Artificial Intelligence In Medicine 143 (2023) 102632

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Generative adversarial networks in electrocardiogram synthesis: Recent developments and challenges

Laurenz Berger [a,b,*], Max Haberbusch [a,b], Francesco Moscato [a,b,c]

[a] Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Währinger Gürtel 18-20, A-1090 Vienna, Austria
[b] Ludwig Boltzmann Institute for Cardiovascular Research, Währinger Gürtel 18-20, A-1090 Vienna, Austria
[c] Austrian Cluster for Tissue Regeneration, Donaueschingenstraße 13, A-1200 Vienna, Austria

ARTICLE INFO

Keywords:
Deep learning
Generative adversarial networks
Data augmentation
Artificial intelligence
Electrocardiogram

ABSTRACT

Training deep neural network classifiers for electrocardiograms (ECGs) requires sufficient data. However, imbalanced datasets pose a major problem for the training process and hence data augmentation is commonly performed. Generative adversarial networks (GANs) can create synthetic ECG data to augment such imbalanced datasets. This review aims at identifying the present literature concerning synthetic ECG signal generation using GANs to provide a comprehensive overview of architectures, quality evaluation metrics, and classification performances. Thirty publications from the years 2019 to 2022 were selected from three separate databases. Nine publications used a quality evaluation metric neglecting classification, eleven performed a classification but omitted a quality evaluation metric, and ten publications performed both. Twenty different quality evaluation metrics were observed. Overall, the classification performance of databases augmented with synthetically created ECG signals increased by 7 % to 98 % in accuracy and 6 % to 97 % in sensitivity. In conclusion, synthetic ECG signal generation using GANs represents a promising tool for data augmentation of imbalanced datasets. Consistent quality evaluation of generated signals remains challenging. Hence, future work should focus on the establishment of a gold standard for quality evaluation metrics for GANs.

1. Introduction

The electrocardiogram (ECG) carries substantial information about cardiac activity that can be used for the detection and diagnosis of various cardiac diseases, such as atrial fibrillation or myocardial infarction (MI) [1,2]. Traditionally, the diagnosis of cardiac pathologies based on ECGs has been performed by medical personnel in an office setting and required medical expertise. However, this approach is often time-consuming and costly, which results in limited accessibility to a larger patient population [3]. Fortunately, recent developments in the miniaturization of bioelectrical amplifiers and their integration into wearables, like smartwatches, have created the opportunity to screen larger populations for early signs of cardiac diseases [4,5]. Therefore, in recent years, the automatic detection and classification of cardiac pathologies from ECGs have risen in popularity. Furthermore, the concurrent growth of computing power has enabled the reasonable use of machine learning methods [6,7] and deep neural networks (DNNs) [8,9] for the automatic classification of ECG signals.

In contrast to traditional machine learning techniques, DNNs have the great advantage of being able to automatically detect features from the data [10]. However, their performance is limited by the amount of available data with well-annotated ground truth which is often not trivial to obtain. Moreover, data imbalance has been a major challenge limiting the performance of DNN-based classifiers [11,12]. Unfortunately, publicly available ECG databases are affected by both data availability and imbalance issues. A possible solution is the use of data augmentation which refers to increasing the size and feature variety of a dataset by adding slightly modified versions of the original data while preserving its inherent nature [13]. Data augmentation has shown to be successful in the medical imaging domain [14]. More recently, its application in the context of medical time series classification has seen increasing interest [15,16].

With the introduction of generative adversarial networks (GANs) by Goodfellow et al. [17] in 2014, a powerful DNN-based technique has been established that is capable of generating synthetic data with astonishing authenticity, resembling the inherent features of real-world data with great accuracy. The fundamental principle of GANs is the use of a generator and a discriminator network competing against each

* Corresponding author at: Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Währinger Gürtel 18-20, A-1090 Vienna, Austria.
E-mail address: laurenz.berger@meduniwien.ac.at (L. Berger).

https://doi.org/10.1016/j.artmed.2023.102632
Received 9 March 2023; Received in revised form 25 July 2023; Accepted 8 August 2023
Available online 10 August 2023

other. The generator network is trained to generate data using a random latent vector as input, while the discriminator network is trained to distinguish real data from the synthetic data produced by the generator network. Hence, the generator tries to fool the discriminator into recognizing generated data as real and thus strives to improve the authenticity of the generated signals [18]. Fig. 1 depicts a schematic representation of the entire process of synthetic ECG signal generation using GANs including the authenticity assessment of the generated data and its application in data augmentation for pathology classifier performance enhancement.

Although GANs have been extensively studied for the processing of medical images and partially for medical time series, there is no comprehensive summary of their application to the synthesis of ECG data in the literature yet. Therefore, this review aims at identifying the present literature on GANs for the generation of synthetic ECG signals. It provides a comprehensive overview of employed DNN architectures, metrics to quantify the authenticity of the synthesized signals, and the respective classification performances for GAN-generated data. While there exist multiple generative methods being used to generate ECGs, such as autoregressive models or time-varying autoregressive models, this review aims to investigate the capabilities and potential of GANs for this particular task. Consequently, we made a deliberate decision to narrow our investigation solely to GANs to comprehensively explore their strengths and limitations in ECG generation.

2. Review structure

This review is structured into six sections. First, the literature research and data extraction process are described. Second, the different variations of GAN architectures commonly used for the generation of synthetic ECG signals within the reviewed publications are outlined. Third, a summary of the commonly applied methods and metrics to assess the authenticity of the GAN-generated ECG signals is provided. Fourth, an overview of the standard ECG databases used for GAN-based ECG generation is presented. Fifth, an overview and discussion of the studies that were included in this review are given. Finally, a summary of the key findings and challenges is presented and an outlook on further developments concludes the review.

3. Literature screening

A literature search was conducted using online databases including Scopus, PubMed, and IEEE Xplore. Since the developments in this field are very dynamic, additionally preprint servers such as arxiv.org were included in the search. The online repositories were searched using the query "(ECG OR electrocardiogram) AND (synth* OR creat* OR augment* OR generat* OR simul*) AND (GAN OR generative adversarial)" in title and abstract of records. This review was conducted following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [19].

All original research and conference articles from the time of the establishment of GANs in 2014 to September 2022 were considered. Duplicates were removed before screening. After initial screening, articles that did not use a GAN, were not available in English language, were conducted in a different field of studies, synthesized from non-ECG signals, or were not specific to cardiovascular diseases were removed. Seven articles could not be retrieved because they were not available for download to the authors. This initial selection led to sixty-two reports, which were further assessed for eligibility. Reports were excluded if they were not specific to ECG; did not involve ECG generation but focused on another GAN-related task; did not utilize time-domain data but instead converted them to images; did not conduct quality evaluation or classification; or did not report essential information regarding their methods.

The literature screening process is summarized in Fig. 2. After screening, the following primary information was extracted from all studies: authors, year of publication, GAN architecture, quality evaluation metrics, ECG database, included pathologies, and generated signal length. If available, additional classifier architecture and performance metrics were extracted.

4. GAN architectures for ECG synthesis

GANs were first proposed by Goodfellow et al. [17] in 2014 and since then were widely used to generate artificial data, such as images [20], videos [21], natural language [22], and music [23]. A GAN consists of two independent models: a generator and a discriminator, which both are composed of neural networks of different architectures (Fig. 3).

In general, the discriminator is used to determine whether the input data is real or not. The task of the generator is to capture the distribution of real data. The training process of this framework is similar to a min-max two-player game, in which the ultimate goal is that the distribution of generated data is as consistent as possible with the distribution of raw data. The loss function of the original GAN is defined as

$$\min_G \max_D V(D,G) = E_{x\sim P_r}[log(D(x))] + E_{x\sim P_g}[log(1-D(x))], \tag{1}$$

where the first instance of $D(x)$ represents the discriminator of real data and the second one represents the discriminator of generated data. Both P_r and P_g are the distributions of real and generated data, respectively.

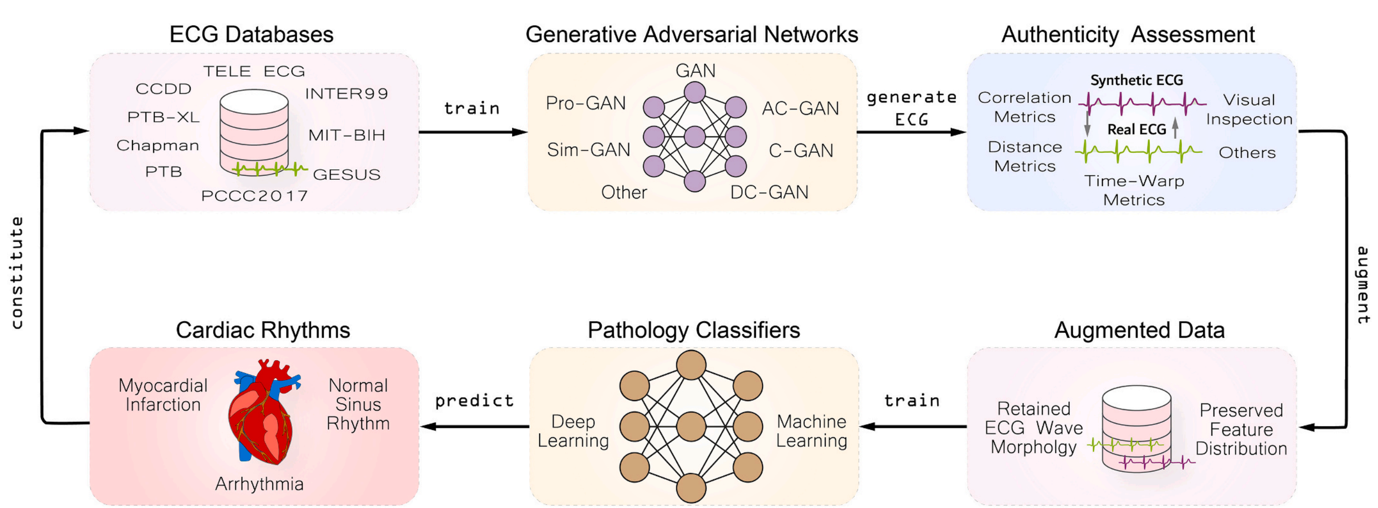

Fig. 1. Schematic representation of the data analysis pipeline used in the selected publications of the review.

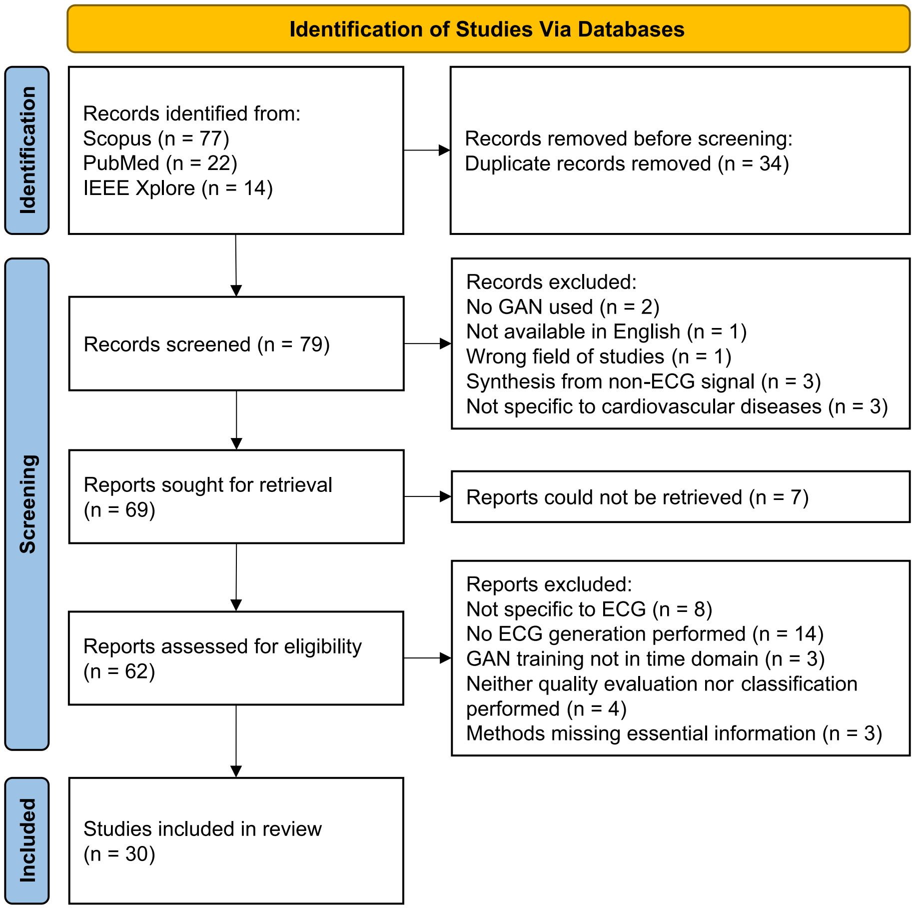

Fig. 2. PRISMA flowchart of the literature screening process.

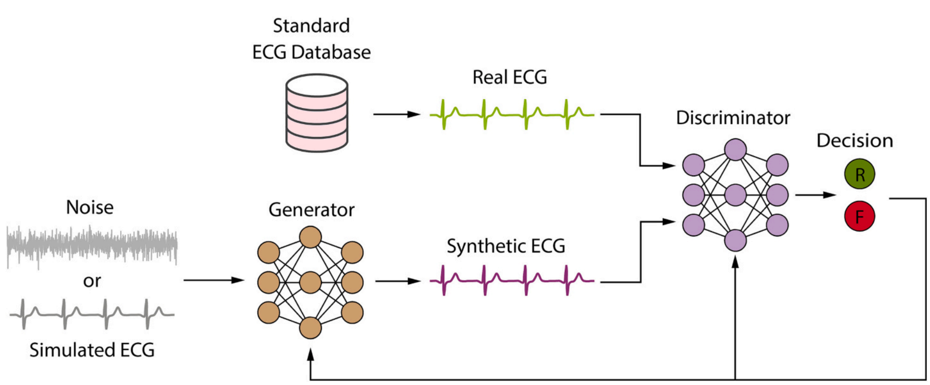

Fig. 3. Schematic representation of a generative adversarial network architecture.

When optimizing this model, the first instance of $D(x)$ should converge close to one while the second instance should approach zero.

The selected publications in this review were analyzed to identify different variations of GAN architectures that were employed for ECG signal generation. A total of seven different variations of GAN architecture were found in the selected studies: basic GANs, auxiliary-classifier GANs (AC-GANs), conditional GANs (C-GANs), deep-convolutional GANs (DC-GANs), simulator-based GANs (Sim-GANs), progressive-growing GANs (Pro-GANs), and other mixed GAN types (Table 1). Although all variations are based on the fundamental GAN architecture, some show slight deviations. While the generator network of a basic GAN receives only a latent vector of random noise as input, AC-GANs also receive conditional information (e.g., class labels) as input to the generator [24]. C-GANs use that additional information not only as input to the generator but also to the discriminator [25]. This enables the distinct generation of underrepresented classes to rebalance datasets. DC-GANs incorporate convolutional layers, batch normalization, and rectified linear unit (ReLu) activation functions to stabilize training of the GAN [26]. Sim-GANs receive synthetic signals as input to the generator rather than using real data [27]. Pro-GANs are designed to evolve iteratively by adding weights to the network layers and optimizing the network structure by the addition and removal of new layers [28].

Table 1
Overview of the generative adversarial network architectures used in the selected publications.

Network architecture	Abbreviation	Description
Basic GAN [17]	GAN	Basic architecture with random noise as generator input.
Auxiliary-classifier GAN [24]	AC-GAN	Conditional information (e.g., class labels) is additionally provided as generator inputs.
Conditional GAN [25]	C-GAN	Conditional information (e.g., class labels) is additionally provided as generator and discriminator inputs.
Deep-convolutional GAN [26]	DC-GAN	Convolutional layers, batch normalization, and ReLu activation functions are added to the architecture to stabilize training of the GAN.
Simulator-based GAN [27]	Sim-GAN	Synthetic signals are used as inputs to the generator.
Progressive-growing GAN [28]	Pro-GAN	Designed to iteratively optimize the network structure by addition and removal of layers.
Other	Other	A mix of different network architectures.

5. GAN quality evaluation metrics

It is crucial that the GAN-generated ECG signals properly resemble the morphologies associated with distinct types of pathologies similar to real-world data. In total, four publications [29–32] included in this review used visual assessment of the generated signals as well as the loss function of the GAN as quality evaluation metrics of their generated signals. While this method holds valuable information regarding the convergence of the model, it does not provide a metric of the resemblance of morphologies. Furthermore, the possibilities for visual evaluation of the signals by an expert are limited by the time required and the amount of generated data. Moreover, this is a qualitative approach prone to subjective bias and does not allow for quantification of the quality of the synthetic signals, in which is crucial in comparing the performance of different GAN architectures for generating authentic ECGs. Therefore, quantitative and meaningful quality evaluation metrics of synthetic ECG data in terms of the resemblance of (patho-) physiological real-world morphologies are required.

The quantitative metrics found in the selected publications were divided into three groups: (1) calculation of distance metrics between real and synthetic ECG signals; (2) calculation of correlation metrics between real and synthetic ECG signals; and, (3) other metrics such as the quantification through the performance of a classifier trained on GAN-generated data in classification of real-world ECG signals (Fig. 4). A detailed summary of the computation of the quality metrics used in the reviewed studies is available in the Supplementary material Tab. 1.

5.1. Distance-based metrics

In the application of DNNs, a widespread approach to assess the capability of a model to replicate real-world data is to calculate the error or distance between the predicted and yet unseen original data not involved in the training process. A total of eleven studies included in this review used distance-based metrics to quantify the resemblance of the GAN-generated ECG signals to the real-world data. The distance metrics predominantly employed in the studies of this review were the Euclidian distance (ED) [33–35], the root mean square error (RMSE) [36–38], and the percentage root mean square error (PRD) [36–39]. Other applied metrics noted were the mean absolute error (MAE) and mean square error (MSE). Additionally, other distance-based metrics were used to quantify the quality of the predicted signals including the Fréchet distance (FD), which calculates the similarity between two curves taking into consideration the location and the order of the points along the curves [40]. The Fréchet inception distance (FID), kernel inception distance (KID), Kullback-Leibler divergence (KLD), and maximum mean discrepancy (MMD) were used to calculate the distance between the probability distributions of two signals [41]. Dynamic time warping (DTW) metrics were also applied in several of the selected publications. In general, DTW yields a discrete matching between existing elements of one time series to another which poses a major improvement compared to other distance metrics [42]. Different variations of DTW were used in the selected publications to quantify the quality of the generated ECG signals including multivariate dynamic time warping (MDTW), soft dynamic time warping (SDTW), and the time warp edit distance (TWED). The SDTW is essentially a smoothed formulation of DTW that computes the soft minimum of all alignment costs [43]. The TWED is a

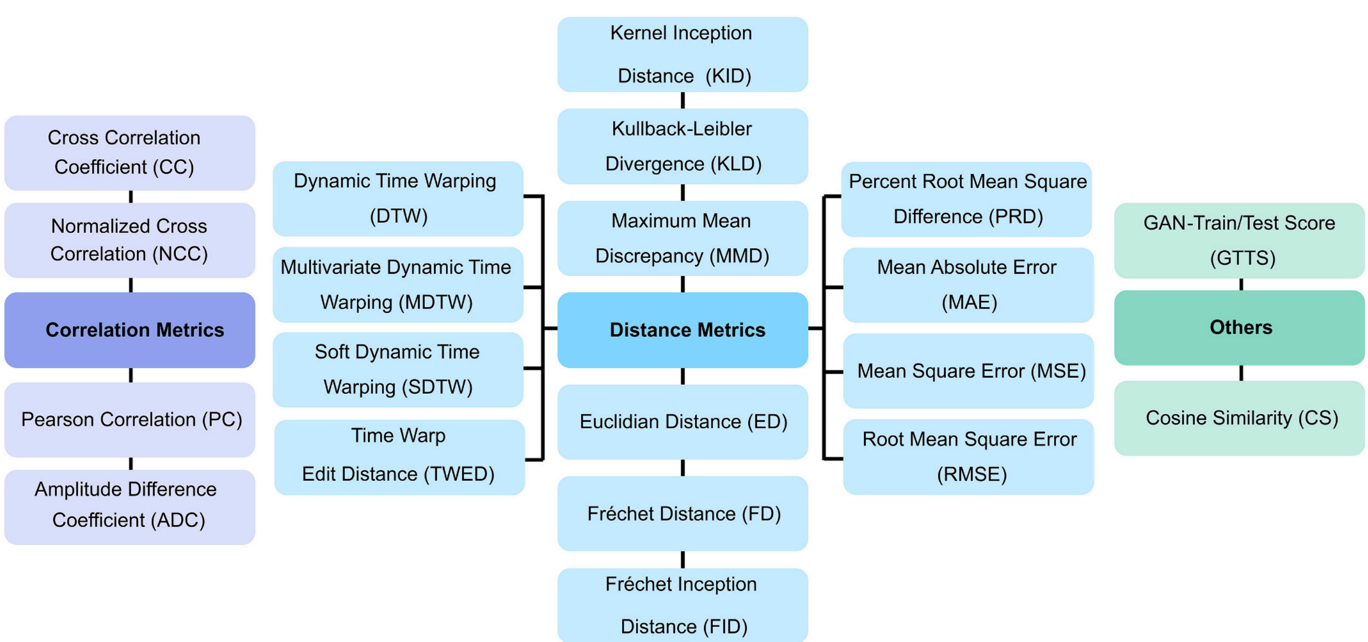

Fig. 4. Overview of quality assessment metrics for generative adversarial network-generated electrocardiograms divided into three subgroups: correlation metrics, distance metrics, and other assessment metrics.

time elasticity metric for discrete time series matching. Lastly, the MDTW is a modified version of the classical approach, adding additional dimensions to the input.

5.2. Correlation-based metrics

Correlation-based metrics were less frequently applied than distance-based metrics in the selected studies. However, two correlation-based metrics, namely normalized cross-correlation (NCC) and Pearson correlation (PC) were employed [34,37,39,44]. It is important to note that PC considers only the amplitudes over the entire signals but omits a possible time shift, because the data is not compared in regards to time. The amplitude difference coefficient (ADC) is mathematically defined as the sum of the product of corresponding values from real and generated time series divided by the sum of the squared value of the real data. Essentially, it can be regarded as the regression coefficient of two time series [45]. While correlation-based metrics are suitable for quantification of signal similarity, they are unable to assess the quality of synthetic ECG signals in terms of the authenticity of ECG wave morphologies.

5.3. Other metrics

There were multiple other metrics employed by some of the reviewed studies that did not fit one of the above categories. One example is the GAN-train/test score (GTTS) that was introduced by Shmelkov et al. [46]. The GAN-train is defined as the accuracy of the classifier trained on the generated data and evaluated on a validation set of real data. Where the GAN-test is the accuracy of a classifier trained on the original training set but tested on the generated data. The cosine similarity (CS) was only applied by a single publication [44]. The CS measures the similarity between two time series which are viewed as vectors in an inner product space by calculating the cosine of the angle between them. Therefore, the CS is independent of the magnitudes of the vectors but depends only on their angle [47].

6. ECG databases used for GAN training

Twelve different standard human ECG databases were used in the studies included in this review. Detailed information on availability, pathologies, lengths of ECG signals, sampling frequency, bit resolution, number of leads, and number of subjects is presented in Table 2. For further analysis, the databases were grouped by their availability.

6.1. Publicly available databases

The majority (n = 23) of the publications [27,29,32–37,44,60–74] employed the popular and publicly available MIT-BIH-Arrhythmia (MIT-BIH-A) database [48], which comprises 2-lead ECG recordings of normal sinus rhythms, paced beats, and 15 different types of arrhythmias. The MIT-BIH-A database is regarded as the gold standard for arrhythmia databases but requires preprocessing in terms of R-peak based beat segmentation due to the recording time of 30 min per record which is unsuitable for input for GAN architectures. A limited number (n = 5) of these studies [27,60–62,71] used a subset of arrhythmias available in the MIT-BIH-A database. Two studies [67,70] also used the MIT-BIH Normal Sinus Rhythm (MIT-BIH-NSR) database [49] to merge with other databases to create a larger dataset with additional regular sinus rhythms included. The PTB diagnostic ECG database [50] was employed in two studies [39,75] and includes 15-lead ECG recordings of normal sinus rhythms and arrhythmias as well as ECG signals of patients with heart failure, myocardial hypertrophy, MI, myocarditis, and valvular heart disease. The PTB-XL database [51] is not just an addition to the PTB but a database on its own, which was also used by two studies [38,76] and includes 12-lead ECG recordings of normal sinus rhythms, arrhythmias, myocardial hypertrophy, MI, and ST/T changes. The Chapman [52] and the TELE ECG [54] databases were used only by a single study [70,76]. The Chapman [52] database includes 12-lead recordings of normal sinus rhythms, arrhythmias, tachycardia, and bradycardia; and the TELE ECG [54] databases includes 1-lead recordings of unspecified ECGs.

6.2. Non-publicly available databases

More than one of the studies included in this review used non-publicly available databases. In the study of Thambawita et al. [30], the authors used the GESUS database [56] which includes 12-lead recordings of normal sinus rhythms. The authors used the INTER99 database [57] in their study, whereby no information on the included types of pathologies is available. The CCDD database [58] includes ECG signals of an unknown number of leads for normal sinus rhythms and was used as an additional dataset in the study of Zhang and Babaeizadeh [38]. The authors also employed the CSE database [59] in their work, which includes ECG signals of an unknown number of leads for

Table 2
Overview of electrocardiogram databases used in the studies.

Name	Country	Availability	Length	Sampling rate [Hz]	Resolution [bit]	Leads	Subjects	Pathologies
MIT-BIH-A [48]	USA	Public	30 min	360	11	2	47	Arrhythmia, sinus rhythm, paced rhythm
MIT-BIH-NSR [49]	USA	Public	~24 h	128	12	2	18	Sinus rhythm
PTB [50]	GER	Public	10 s	1000	16	15	290	Arrhythmia, heart failure, myocardial hypertrophy, sinus rhythm, myocardial infarction, myocarditis, valvular heart disease
PTB-XL [51]	GER	Public	10 s	500	16	12	18,885	Arrhythmias, myocardial hypertrophy, myocardial infarction, sinus rhythm, ST/T change
Chapman [52]	CHN	Public	30	500	32	12	10,646	Arrhythmia, sinus rhythm, tachycardia, bradycardia
PCCC2017 [53]	USA	Public	9–60 s	300	16	6	n.a.	Arrhythmia, sinus rhythm
TELE ECG [54]	USA	Public	~48 min	500	n.a.	1	250	n.a.
CPSC2018 [55]	CHN	Public	6–60 s	500	n.a.	12	6877	Arrhythmia, sinus rhythm
GESUS [56]	DEN	Non-public	n.a.	150	n.a.	12	10,618	Sinus rhythm
INTER99 [57]	DEN	Non-public	n.a.	n.a.	n.a.	n.a.	60,000	Sinus rhythm
CCDD [58]	CHN	Non-public	n.a.	500	n.a.	12	n.a.	n.a.
CSE [59]	n.a.	Non-public	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

n.a. – not available.

unspecified types of pathologies.

7. Detailed review of selected studies

For a detailed analysis, the included studies ($n = 30$) were divided into three groups: (1) GAN-generated signal quality was quantified but no classification was performed ($n = 9$) [33,34,36,37,39,61,66,74,76]; (2) classification was performed but GAN-generated signal quality was not quantified ($n = 11$) [27,32,60,62–65,69,70,72,77]; and, (3) GAN-generated signal quality was quantified and classification was performed ($n = 10$) [29–31,35,44,62,67,71,75,78]. All included publications were analyzed based on the employed GAN architectures. For groups (2) and (3), additionally, the used classifier architectures and the GAN-generated signal quality metrics were analyzed. While publications are presented in Table 3 in chronological order, detailed analysis of the studies in their respective groups was performed based on similarity of methodology. Furthermore, classification performances before and after database augmentation using GAN-generated data has been summarized in Table 4.

7.1. Studies quantifying signal quality omitting classification

In 2019, Zhu et al. [36] were among the first to generate ECG signals using GANs. They proposed a bidirectional long short-term memory

Table 3
List of the selected studies included in this review with details on the used generative adversarial network architecture, databases, pathologies, generated signal length, preprocessing steps and whether electrocardiogram signal quality was quantified, and classification was performed.

Year	Study	GAN architecture	Database	Pathologies	Signal length	Preprocessing	Quality quantified	Classification performed
2019	Zhu et al. [36]	GAN	MIT-BIH-A	N, S, V, F, Q	0.6–1 s	n.a.	Yes	No
2019	Wang et al. [29]	AC-GAN	MIT-BIH-A, CPSC2018	N, S, V, F	0.6–60 s	Beat segmentation, filtering	Yes	Yes
2019	Lee et al. [39]	GAN	PTB	All of the database	1 s	Down sampling	Yes	No
2019	Delaney et al. [74]	GAN	MIT-BIH-A	None	1 s	Gain-removal, down sampling, normalization, beat segmentation	Yes	No
2019	Ye et al. [33]	DC-GAN	MIT-BIH-A	N, S, V, F, Q	0.5–2 s	Normalization	Yes	No
2020	Shaker et al. [63]	GAN	MIT-BIH-A	N, S, V, F, Q	0.83 s	Band-pass filtering, normalization, beat segmentation	No	Yes
2020	Golany et al. [60]	DC-GAN	MIT-BIH-A	N, S, V, F	0.6 s	n.a.	No	Yes
2020	Hatamian et al. [77]	DC-GAN	PCCC2017	N, AF	5 s	Beat segmentation	No	Yes
2020	Wulan et al. [61]	Other	MIT-BIH-A	N, LBBB, RBBB	4–20 s	Band-pass filtering, data rebalancing	Yes	No
2020	Golany et al. [27]	Sim-GAN	MIT-BIH-A	S, V, F	0.6 s	n.a.	No	Yes
2020	Lan et al. [65]	DC-GAN	MIT-BIH-A	N, S, V, F, Q	6 s	Down sampling, high-pass filter	No	Yes
2020	Nankani and Baruah [34]	DC-GAN	MIT-BIH-A	N, S, V, F, Q	0.72 s	Notch filtering, median filtering, beat segmentation, data augmentation	Yes	No
2020	Hazra and Byun [37]	GAN	MIT-BIH-A	N, S, V, F, Q	n.a.	Discrete wavelet transform, thresholding, z-score transformation, Beat segmentation	Yes	No
2020	He et al. [64]	DC-GAN	MIT-BIH-A	N, S, V, F, Q	0.69 s	Beat segmentation	No	Yes
2021	Zhou et al. [71]	AC-GAN	MIT-BIH-A	S, V	0.35 s	Beat segmentation	Yes	Yes
2021	Kim and Pan [44]	AC-GAN	MIT-BIH-A	N, S, V, F, Q	0.48 s	n.a.	Yes	Yes
2021	Yang et al. [35]	Pro-GAN	MIT-BIH-A	N, S, V, F, Q	0.71 s	Beat segmentation	Yes	Yes
2021	Dasgupta et al. [66]	DC-GAN	MIT-BIH-A, MIT-BIH-NSR	N, S, V, F, Q	2.84 s	Down sampling	Yes	No
2021	Golany et al. [62]	Sim-GAN	MIT-BIH-A	S, V, F	0.6 s	n.a.	No	Yes
2021	Rath et al. [69]	DC-GAN	MIT-BIH-A, PTB	N, S, V, F, Q	0.6–1 s	n.a.	No	Yes
2021	Zhang and Babaeizadeh [38]	GAN	PTB-XL, CCDD, CSE, Chapman	LBBB, LVH, ACUTMI	10 s	Down sampling	Yes	Yes
2021	Zhou et al. [70]	C-GAN	MIT-BIH-A, MIT-BIH-NSR, PCCC2017, TELE ECG	All of all databases	1 s and 10 s	Down sampling, beat segmentation	No	Yes
2021	Hossain et al. [68]	C-GAN	MIT-BIH-A	N, S, V, F	0.77 s	Beat segmentation, normalization	Yes	Yes
2021	Thambawita et al. [30]	DC-GAN	GESUS, Inter99	None	10 s	n.a.	Yes	Yes
2021	Brophy et al. [67]	GAN	MIT-BIH-A, MIT-BIH-NSR	N, S, V, F, Q	5 s	Down sampling, R-peak alignment, and segmentation	Yes	Yes
2022	Li et al. [75]	DC-GAN	PTB	MI	10 s	n.a.	Yes	Yes
2022	Seo et al. [76]	Other	Chapman, PTB-XL	LVH, RBBB, LBBB, WPW, RVH, MI	2.5 s	None	Yes	No
2022	Wang et al. [32]	Sim-GAN	MIT-BIH-A	N, S, V, F, Q	10 s	None	No	Yes
2022	Islam et al. [73]	GAN	MIT-BIH-A	N, S, V, F, Q	0.6 s	Beat segmentation, Band-pass filtering, wavelet filtering, z-score transformation	Yes	Yes
2022	Ma et al. [72]	DC-GAN	MIT-BIH-A	N, S, V, F, Q	0.41 s	Discrete wavelet transform, beat segmentation	No	Yes

LVH – left ventricular hypertrophy; RVH – right ventricular hypertrophy; RBBB – right bundle branch block; LBBB – left bundle branch block; WPW – Wolff-Parkinson-White-syndrome; MI – myocardial infarction; ACUTMI – acute myocardial infarction; N – normal beat; S – supraventricular premature beat; V – premature ventricular contraction; F – fusion beat; Q – other beats; AF – atrial fibrillation; n.a. – not available.

Table 4
Summary of the classification performance before and after augmentation of studies performing a classification.

Study	GAN architecture	Database	Classification performance			Augmented data (% database size)
			Parameter	Before augmentation	After augmentation	
Wang et al. [29]	AC-GAN	MIT-BIH-A, CPSC2018	Accuracy:	99 %	99 %	77 %
			Sensitivity:	99 %	99 %	
			Specificity:	99 %	99 %	
Shaker et al. [63]	GAN	MIT-BIH-A	Sensitivity:	95 %	97 %	n.a.
			Specificity:	95 %	97 %	
			Precision:	85 %	94 %	
Hatamian et al. [77]	DC-GAN	PCCC2017	Accuracy:	90 %	93 %	n.a.
			F1-Score:	86 %	87 %	
Lan et al. [65]	DC-GAN	MIT-BIH-A	Sensitivity:	99 %	97 %	178 %
			Specificity:	99 %	99 %	
			F1-Score:	97 %	97 %	
He et al. [64]	DC-GAN	MIT-BIH-A	Accuracy:	62 %	91 %	n.a.
Yang et al. [35]	Pro-GAN	MIT-BIH-A	Accuracy:	96 %	99 %	n.a.
			Sensitivity:	75 %	99 %	
			Precision:	86 %	93 %	
Rath et al. [69]	DC-GAN	MIT-BIH-A, PTB	Accuracy:	97 % and 97 %	99 % and 99 %	39 % and 75 %
			F1-Score:	94 % and 97 %	99 % and 99 %	
			AUROC:	96 % and 92 %	98 % and 99 %	
Li et al. [75]	DC-GAN	PTB	Accuracy:	89 %	99%	43 %
			Sensitivity:	84 %	99 %	
			Specificity:	99 %	99 %	
			Precision:	99 %	99 %	
			F1-Score:	91 %	99 %	
Wang et al. [32]	Sim-GAN	MIT-BIH-A	F1-Score	88 %	92 %	n.a.
Islam et al. [73]	GAN	MIT-BIH-A	Accuracy:	97 %	98 %	n.a.
			Sensitivity:	97 %	97 %	
			Precision:	95 %	96 %	
			F1-Score:	96 %	96 %	

n.a. – not available.

(BiLSTM) network as the generator and a convolutional neural network (CNN) as the discriminator for the synthesis of single-beat ECG signals of 0.6–1 s length based on raw data from the MIT-BIH-A database. They compared their proposed network to other GAN architectures using different variants of autoencoders. The authors not only showed that for the BiLSTM-CNN GAN, the loss function converged the fastest. It was consistently superior to the other tested architectures concerning the quality of the synthesized ECG signals with FD = 0.76, PRD = 66.41 %, and RMSE = 0.28 being at least 45 %, 45 %, and 29 % lower than those of the other tested architectures, respectively. Overall, the ECG signals synthesized using their method showed great authenticity with high morphological similarity to the original data.

In a similar approach, Delaney et al. [74] also used the MIT-BIH-A database and investigated two different generator (LSTM and BiLSTM) and discriminator (LSTM and CNN) architectures in the generation of regular sine waves as well as ECG signals. However, in contrast to others, they included only normal sinus rhythms, therefore omitting any pathologies. Gain removal, down sampling, and R-peak alignment were performed before GAN training. Similar to other publications in this group, they generated ECG signals of about 1 s duration. The authors showed that a GAN employing a LSTM generator and CNN discriminator performed best resulting in MMD $<$ 0.01 and DTW = 11.66. They concluded that DTW is a more suitable evaluation metric than MMD, since it has a higher sensitivity to the relative scale between the synthetic and test data and is furthermore more robust if training fails.

Hazra and Byun [37] proposed a network architecture to generate not only ECGs but also electroencephalograms, photoplethysmograms as well as electromyograms. They used bidirectional grid LSTM units and termed their network “SynSigGAN”. They used the MIT-BIH-A database and applied the discrete wavelet transformation, thresholding, inverse discrete wavelet transformation, z-score transformation, and beat segmentation prior to training. Their approach performed very well in terms of generated signal quality with FD = 0.94, MAE = 0.22, PC = 0.99, PRD = 6.34 %, and RMSE = 0.13. Although the authors compared the performance of their architecture to that of other GANs with the conclusion that their approach performs significantly better, it remained unclear whether the comparison was done for combined measures for all types of physiological signals or for single modalities.

Over time, GAN architectures have evolved and Nankani and Baruah [34] used a DC-GAN. They trained on the MIT-BIH-A database and applied various techniques such as a notch and median filter, beat segmentation, and traditional data augmentation methods to the data before training. The authors of the study investigated the impact of different loss functions (Wasserstein, least squares, binary cross entropy) and input noise distributions (uniform, Gaussian) on the quality of the generated ECG signals for two separate subsets of beat classes: (1) normal sinus beat (N), supraventricular premature beat (S), premature ventricular contraction (V), and fusion beat (F), and (2) S, V, F. For atrial arrhythmia, the GAN with a uniform noise input and a cross-entropy loss function outperformed all other configurations with FID = 12.47, MMD = 0.1, ED = 0.22, KLD = 0.08. For ventricular arrhythmias, the GAN using uniform noise input, cross-entropy loss function, Gaussian noise input, and least squares loss function performed the best, with FID = 14.41, MMD = 0.11, ED = 0.24, KLD = 0.08, and TWED = 5.13. It is interesting to note that this study is the only one that investigated the effect of uniform noise inputs on the generated signal quality in GANs.

Also, Dasgupta et al. [66] utilized a DC-GAN to augment the MIT-BIH-A database and named their model “CardioGAN”. The authors compared their model to autoencoder models and to the model of Zhu et al. [36]. Compared to the work of Zhu et al. [36], the authors reported generated signals with a PRD = 38.57 % (−42 %), RMSE = 0.16 (−43 %), FD = 0.61 (−20 %), and DTW = 4.15 (−56 %), therefore showing increased quality of their generated ECG signals.

Ye et al. [33] used a DC-GAN based on sequence GAN algorithm with a policy gradient reinforcement learning to synthesize ECG signals from only female records (n = 5) in the MIT-BIH-A database. Female data is underrepresented in the MIT-BIH-A database and synthesizing them could improve gender-specific data balancing. Generated signals with lengths of 0.5 s, 1 s, 1.5 s, and 2 s were compared to three other networks (Wasserstein GAN using weight clipping, Wasserstein GAN using gradient penalty, and a BiLSTM-CNN GAN similar to Zhu et al. [36]), and their model outperformed the others, with the exception of Wasserstein GAN with weight clipping in ED for generated signals of length 1 s and 1.5 s. They also assessed the TWED, DTW, and SDTW for different signal lengths, concluding that all three time-warping metrics increased with signal length. They found a TWED between 27.47 and 112.54 for the four generated signal durations for one of the five clinical records, while DTW and SDTW showed the lowest values (81.24–323.32 and 46.25–203.75) for another clinical record. Lee et al. [39] applied a basic GAN to synthesize 5-lead ECG signals from limb leads using the PTB database, down sampling the data, and applying ordered time series embedding before training. Their approach produced ECG signals of 1 s duration with high signal quality (PRD = 7.21 %, PC = 1.00, ADC = 0.02). The generated signals were visually compared to signals of similar pathologies and found no visually distinguishable differences.

Also, Seo et al. [76] proposed a GAN for the generation of multi-lead signals from single-lead ECG. They employed the PTB-XL and the Chapman databases separately, then combined them to train a GAN with a U-Net generator and a PatchGAN discriminator. Data was segmented based on R-peaks into signals of 2.5 s length, including six different pathologies (left ventricular hypertrophy (LVH), left and right bundle branch block (LBBB and RBBB), Wolff-Parkinson-White-syndrome (WPW), right ventricular hypertrophy (RVH), and MI). For further evaluation, the 2.5 s segments were stacked to form the original recording length of 10 s. Overall, the proposed network showed satisfactory performance with FD = 6.70 and MSE = 0.02. Additionally, the generated signals were labeled by a specialist and a percentage of correctly labeled signals was given as a measure of generated signal quality (LVH = 95 %, RBBB = 89 %, LBBB = 86 %, WPW = 63 %, RVH = 20 %, and MI = 0 %).

Wulan et al. [61] compared three models for ECG signal generation: the WaveNet-based model, the SpectroGAN model, and the WaveletGAN model. The WaveNet-based GAN preprocesses the data using a μ-law companding transformation followed by a sequence of convolutional layers with dilation. The SpectroGAN and WaveletGAN models use short-term Fourier transform and stationary wavelet transform, respectively. Only N beat classes, LBBB pathologies, and RBBB pathologies were extracted from the MIT-BIH-A database, then band-pass filtered, and segmented by R-peaks. Duration of generated signals was between 4 and 20 s. The WaveletGAN model performed best in terms of the quality of generated ECGs with a GTTS = 89.07 % and 92.33 %. The authors highlighted that their WaveNet-based GAN could produce continuous ECG signals of up to 20 s rather than concatenating single-beat ECGs, which they considered a major advantage over other approaches. Nonetheless, the conclusiveness of their study is compromised by the rather small sample size.

7.2. Studies with classification without signal quality quantification

Shaker et al. [63] used a GAN to generate ECG signals to augment the MIT-BIH-A database. Classification was performed directly into 15 arrhythmias using a CNN and a two-step approach. The two-step approach classified a CNN into the five main beat classes (N, S, V, F, other beats (Q)), and further classified individual CNNs into 15 subcategories. Prior to training, a bandpass filter, zero-one normalization, and beat segmentation were executed. Data augmentation was performed by adding 47 % of the original data to the database. For the direct classification approach, the average precision increased from 81 % to 90 % (+9 %), sensitivity stayed at 99 %, and specificity increased from 98 % to 99 % (+1 %). When classifying into the five main categories, the average precision increased from 90 % to 94 % (+4 %), sensitivity stayed at 98 %, and specificity increased from 97 % to 98 % (+1 %). When classifying further into the 15 subcategories, the average precision increased from 85 % to 94 % (+9 %), sensitivity from 95 % to 98 % (+3 %), and specificity from 95 % to 97 % (+2 %). Additionally, comparison of GAN-based augmentation with four other methods (weighted loss, random oversampling, synthetic minority oversampling technique, and adaptive synthetic sampling) was performed. Compared to the second-best augmentation method (random oversampling), the GAN-based augmentation reached 90 % precision (+4 %), 99 % (+0 %) sensitivity, and 99 % (+0 %) specificity.

Golany et al. [60] investigated the possible improvement of an LSTM-based classifier for the detection of atrial arrhythmias (N, S, V, F) through GAN-based data augmentation using the MIT-BIH-A database. They employed a DC-GAN to generate ECG signals of 0.6 s length. They evaluated classification performance using a fully connected network, an LSTM network for augmentation of the database with jittered samples, and GAN-generated signals with a maximum of 15,000 samples (+33 %) added. Overall, by maximum augmentation with the GAN-generated signals, the classification performance in terms of the area under the receiver operating curve (AUROC) increased from 87 % to 91 % (+4 %), from 82 % to 92 % (+10 %), from 97 % to 98 % (+1 %), and from 95 % to 96 % (+1 %) for classification of N, S, V, and F beat classes. In contrast, data augmentation with jittered ECG samples reduced the classification results for all beat types.

Also, He et al. [64], Ma et al. [72], and Lan et al. [65] used DC-GANs to augment the MIT-BIH-A database to increase pathology classification performance. He et al. [64] employed pathology classification in a mobile device together with cloud computing and applying beat segmentation to the data resulting in signals of length 0.69 s. When compared to non-augmented datasets, the authors reported an average increase in accuracy from 62 % to 91 % (+29 %). No comment on the amount of added data was given.

Ma et al. [72] augmented with 157 % of the original data by creating ECG signals of approximately 0.41 s length. The data was filtered using a discrete wavelet transform and segmented by R-peaks prior to training. Compared to the non-augmented dataset the authors reported an increase in accuracy from 94 % to 99 % (+5 %), in sensitivity from 96 % to 98 % (+2 %), and in specificity from 92 % to 99 % (+7 %).

Lan et al. [65] created ECG signals of 6 s duration by down sampling and high-pass filtering the data before training, and augmenting the signal with 178 % of the original size. The authors compared three pretrained nets (LeNet, AlexNet, ZFNet and VGG13) for classification.

The best performing classifier showed a <1 % increase in sensitivity, specificity, and F1-score when compared to the non-augmented dataset. In contrast to other publications in this group, Zhou et al. [70], Hatamian et al. [77], and Rath et al. [69] used different databases to augment and applied different GAN architectures. Zhou et al. [70] used four different databases merging the PCC2017 and TELE ECG into a combined set (COMD) and the two MIT-BIH databases into another combined set (RECD). Data was resampled and beat segmentation was performed. Data was generated with durations of 1 s and 10 s. The COMD database was used to train the C-GAN, classification was performed on a combination of all databases. Augmentation was performed by doubling the original amount of data. When compared to non-augmented datasets accuracy increased from 94 % to 97 % (+3 %), sensitivity from 97 % to 99 % (+2 %), and specificity from 93 % to 96 % (+3 %) for the COMD database; and from 94 % to 96 % (+2 %), 98 % to 99 % (+1 %), and 92 % to 95 % (+3 %) for the RECD database for accuracy, sensitivity, and specificity.

Hatamian et al. [77] used a DC-GAN to augment the PCCC2017 database with ECG samples of 5 s length. Beat segmentation was performed prior to training and the time series was converted into frequency domain. Oversampling, Gaussian mixed models, and the DC-GAN generated signals were compared as augmentation strategies for classification. In comparison with the non-augmented dataset, the F1-score increased from 86 % to 87 % (+1 %) and accuracy on average for both beat classes increased from 90 % to 93 % (+3 %).

Rath et al. [69] developed an ensemble model of an LSTM classifier and DC-GAN to augment the MIT-BIH-A and PTB databases. The authors compared classification performance to standard machine learning (support vector machine, naïve Bayes, and multi-layer perceptron) and a

LSTM network. Data was segmented by R-peaks resulting in ECG signals of 0.6 s to 1 s length. Databases were augmented with 39 % and 75 % for the MIT-BIH-A and the PTB database, respectively. Overall, when compared to non-augmented datasets, the authors reported an increase in accuracy from 97 % to 99 % (+2 %), in F1-Score from 94 % to 99 % (+5 %) and in AUROC from 96 % to 98 % (+2 %) for the MIT-BIH-A database. For the PTB database, accuracy increased from 97 % to 99 % (+2 %), F1-Score from 97 % to 99 % (+2 %) and in AUROC from 92 % to 99 % (+7 %).

In 2021, Golany et al. [27] proposed using ECG signals as an input to the generator of the GAN rather than noise. The authors aptly named their approach Sim-GAN. This study employed the raw ECG traces of the MIT-BIH-A database to train the GAN to generate S, V and F beats of 0.6 s length. To evaluate the performance of their approach, the authors used Sim-GAN-generated ECGs for data augmentation to improve the performance of a ResNet arrhythmia classifier. They compared the augmentation performance that resulted from using the Sim-GAN to that using other approaches including ResNet, basic GANs, DC-GANs, and Sim-GANs without convolutional layers. The classifier performance was quantified by sensitivity and precision. Augmenting with Sim-GAN was compared to a non-augmented dataset, and it was found that precision increased from 43 % to 80 % (+37 %) for the S beat class, from 79 % to 84 % (+5 %) for the V beat class, and from 4 % to 40 % (+36 %) for the F beat class. However, sensitivity did not change between any of the tested GAN-architectures.

Based on their previous work, Golany et al. [62] developed a Sim-GAN with ECG signals created by a set of ordinary differential equations as input to the generator using the S, V, and F beat classes of the MIT-BIH-A database without preprocessing the data. The quality of the generated ECG signals was assessed visually by comparison of the morphological differences between individual samples of the ECG signals and samples of the database. The authors compared the classification performance with four datasets: (1) signals augmented with synthetically generated ECG signals of their proposed model to their previous work of the Sim-GAN [27], (2) a DC-GAN, (3) a Wasserstein GAN, and (4) the generated samples of the ordinary differential equation generator. Compared to a non-augmented classification performance of another study by Kachuee et al. [79], it was found that precision increased from 2 % to 83 % (+81 %) for the S beat class, from 88 % to 90 % (+2 %) for the V beat class, and from 4 % to 45 % (+41 %) for the F beat class. Sensitivity did not change between any of the tested GAN-architectures. The authors concluded that they have shown how an ordinary differential equation model describing cardiac cycles can learn with a GAN and they can further improve classification performance.

Also, Wang et al. [32] employed a Sim-GAN to synthesize ECG signals based on the MIT-BIH-A database. In this approach, akin to the method employed by Golany et al. [27], the generator receives a simulated ECG as input and refined it to better resemble real-world data. They employed a DC-GAN-based architecture for both the generator (refiner) and the discriminator using ResNet blocks as convolutions. They optimized the network using Easy Cartesian Genetic Programming to evolve the architecture and training hyperparameters. The data were split into 10-s segments of normal and abnormal ECG signals. The Sim-GAN was trained on 32 hand-selected pathological ECGs. To assess the performance, they trained multiple classifiers including a 1D-SqueezeNet to distinguish normal from abnormal ECGs using real and augmented real data. The augmentation with synthetic data of the optimized Sim-GAN improved classification performance in terms of F1-score from 88 % to 92 % (+4 %).

7.3. Studies with signal quality quantification and classification

Islam et al. [73] used a GAN to augment the MIT-BIH-A database using a hybrid framework based on a bidirectional recurrent neural network with a multilayered dilated CNN for classification. Data was preprocessed using beat segmentation, bandpass filtering, and z-transformation prior to training. They used MMD to quantify the quality of the generated signals, however they did not report any results. When compared to non-augmented datasets using a LSTM classifier, their hybrid network with GAN augmentation increased accuracy, precision, sensitivity and F1-score <1 % on average for the test dataset.

Hossain et al. [68] used a C-GAN to synthesize 0.77 s beats using the MIT-BIH-A database. Prior to training, the data was segmented by R-peaks and zero-one normalization was performed. Signals showed a MSE = 0.00038, CC coefficient = 0.9986, and PRD = 0.0136 %. Classification results of the augmented database were compared with other studies that used non-GANs-based augmentation. The authors compared their results to another study by Acharya et al. [80] showing that accuracy increased from 80 % to 98 % (+18 %), sensitivity on average for from 79 % to 94 % (+15 %), and specificity from 89 % to 98 % (+9 %).

Wang et al. [29], Kim and Pan [44], and Zhou et al. [71] all used an AC-GAN to synthesize ECG signals. To evaluate single heartbeat performance, Wang et al. [29] used the MIT-BIH-A database, and then used the CPSC2018 database to evaluate consecutive heartbeat detection performance. Data was segmented by R-peaks and filtered before training and signals were generated with lengths between 0.6 s and 60 s. Quality of the generated signals using the MIT-BIH-A database was ED = 3.68, PC = 0.87, and KLD = 3.40. Augmentation of the MIT-BIH-A database was 77 % of the original size while the CPSC2018 database was augmented with 171 %. When classifying with a LSTM network using the MIT-BIH-A database, accuracy, sensitivity, and specificity were above 99 % without augmenting the dataset and increased <0.3 % on average when augmented with GAN-generated signals. Furthermore, no augmentation was compared to synthetic minority oversampling and to GAN-based augmentation with an increase of F1-Score from 0.524 to 0.718 (+37 %) and to 0.883 (+23 %), respectively.

Kim and Pan [44] created ECG signals of 0.48 s length without preprocessing using only the MIT-BIH-A database with an ensemble network design of parallel structures and compared the generated ECG signals to self-recorded ECG signals of 89 subjects. Furthermore, real data was substituted with generated data instead of augmentation. Classification accuracy was 96 % and 98 % for different ratios of generated and real data in the dataset. Signal quality was evaluated with CS = 99 % and ED = 0.25. The authors concluded that the generated synthetic ECG signals were similar to the real ECG signals even if the comparison data did not have the same size. Furthermore, they compared the classification performance of their ECG generation method on a test set of the MIT-BIH-A database, which resulted in an accuracy of 99 %, specificity of 99 %, and sensitivity of 99 %.

Zhou et al. [71] used the MIT-BIH-A database to generate synthetic ECG signals with segmented data resulting in 0.35 s long signals. The quality of the generated ECG signals was evaluated with FD = 42 (N beats), 58 (S beats), 91 (V beats), and 101 (F beats). Classification performance for the detection of S and V both resulted in accuracies of 99 %, sensitivities of 87 % and 93 %, specificities of 99 % and 99 %, precisions of 85 % and 94 %, and F1-Scores of 86 % and 93 % for S and V beats. Furthermore, the influence of different data augmentation rates (4 %, 16 %, 40 %, and 158 %) on the classification performance was compared with improvement by up to 5 % using an augmentation of 158 %.

Yang et al. [35] were the only group that used a Pro-GAN with the MIT-BIH-A database to create synthetic ECG signals with approximately 0.71 s length. The training process was divided into four parts, starting with ECGs with 32 samples and up to 256 samples in the last stage. New convolutional structures were added at each stage where a new high-resolution block complements the network with a weight (linearly increasing from zero to one) following a network parameter optimization. They reported an ED = 7.38, DTW = 6.44, PC = 0.99, and KLD = 0.01 and compared their results with Shaker et al. [63] and Golany et al. [60], where both reported superior results for ED and DTW. The database was augmented with 50,000 recordings (+53 %). Their Pro-GAN showed a GTTS = 93 % and 98 %, respectively. The augmented dataset with generated ECG signals resulted in an increase in accuracy from 96 % to 99 % (+2 %), in sensitivity from 75 % to 99 % (+24 %), and in precision from 86 % to 93 % (+7 %) when using a CNN classifier network.

Thambawita et al. [30] synthesized only regular sinus rhythms, similar to Delaney et al. [74]. The authors tested a WaveGAN, which is a stack of deconvolutional layers, and a Pulse2Pulse GAN, which is a U-net-based DC-GAN. Both WaveGAN and Pulse2Pulse GAN use a discriminator that is realized as a stack of convolutional layers, each followed by a Leaky ReLU-activation and a phase shuffle layer. A combined dataset of the GESUS and the Inter99 databases that were divided into 10 s segments was used. All 8-lead ECGs were converted to 12-lead ECGs by calculation of the four missing channels based on the first two leads using trigonometric functions. The quality of the synthetic signals was assessed by the ratio of which the MUSE 12SL algorithm classified the generated ECGs as normal sinus rhythms. Initially, 10,000 synthetic ECGs were generated from every 500 epochs until reaching 3000 epochs from both GAN architectures. The best-performing epoch was then used to generate 150,000 synthetic signals for the final evaluation which resulted in 81 % being deemed as normal ECGs by the MUSE 12SL algorithm. Finally, the authors also reported that the correlation between QT and R-R intervals closely resembled that of real-world data.

Li et al. [75] aimed to improve the performance of classifiers to detect MI through data augmentation using synthetic single-lead ECG signals generated by a DC-GAN using the PTB database, where they split the ECG records into segments consisting of 1024 samples. First, to assess the performance of their DC-GAN, they compared it to that of the well-established Wasserstein GAN (WGAN) and WGAN with gradient penalty (WGAN-GP). Second, to assess the quality of the synthetic ECGs, they trained a CNN-based classifier to detect MI with different-sized sets of real and augmented ECG signals. In terms of FID, with a value of 186.83, their network outperformed the WGAN (206.12) and WGAN-GP (188.97). Also, in terms of KID (0.2703 ± 0.0027), their GAN performed slightly better in a range of 1–10 % compared to the WGAN and WGAN-GP. By augmenting the original data consisting of 200 MI and 80 normal ECGs with 40 (+25 %) and 120 (+43 %) synthetic normal ECGs, the classification accuracy increased from 89 % to 95 % (+6 %) and to 99 % (+10 %), respectively. Sensitivity increased from 84 % to 94 % (+10 %) and to 99 % (+15 %), respectively. Overall, classification performance was above 98 % for all metrics for the augmented dataset. Finally, performance was compared to other machine learning approaches in 5-fold cross-validation approaches showing consistently better results in terms of accuracy, sensitivity, specificity, precision, and F1-score.

Zhang and Babaeizadeh [38] proposed a 2D BiLSTM GAN to synthesize 12-lead ECGs for LVH, LBBB, acute MI, and normal sinus rhythms. They used a combined database consisting of the PTB-XL, CCDD, CSE, and the Chapman databases. All ECG signals of 10-s length were re-classified using a commercial algorithm which served as a ground truth for the training and evaluation of their network. The evaluation of the network was done by classification of about 1000 generated ECG signals into pathologies, again using the same algorithm previously used for data re-classification. The success rates were 98 % for LVH, 93 % for LBBB, 79 % for acute MI, and 59 % for normal sinus rhythms. The statistical results suggest that the synthetic ECGs are neither overfit nor biased towards the training data and contain a large set of wave morphologies.

Brophy et al. [67] used a multivariate GAN and investigated five different loss functions (least squares, least square with DTW term, GAN with DTW criterion, loss sensitive GAN, loss sensitive GAN with DTW) being merged with the MIT-BIH-A and the MIT-BIH-NSR databases. Data were down sampled, and R-peak aligned prior to training. Furthermore, they compared their model to a variational autoencoder and an LSTM network in terms of creating ECG signals. Their best-performing model type was the GAN with a least square loss function together with a DTW distance metric. The authors evaluated a MDTW = 3.82 and an MMD = 0.01 in comparison with 47.14 and 0.96 for the variational autoencoder model and 27.87 and 1.09 for the LSTM model. Additionally, classification was performed using a support vector machine and a LSTM network. Again, the GAN with a least squares loss function together with a DTW distance metric performed best with an average accuracy of 51 % for both classifiers. This is in comparison with the variational autoencoder with 79 % and the LSTM with 78 %. Regarding this classification result, the authors wanted to demonstrate that the poorer the performance of the classifier, the closer the generated data is to the training data.

8. Discussion

This review summarizes the current literature dealing with synthetic ECG signal generation using GANs while taking into consideration commonly applied quality evaluation metrics and the impact of data augmentation with GAN-generated signals on classification performance. Although similar in their general approach, the abundance of selected publications shows particular differences in the quality evaluation metrics used and the application of classifiers.

Adding these synthetically generated signals to imbalanced datasets, such as the MIT-BIH-A database, has proven to increase the classification performance of commonly used DNN classifiers. Classification accuracy and sensitivity of augmented datasets reached values ranging from 97 to 98 %, with increases of 6–7 % compared to non-augmented training sets in four publications [35,44,71,75]. This leads to the conclusion that augmenting unbalanced datasets with GAN-generated data improves classification performance. In addition to traditional data augmentation techniques, the utilization of GANs for generating similar yet slightly distinct data has emerged as a valuable alternative. However, almost all selected publications use different parameters when training and evaluating their classification networks making direct comparisons challenging. Notably, there was a discernible disparity and lacking accompanying justification in the choice of suitable classification performance metrics as well as the omission of certain metrics that could have provided valuable insights. Hence, future studies highlighting classification performance should include the confusion matrix based on which the readers can potentially calculate the most commonly used metrics themselves.

When discussing data augmentation, it is worth questioning how much data is required to enhance classifier performance. Based on the selected publications a minimum of 40 % (with up to 171 %) of added data was observed to increase classification performance [35,63,70,75]. However, not all authors in the selected publications of this review reported the amount of data added during augmentation, which makes a comparison between these studies challenging. Furthermore, not only is the number of generated samples crucial but also the variety of features present in the added data. However, none of the publications in this review commented on this property.

While calculating errors between generated and original data is a straightforward method and serves as an indicator of signal similarity to some extent, the ability to accurately measure the authenticity of a generated signal is limited, thereby casting doubt on its efficacy as a reliable metric. However, synthetically generated signals must prove to be realistic yet different from the original data they stem from. While there are currently multiple different quality evaluation metrics used (Fig. 4), there still seems to be no consensus on the gold standard among them. While some of these metrics hold valuable information regarding the similarity between real and generated signals, their appropriate use is essential. For example, the PC as a linear correlation between real and generated data has shown an almost perfect correlation in all studies where it was applied [35,37,39]. However, the PC only takes into consideration the amplitudes over the entire signals but omits a possible time shift, because data is not compared with regards to time. Metrics comparing the difference between each data point of a time series (such as the RMSE) indicate how close a generated signal is to a real one and can be appropriate if the goal is to synthesize one lead based on another

[38,76]; and hence, the generated data should resemble the real data as close as possible. However, a phase shift of the generated signal results in a major difference even though the morphology of the signal could be close to the real-world data. A small distance metric would indicate that the two signals are similar in amplitude and overall shape, meaning that they are almost identical, which defies the purpose of artificially generating signals and results in copying the same signal instead. However, differences in shape and amplitude are essential to some extent if data is intended to be used to augment datasets and increase feature variability in the dataset. Therefore, if one's goal is to create synthetic data to augment and increase variability in a dataset, then such a metric might not be suitable at all. Further studies synthesizing data should choose their quality metrics carefully. Research should investigate in more detail the differences between these metrics and compare their advantages when evaluating time series data. Similar remarks have been made by Jeong et al. [81] in their review for GANs applied in image classification and segmentation. Furthermore, it could be beneficial to clinical application to evaluate the generated signals regarding their physiological properties, e.g., if R-R interval distribution is preserved for consecutive beats. This could enable the use of generated ECG signals for further applications.

However, a further aspect that has not been addressed in the reviewed publications is the general limitation of GANs to create only data of comparable size to the real data received by the discriminator. Since most publications segment their input data into single heart beats generated data are limited to this length. Only two publications [29,61] were able to create longer signals (up to 20 s). While this might not necessarily influence the training performance of classification algorithms, it can certainly be a problem if continuous time series of consecutive heartbeats are required. For example, any arrhythmia that is intermittent or occurs sporadically may require multiple beats to be detected, such as atrial fibrillation, supraventricular tachycardia, or ventricular tachycardia.

Overall, a trend of adapting the basic GAN architecture to fulfill the needs of time series generation was observed. In particular, using conditional instead of random input has proven to be a valuable upgrade in GAN architecture for ECG signal generation. Generating a specific class using a C-GAN or Sim-GAN seems to be promising, especially for data rebalancing.

Another important aspect in data generation using GANs is noisy input data, which can have negative impact on the quality and stability of data generation using GANs. The generated outputs may exhibit artifacts, inconsistencies, or poor fidelity to the desired data distribution. Furthermore, excessive noise may introduce instability, making it more challenging for the GAN to converge to an optimal solution. Proper preprocessing and noise reduction techniques can help mitigate these issues and improve training stability. Data preprocessing has been observed to be a controversial part of GAN training within the reviewed publications. Except for the segmentation of ECG time series into individual heart beats, there seems to be no other method commonly applied and none of the applied methods seems to provide any benefit for signal generation. If preprocessing is not feasible or does not result in the desired noise reduction, GAN training often includes regularization techniques to mitigate the adverse effects of noise. For instance, dropout or weight decay can help the GAN learn to generate meaningful data even in the presence of noise. These regularization methods encourage the model to focus on more robust and salient features, reducing the influence of noisy input. In general, a majority of the reviewed publications utilized such methods. While noisy input data is generally undesired, deliberate noise injection during training can be beneficial in some cases. Noise augmentation techniques, such as adding Gaussian noise or random perturbations to the input data, can improve the GAN's generalization capability and robustness by exposing it to a wider range of data variations.

Finally, the limitations of this review must be addressed. Even though GANs find applications in various fields, this review is limited to the specific application of ECG signal generation. Furthermore, the period for publications was limited to papers published after the establishment of GANs themselves in 2014. Seven pre-print publications [27,32,60,62,66,68,74] were included from arXiv.org, which have not been published in a peer-reviewed journal at the time of writing this review. However, this approach was unavoidable due to the immense pace at which the field is evolving.

9. Conclusion

The emergence of GANs has led to significant interest in both image and time series generation in recent years. This review summarizes the current applications of GANs in ECG signal generation to augment imbalanced datasets. A comprehensive overview of the used architectures, evaluation metrics, and classification performances is presented. The review shows that synthetic ECG signals can be effectively generated using GANs. Data augmentation based on synthetically generated ECG signals has proven to increase ECG classification accuracy and sensitivity significantly in the range of 6–7 %. However, single heartbeat segmentation as currently performed by many, poses the challenge of assessing intermittent arrhythmias or other pathologies which require consecutive heart beats. Furthermore, the absence of a gold standard for quality evaluation metrics of synthetically generated signals and the usage of diverse training parameters in classifier training impedes direct comparison between current approaches. Further research should focus on the establishment of a common quality evaluation metric for synthetically generated ECG signals of GANs.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.artmed.2023.102632.

CRediT authorship contribution statement

L.B., M.H., and F.M. conceived the study. L.B. conducted the review process; L.B. and M.H. drafted the manuscript; L.B., M.H., and F.M. edited the manuscript.

Declaration of competing interest

The authors do not have any financial or personal relationships that could be perceived as a potential conflict of interest.

Acknowledgements

This work was supported by the Austrian Research Promotion Agency (FFG) [grant number 883859]. We would like to express our sincere appreciation to Carina Butterworth from University of Calgary and SAIT for proofreading the manuscript.

References

[1] Malik M, John Camm A, Thomas Bigger J, Breithardt G, Cerutti S, Cohen RJ, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 1996;93:1043–65. https://doi.org/10.1161/01.CIR.93.5.1043.

[2] Brush JE, Brand DA, Acampora D, Chalmer B, Wackers FJ. Use of the initial electrocardiogram to predict in-hospital complications of acute myocardial infarction. N Engl J Med 1985;312:1137–41. https://doi.org/10.1056/nejm198505023121801.

[3] Celermajer DS, Chow CK, Marijon E, Anstey NM, Woo KS. Cardiovascular disease in the developing world: prevalences, patterns, and the potential of early disease detection. J Am Coll Cardiol 2012;60:1207–16. https://doi.org/10.1016/J.JACC.2012.03.074.

[4] Reeder B, David A. Health at hand: a systematic review of smart watch uses for health and wellness. J Biomed Inform 2016;63:269–76. https://doi.org/10.1016/J.JBI.2016.09.001.

[5] Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019;381:1909–17. https://doi.org/10.1056/NEJMOA1901183.

[6] Lih OS, Jahmunah V, San TR, Ciaccio EJ, Yamakawa T, Tanabe M, et al. Comprehensive electrocardiographic diagnosis based on deep learning. Artif Intell Med 2020:103. https://doi.org/10.1016/j.artmed.2019.101789.

[7] Celin S, Vasanth K. ECG signal classification using various machine learning techniques. J Med Syst 2018:42. https://doi.org/10.1007/s10916-018-1083-6.
[8] Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. Expert Syst Appl X 2020;7:100033. https://doi.org/10.1016/j.eswax.2020.100033.
[9] Somani S, Russak AJ, Richter F, Zhao S, Vaid A, Chaudhry F, et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. Europace 2021; 23:1179–91. https://doi.org/10.1093/EUROPACE/EUAA377.
[10] Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: a data perspective. ACM Comput Surv 2017:50. https://doi.org/10.1145/3136625.
[11] Kaur H, Pannu HS, Malhi AK. A systematic review on imbalanced data challenges in machine learning: applications and solutions. ACM Comput Surv 2019:52. https://doi.org/10.1145/3343440.
[12] Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA, Tourassi GD. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. Neural Netw 2008;21:427–36. https://doi.org/10.1016/J.NEUNET.2007.12.031.
[13] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019:6. https://doi.org/10.1186/s40537-019-0197-0.
[14] Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. J Med Imaging Radiat Oncol 2021;65:545–63. https://doi.org/10.1111/1754-9485.13261.
[15] Kim H, Phan TQ, Hong W, Chun SY. Physiology-based augmented deep neural network frameworks for ECG biometrics with short ECG pulses considering varying heart rates. Pattern Recogn Lett 2022;156:1–6. https://doi.org/10.1016/j.patrec.2022.02.014.
[16] Do E, Boynton J, Lee BS, Lustgarten D. Data augmentation for 12-Lead ECG beat classification. SN Comput Sci 2022;3:1–17. https://doi.org/10.1007/s42979-021-00924-x.
[17] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Adv Neural Inf Proces Syst 2014;27. https://doi.org/10.48550/arXiv.1406.2661.
[18] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM 2020;63:139–44. https://doi.org/10.1145/3422622.
[19] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Statement: an updated guideline for reporting systematic reviews. BMJ 2020;2021: 372. https://doi.org/10.1136/bmj.n71.
[20] Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B. High-resolution image synthesis and semantic manipulation with conditional GANs. Proc IEEE Comput Soc Conf Comput Vis Pattern Recogn 2018:8798–807. https://doi.org/10.1109/CVPR.2018.00917.
[21] Vondrick C, Pirsiavash H, Torralba A. Generating videos with scene dynamics. Adv Neural Inf Proces Syst 2016:613–21.
[22] Li J, Monroe W, Shi T, Jean S, Ritter A, Jurafsky D. Adversarial learning for neural dialogue generation. EMNLP 2017 - Conf Empir Methods Nat Lang Process Proc. 2017. p. 2157–69. https://doi.org/10.18653/V1/D17-1230.
[23] Kumar K, Kumar R, de Boissiere T, Gestin L, Teoh WZ, Sotelo J, et al. MelGAN: Generative adversarial networks for conditional waveform synthesis. In: Adv. Neural Inf. Process. Syst. 32. Neural Information Processing Systems FEoundation; 2019.
[24] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs. 34th Int Conf Mach Learn ICML 20176; 2016. p. 4043–55. https://doi.org/10.48550/arxiv.1610.09585.
[25] Mirza M, Osindero S. Conditional Generative Adversarial Nets. 2014. https://doi.org/10.48550/arxiv.1411.1784.
[26] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. In: 4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc. International Conference on Learning Representations, ICLR; 2016. https://doi.org/10.48550/arxiv.1511.06434.
[27] Golany T, Freedman D, Radinsky K. SimGANs: Simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification. In: 37th Int. Conf. Mach. Learn. ICML 2020. International Machine Learning Society (IMLS); 2020. p. 3555–64. https://doi.org/10.48550/arxiv.2006.15353. vol. PartF16814.
[28] Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. In: 6th Int Conf Learn Represent ICLR 2018 - Conf Track Proc; 2017. https://doi.org/10.48550/arxiv.1710.10196.
[29] Wang P, Hou B, Shao S, Yan R. ECG arrhythmias detection using auxiliary classifier generative adversarial network and residual network. IEEE Access 2019;7: 100910–22. https://doi.org/10.1109/ACCESS.2019.2930882.
[30] Thambawita V, Isaksen JL, Hicks SA, Ghouse J, Ahlberg G, Linneberg A, et al. DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. Sci Rep 2021:11. https://doi.org/10.1038/s41598-021-01295-2.
[31] Zhang YH, Babaeizadeh S. Synthesis of standard 12-lead electrocardiograms using two-dimensional generative adversarial networks. J Electrocardiol 2021;69:6–14. https://doi.org/10.1016/j.jelectrocard.2021.08.019.
[32] Wang G, Thite A, Talebi R, D'Achille A, Mussa A, Zutty J. Evolving SimGANs to improve abnormal electrocardiogram classification. In: GECCO 2022 Companion - Proc. 2022 Genet. Evol. Comput. Conf; 2022. p. 1887–94. https://doi.org/10.1145/3520304.3534048.
[33] Ye F, Zhu F, Fu Y, Shen B. ECG generation with sequence generative adversarial nets optimized by policy gradient. IEEE Access 2019;7:159369–78. https://doi.org/10.1109/ACCESS.2019.2950383.
[34] Nankani D, Baruah RD. Investigating deep convolution conditional GANs for electrocardiogram generation. In: Proc. Int. Jt. Conf. Neural Networks. Institute of Electrical and Electronics Engineers Inc.; 2020. https://doi.org/10.1109/IJCNN48605.2020.9207613.
[35] Yang H, Liu J, Zhang L, Li Y, Zhang H. ProEGAN-MS: a progressive growing generative adversarial networks for electrocardiogram generation. IEEE Access 2021;9:52089–100. https://doi.org/10.1109/ACCESS.2021.3069827.
[36] Zhu F, Ye F, Fu Y, Liu Q, Shen B. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. Sci Rep 2019:9. https://doi.org/10.1038/s41598-019-42516-z.
[37] Hazra D, Byun YC. Synsiggan: generative adversarial networks for synthetic biomedical signal generation. Biology (Basel) 2020;9:1–20. https://doi.org/10.3390/biology9120441.
[38] Zhang YH, Babaeizadeh S. Synthesis of standard 12-lead electrocardiograms using two dimensional generative adversarial network. J Electrocardiol 2021;69:6–14. https://doi.org/10.1016/j.jelectrocard.2021.08.019.
[39] Lee J, Oh K, Kim B, Yoo SK. Synthesis of electrocardiogram V-Lead signals from limb-lead measurement using R-peak aligned generative adversarial network. IEEE J Biomed Heal Informatics 2019;24:1265–75. https://doi.org/10.1109/JBHI.2019.2936583.
[40] Alt H, Godau M. Computing the Fréchet distance between two polygonal curves. Int J Comput Geom Appl 1995;05:75–91. https://doi.org/10.1142/S0218195995000064.
[41] Dowson DC, Landau BV. The Fréchet distance between multivariate normal distributions. J Multivar Anal 1982;12:450–5. https://doi.org/10.1016/0047-259X(82)90077-X.
[42] Müller M. Dynamic time warping. Inf Retr Music Motion 2007:69–84. https://doi.org/10.1007/978-3-540-74048-3_4.
[43] Cuturi M, Blondel M. Soft-DTW: A differentiable loss function for time-series. In: 34th Int Conf Mach Learn ICML 2017. 2; 2017. p. 1483–505.
[44] Kim MG, Pan SB. A study on user recognition using the generated synthetic electrocardiogram signal. Sensors 2021;21:1–13. https://doi.org/10.3390/s21051887.
[45] Vemishetty N, Gundlapalle V, Acharyya A, Chakravarti B. A novel 2-lead to 12-lead ECG reconstruction methodology for remote health monitoring applications. In: 2019 Comput Cardiol Conf. 45; 2019. https://doi.org/10.22489/CINC.2019.231.
[46] Shmelkov K, Schmid C, Alahari K. How good is my GAN? Lect Notes Comput Sci 2018;11206:218–34. https://doi.org/10.48550/arxiv.1807.09499. LNCS.
[47] Apaydin T, Ferhatosmanoglu H. Access structures for angular similarity queries. IEEE Trans Knowl Data Eng 2006;18:1512–25. https://doi.org/10.1109/TKDE.2006.165.
[48] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag 2001;20:45–50. https://doi.org/10.1109/51.932724.
[49] Moody GB. MIT-BIH Normal Sinus Rhythm Database. 1999. https://doi.org/10.13026/C2NK5R.
[50] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomed Tech Biomed Eng 1995;40:317–8. https://doi.org/10.1515/bmte.1995.40.s1.317.
[51] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL, a large publicly available electrocardiography dataset. Sci Data 2020;71(7): 1–15. https://doi.org/10.1038/s41597-020-0495-6.
[52] Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Sci Data 2020;71(7):1–8. https://doi.org/10.1038/s41597-020-0386-x.
[53] Clifford GD, Liu C, Moody B, Lehman LH, Silva I, Li Q, et al. AF classification from a short single lead ECG recording: the PhysioNet/computing in cardiology challenge 2017. Comput Cardiol 2010;2017(44):1–4. https://doi.org/10.22489/CINC.2017.065-469.
[54] Khamis H, Weiss R, Xie Y, Chang C-W, Lovell NH, Redmond SJA-ARC. TELE ECG Database: 250 telehealth ECG records (collected using dry metal electrodes) with annotated QRS and artifact masks, and MATLAB code for the UNSW artifact detection and UNSW QRS detection algorithms n.d. doi:10.7910/DVN/QTG0EP.
[55] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. J Med Imaging Heal Informatics 2018;8:1368–73. https://doi.org/10.1166/JMIHI.2018.2442.
[56] Juhl CR, Miller IM, Jemec GB, Kanters JK, Ellervik C. Hidradenitis suppurativa and electrocardiographic changes: a cross-sectional population study. Br J Dermatol 2018;178:222–8. https://doi.org/10.1111/BJD.15778.
[57] Ghouse J, Have CT, Weeke P, Bille Nielsen J, Ahlberg G, Balslev-Harder M, et al. Rare genetic variants previously associated with congenital forms of long QT syndrome have little or no effect on the QT interval. Eur Heart J 2015;36:2523–9. https://doi.org/10.1093/EURHEARTJ/EHV297.
[58] Jia-Wei Z, Li-Ping W, Xia L, Hong-Hai Z, Jun D. Chinese Cardiovascular Disease Database (CCDD) and its management tool. In: 10th IEEE Int Conf Bioinforma Bioeng 2010. BIBE 2010; 2010. p. 66–72. https://doi.org/10.1109/BIBE.2010.19.
[59] Willems JL, Arnaud P, Van Bemmel JH, Degani R, MacFarlane PW, Zywietz C. Common standards for quantitative electrocardiography: goals and main results. Methods Inf Med 1990;29:263–71. https://doi.org/10.1055/S-0038-1634793/ID/JR1634793-25.
[60] Golany T, Lavee G, Yarden ST, Radinsky K. Improving ECG Classification Using Generative Adversarial Networks. 2020. p. 13280–5.
[61] Wulan N, Wang W, Sun P, Wang K, Xia Y, Zhang H. Generating electrocardiogram signals by deep learning. Neurocomputing 2020;404:122–36. https://doi.org/10.1016/j.neucom.2020.04.076.

[62] Golany T, Freedman D, Radinsky K. ECG ODE-GAN: learning ordinary differential equations of ECG dynamics via generative adversarial learning. In: 35th AAAI Conf. Artif. Intell. AAAI 2021. 1; 2021. p. 134–41.
[63] Shaker AMM, Tantawi M, Shedeed HAA, Tolba MFF. Generalization of convolutional neural networks for ECG classification using generative adversarial networks. IEEE Access 2020;8:35592–605. https://doi.org/10.1109/ACCESS.2020.2974712.
[64] He Y, Fu B, Yu J, Li R, Jiang R. Efficient learning of healthcare data from IoT devices by edge convolution neural networks. Appl Sci 2020;10:1–19. https://doi.org/10.3390/app10248934.
[65] Lan T, Hu Q, Liu X, He K, Yang C. Arrhythmias classification using short-time Fourier transform and GAN based data augmentation. In: Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS. 2020. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 308–11. https://doi.org/10.1109/EMBC44109.2020.9176733. July.
[66] Dasgupta S, Das S, Bhattacharya U. Cardiogan: An attention-based generative adversarial network for generation of electrocardiograms. In: Proc. - Int. Conf. Pattern Recognit. Institute of Electrical and Electronics Engineers Inc.; 2021. p. 3193–200. https://doi.org/10.1109/ICPR48806.2021.9412905.
[67] Brophy E, De Vos M, Boylan G, Ward T, De Vos M, Boylan G, et al. Multivariate generative adversarial networks and their loss functions for synthesis of multichannel ECGs. IEEE Access 2021;9:158936–45. https://doi.org/10.1109/ACCESS.2021.3130421.
[68] Hossain KFF, Kamran SAA, Tavakkoli A, Pan L, Ma X, Rajasegarar S, et al. ECG-Adv-GAN: detecting ECG adversarial examples with conditional generative adversarial networks. In: 2021 20th IEEE Int. Conf. Mach. Learn. Appl; 2021. p. 50–6. https://doi.org/10.1109/ICMLA52953.2021.00016.
[69] Rath A, Mishra D, Panda G, Satapathy SC. Heart disease detection using deep learning methods from imbalanced ECG samples. Biomed Signal Process Control 2021:68. https://doi.org/10.1016/j.bspc.2021.102820.
[70] Zhou X, Zhu X, Nakamura K, Noro M. Electrocardiogram quality assessment with a generalized deep learning model assisted by conditional generative adversarial networks. Life 2021:11. https://doi.org/10.3390/life11101013.
[71] Zhou Z, Zhai X, Tin C. Fully automatic electrocardiogram classification system based on generative adversarial network with auxiliary classifier. Expert Syst Appl 2021:174. https://doi.org/10.1016/j.eswa.2021.114809.
[72] Ma S, Cui J, Xiao W, Liu L. Deep learning-based data augmentation and model fusion for automatic arrhythmia identification and classification algorithms. Comput Intell Neurosci 2022;2022. https://doi.org/10.1155/2022/1577778.
[73] Islam MS, Islam MN, Hashim N, Rashid M, Bari BS, Al Farid F. New hybrid deep learning approach using BiGRU-BiLSTM and multilayered dilated CNN to detect arrhythmia. IEEE Access 2022;10:58081–96. https://doi.org/10.1109/ACCESS.2022.3178710.
[74] Delaney AM, Brophy E, Ward TE. Synthesis of realistic ECG using generative adversarial networks. 2019. https://doi.org/10.48550/arxiv.1909.09150.
[75] Li W, Tang YM, Yu KM, To S. SLC-GAN: an automated myocardial infarction detection model based on generative adversarial networks and convolutional neural networks with single-lead electrocardiogram synthesis. Inf Sci (NY) 2022;589:738–50. https://doi.org/10.1016/j.ins.2021.12.083.
[76] Seo HC, Yoon GW, Joo S, Nam GB. Multiple electrocardiogram generator with single-lead electrocardiogram. Comput Methods Prog Biomed 2022:221. https://doi.org/10.1016/j.cmpb.2022.106858.
[77] Hatamian FN, Ravikumar N, Vesal S, Kemeth FP, Struck M, Maier A. The effect of data augmentation on classification of atrial fibrillation in short single-lead ECG signals using deep neural networks. In: ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc. 2020. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 1264–8. https://doi.org/10.1109/ICASSP40776.2020.9053800. May.
[78] Islam M, Tayan O, Islam R, Islam S, Nooruddin S, Kabir MN, et al. Deep learning based systems developed for fall detection: a review. IEEE Access 2020;8:166117–37. https://doi.org/10.1109/ACCESS.2020.3021943.
[79] Kachuee M, Fazeli S, Sarrafzadeh M. ECG heartbeat classification: a deep transferable representation. In: Proc - 2018 IEEE Int Conf Healthc Informatics. ICHI 2018; 2018. p. 443–4. https://doi.org/10.1109/ICHI.2018.00092.
[80] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A, et al. A deep convolutional neural network model to classify heartbeats. Comput Biol Med 2017;89:389–96. https://doi.org/10.1016/J.COMPBIOMED.2017.08.022.
[81] Jeong JJ, Tariq A, Adejumo T, Trivedi H, Gichoya JW, Banerjee I. Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation. J Digit Imaging 2022;35:137–52. https://doi.org/10.1007/s10278-021-00556-w.