[8] S. K. Pandey and R. R. Janghel, ‘Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE’, *Australasian physical & engineering sciences in medicine*, vol. 42, no. 4, pp. 1129–1139, 2019.

[9] H. Ge, K. Sun, L. Sun, M. Zhao, and C. Wu, ‘A selective ensemble learning framework for ECG-based heartbeat classification with imbalanced data’, in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2018, pp. 2753–2755.

[10] C. Du, P. X. Liu, and M. Zheng, ‘Classification of imbalanced electrocardiosignal data using convolutional neural network’, *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106483, 2022.

[11] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, ‘Synthetic data augmentation using GAN for improved liver lesion classification’, in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, IEEE, 2018, pp. 289–293.

[12] Z. Qin, Z. Liu, P. Zhu, and Y. Xue, ‘A GAN-based image synthesis method for skin lesion classification’, *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105568, 2020.

[13] T. Golany, G. Lavee, S. T. Yarden, and K. Radinsky, ‘Improving ECG classification using generative adversarial networks’, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 13280–13285.

[14] T. D. Tran, T. T. K. Dang, and N. Q. Tran, ‘An Innovative Approach for Long ECG Synthesis with Wasserstein GAN Model’, in *Computational Science and Its Applications – ICCSA 2024*, O. Gervasi, B. Murgante, C. Garau, D. Taniar, A. M. A. C. Rocha, and M. N. Faginas Lago, Eds., Cham: Springer Nature Switzerland, 2024, pp. 339–351. doi: 10.1007/978-3-031-64608-9_22.

[15] F. Zhou and J. Li, ‘ECG data enhancement method using generate adversarial networks based on Bi-LSTM and CBAM’, *Physiol. Meas.*, vol. 45, no. 2, p. 025003, Feb. 2024, doi: 10.1088/1361-6579/ad2218.

[16] C. Esteban, S. L. Hyland, and G. Rätsch, ‘Real-valued (medical) time series generation with recurrent conditional gans’, *arXiv preprint arXiv:1706.02633*, 2017.

[17] N. Wulan, W. Wang, P. Sun, K. Wang, Y. Xia, and H. Zhang, ‘Generating electrocardiogram signals by deep learning’, *Neurocomputing*, vol. 404, pp. 122–136, 2020.

[18] D. Hazra and Y.-C. Byun, ‘SynSigGAN: Generative adversarial networks for synthetic biomedical signal generation’, *Biology*, vol. 9, no. 12, p. 441, 2020.

[19] E. Piacentino, A. Guarner, and C. Angulo, ‘Generating synthetic ecgs using gans for anonymizing healthcare data’, *Electronics*, vol. 10, no. 4, p. 389, 2021.

[20] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, ‘Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network’, *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[21] Y. Xia, Y. Xu, P. Chen, J. Zhang, and Y. Zhang, ‘Generative adversarial network with transformer generator for boosting ECG classification’, *Biomedical Signal Processing and Control*, vol. 80, p. 104276, 2023.

[22] R. R. Sarra, A. M. Dinar, M. A. Mohammed, M. K. A. Ghani, and M. A. Albahar, ‘A Robust Framework for Data Generative and Heart Disease Prediction Based on Efficient Deep Learning Models’, *Diagnostics*, vol. 12, no. 12, p. 2899, 2022.

[23] S. Ma, J. Cui, W. Xiao, and L. Liu, ‘Deep Learning-Based Data Augmentation and Model Fusion for Automatic Arrhythmia Identification and Classification Algorithms’, *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[24] S. Huang, P. Wang, and R. Li, ‘Noise ECG generation method based on generative adversarial network’, *Biomedical Signal Processing and Control*, vol. 81, p. 104444, 2023.

[25] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, ‘SMOTE: synthetic minority over-sampling technique’, *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[26] G. B. Moody and R. G. Mark, ‘The impact of the MIT-BIH Arrhythmia Database’, *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, May 2001, doi: 10.1109/51.932724.

[27] Ö. Yakut, S. Solak, and E. Bolat, ‘IIR Based Digital Filter Design for Denoising the ECG Signal’, *Journal Of Polytechnic*, vol. 21, no. 1, 2018, doi: 10.2339/politeknik.386970.

[28] M. A. Kızrak and B. Bolat, ‘Derin öğrenme ile kalabalık analizi üzerine detaylı bir araştırma’, *Bilişim Teknolojileri Dergisi*, vol. 11, no. 3, pp. 263–286, 2018.

[29] A. Courville and Y. Bengio, ‘Generative adversarial nets’, *Advanc in Neural*, 2014.

[30] Y. Xiong, L. Wang, Q. Wang, S. Liu, and B. Kou, ‘Improved convolutional neural network with feature selection for imbalanced ECG Multi-Factor classification’, *Measurement*, vol. 189, p. 110471, 2022.

[31] C. Güzel Turhan and H. Bilge, ‘Scalable image generation and super resolution using generative adversarial networks’, *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 35, no. 2, 2020.