scientific reports

Check for updates

OPEN

A performance analysis of convolutional autoencoder modified WaveGAN architectures for realistic 12 lead electrocardiogram synthesis

Jiaqi Liu[1], Kwok Tai Chui[1]✉, Lap-Kei Lee[1], Kenneth Lo[2,3], Aimin Yang[4] & Edmond King Sing Fong[1]

The burgeoning necessity for copious and diverse electrocardiogram (ECG) datasets for deep learning applications in clinical diagnostics has been impeded by the confidential nature of patient data. Related works have shown the effectiveness of additional data generation in enhancing the deep learning models' performance. This research study introduces a novel Convolutional Autoencoder-WaveGAN (CAE-WaveGAN) technique for generating synthetic but realistic 12-lead ECG images to address data scarcity. The proposed model leverages a convolutional autoencoder for efficient feature extraction from ECG signals, which is then utilized by a WaveGAN generator to synthesize high-fidelity ECG images. The method provides a practical solution for expanding ECG training datasets where patient privacy constraints and data scarcity limit the development of robust deep learning models for cardiovascular diagnosis. We conducted a comprehensive performance analysis of various CAE-WaveGAN configurations through an ablation study on the CODE-15% dataset. Experimental results demonstrate that CAE-WaveGAN achieves superior performance across all evaluation metrics, with 19.8% improvement in PSNR and 59.3% enhancement in SSIM compared to baseline methods. Our findings from the ablation study reveal that the optimal CAE-WaveGAN architecture significantly surpasses the traditional WaveGAN in terms of stability and loss metrics, offering a promising solution for generating realistic ECG data in clinical machine learning applications.

Keywords 12-Lead ECG synthesis, Convolutional autoencoder, Deep learning, ECG data generation, Feature extraction, Generative adversarial networks

An electrocardiogram (ECG) is indispensable in the treatment of heart diseases (also known as cardiovascular diseases) and helps diagnose and differentiate among various heart conditions[1–3]. Amidst the ever-advancing fields of artificial intelligence, using deep learning algorithms to classify ECGs is becoming increasingly common in clinical medicine[4,5]. Deep learning methods typically require a large amount of data for training[6,7]; however, ECG data, being sensitive patient information, is usually not disclosed to the public. This leads to a scarcity of ECG data for deep learning[8]. Since deep learning models require large and diverse datasets to achieve robust performance, the lack of accessible ECG data presents a significant barrier to their development and application. To address this challenge, data augmentation and synthetic data generation have become essential strategies for expanding available training datasets. Among various synthetic data generation techniques, generative adversarial networks (GANs) have attracted significant attention due to their remarkable ability to learn complex data distributions and generate highly realistic samples. As a result, GANs have emerged as the most popular and effective approach for generating synthetic ECG data to compensate for this scarcity. GANs consist of a generator and a discriminator. The discriminator aims to accurately classify real and fake images,

[1]School of Science and Technology, Hong Kong Metropolitan University, Ho Man Tin, Kowloon, Hong Kong SAR, China. [2]Department of Food Science and Nutrition, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China. [3]Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China. [4]Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China. ✉email: jktchui@hkmu.edu.hk

while the generator strives to produce realistic images that the discriminator cannot accurately classify. Through iterative training of both modules, their performance alternately improves due to the competitive nature of the game[9]. Notably, GANs provide additional and necessary constraints for training the generator, leading to significant improvements in its performance. Furthermore, GAN training mechanisms do not impose specific requirements on the generator's specific structure[10].

Over the past decade, deep learning algorithms have advanced rapidly, with a notable increase in the generation of ECG images[9,11]. While many algorithms have been developed to produce functional ECG images, many are limited to creating images for single-lead ECGs[12–14]. Regarding 12-lead ECG images, different leads exhibit distinct features and correlations[15]. However, current ECG synthesis approaches face several critical limitations that restrict their clinical applicability. First, traditional GAN-based methods, such as DCGAN and standard WGAN, suffer from training instability issues, particularly when applied to complex multi-lead ECG data, resulting in mode collapse and poor convergence. Second, existing approaches lack sophisticated feature extraction mechanisms that can capture the intricate morphological relationships between different ECG leads, leading to generated signals that may preserve individual lead characteristics but fail to maintain physiologically accurate inter-lead correlations. Third, most current methods, including the original WaveGAN, do not adequately address the unique temporal dependencies and periodic patterns inherent in ECG signals, resulting in synthetic data that lacks clinical realism for diagnostic applications.

In addition, a GAN may struggle to replicate each lead perfectly. Without feature extraction, the GAN-generated ECG image could lack the specific features of the leads and the correlations between them, resulting in a significant discrepancy compared to a real sample and potentially introducing unrealistically poor features. Therefore, designing GAN-derived models specifically for the 12 leads is crucial to ensure accurate extraction of features and associations during training. This approach is instrumental in producing high-quality and realistic ECG images.

The presented study proposes a CAE-WaveGAN technique for generating high-fidelity 12-lead ECG signals using model architecture specifically designed for time-series data, leveraging a Convolutional Autoencoder (CAE) for feature extraction combined with a generative model, these generated signals are subsequently converted into ECG images, which are commonly used in clinical settings and for image-based quality assessment. The proposed CAE-WaveGAN model is specifically designed to process one-dimensional ECG signal data, not images. The WaveGAN architecture is inherently tailored for generating raw time-series signals, making it particularly well-suited for ECG data synthesis in its native temporal format. Throughout the training process, our model operates exclusively on raw ECG signal sequences to learn the underlying temporal patterns and physiological characteristics. The conversion from generated ECG signals to 12-lead ECG images occurs post-generation and serves two primary purposes: visualization and comprehensive evaluation. This transformation enables clinicians to assess the diagnostic quality and morphological fidelity of synthetic signals using familiar clinical presentation formats. Additionally, image-based evaluation facilitates the application of established computer vision metrics and enables seamless integration with existing clinical workflows, where ECG images are a standard practice. CAE is an unsupervised learning algorithm within the autoencoder framework, offering the advantage of efficiently capturing spatial and temporal features in data through convolutional operations. WaveGAN's one-dimensional convolution operations are well-suited for processing the periodicity and frequency content of ECG signals, making it effective for ECG signal generation. This ability positions it as a promising starting point from which inspiration can be drawn and adapted for the task of ECG signal generation. Its ability to produce high-quality raw waveform data, combined with techniques like phase shuffling for generating diverse data, ensures robust performance in synthesizing realistic ECG signals.

An autoencoder is designed to encode input data into a lower-dimensional representation and reconstruct the output data from this representation. The reconstruction is then compared to the original input data to minimise their difference. This process of comparison and optimization is repeated across numerous iterations[16]. A large amount of unlabelled data is used to train a feature learning network to extract features from signals via CAE. The proposed method addresses the intricacies of the 12-lead ECG, encompassing six limb leads (I, II, III, aVR, aVL, aVF) and six precordial leads (V1–V6) and their inter-lead dependencies. An ablation study was conducted using the CODE-15% dataset to benchmark the CAE-WaveGAN against traditional WaveGAN approaches, assessing the quality of the generated ECG signals and the stability of the training process. It was observed that the proposed generator synthesizes ECG signals with high visual realism, is suitable for clinical diagnostics and monitoring and exhibits enhanced stability during training.

These limitations collectively create a significant gap in the field. While various generative models have shown promise for ECG synthesis, no one adequately combines stable training processes with sophisticated feature extraction capabilities specifically designed for multi-lead ECG generation. The absence of such a framework limits the practical utility of synthetic ECG data for augmenting clinical datasets and training robust diagnostic algorithms.

To sum up, the research contributions of this paper are as follows:

- A CAE-WaveGAN-based method is proposed to generate 12-lead ECG images, and a comparative study of various CAE-WaveGAN configurations has been conducted to identify an optimal structure.
- Employing a CAE as the feature extractor within the generator was refined through parameter optimization and design adjustments. Compared to alternative configurations, it demonstrated superior stability during training and enhanced accuracy in generation.
- Extensive testing on the CODE-15% dataset revealed that our model's generator significantly outperforms the original WaveGAN generator regarding loss metrics during both the training and validation phases.
- The rest of the paper is structured as follows. Section 2 presents the related works in ECG synthesis and denoising using GANs, providing context for research contributions. Section 3 details the proposed meth-

odology, including the WaveGAN architecture, the convolutional autoencoder design, and the proposed CAE-WaveGAN framework. Section 4 describes our experimental setup, evaluation procedures, tools, datasets, and data preprocessing steps. Section 5 presents our results and discussions, including ablation studies on hyperparameters, comparative analysis with original WaveGAN, WGAN, ECGGAN, DCGAN, and the impact of different CAE layer configurations and skip connections. Finally, Sect. 6 concludes the paper with key findings, study limitations, and future research directions.

Related work

Recent works have applied GANs to ECG synthesis and denoising, demonstrating improved performance in data-limited scenarios. These studies are reviewed in the ensuing subsection.

Synthesizing and denoising with GANs

Antczak proposed a deep 2D convolutional network trained on real ECG data to classify signals based on diagnostic groups[17]. It also evaluated the denoising autoencoders. This allowed the calculation of the data's "realism," closely related to human perception. High-quality ECG data were generated using a Wasserstein GAN and processed through a parametric noise model. Finally, a convolutional denoising autoencoder was trained to reconstruct the noise-free initial signal from noisy inputs. This approach significantly improved the performance of models trained on synthetic data compared to those trained solely on real data. It also examined the potential advantages of transfer learning by reusing a trained discriminator network in the denoising models. Although this method achieved good performance with both synthetic and real data, the detailed effects of transfer learning strategies on model performance were not thoroughly analysed, indicating a potential area for further research to optimize model performance and enhance its applicability in medical diagnosis.

Shaik and Bhavanam contributed to ongoing research in arrhythmia detection by introducing a Prioritized Feature Subset Vector-Associated Generative Adversarial Network (PFSV-AGAN)[18]. Their approach aimed at resolving the issue of inconsistent data sizes in ECG signal datasets and the inherent problem of insufficient labeled data for robust classifier training. The proposed model leveraged the power of GANs to synthesize realistic ECG signals for multiple classes, focusing on generating class-specific adversarial examples that could enhance the training process of deep learning models for arrhythmia detection. While the authors asserted improved classification accuracy, the research primarily concentrated on generating synthetic ECG data. However, it did not address the challenge of noise reduction in existing ECG signals, which is critical in practical and clinical scenarios. One notable limitation of this study is that the generation of synthetic data lacks a thorough exploration of the impact of synthetic noise on the classification models. Although the proposed PFSV-AGAN framework aims to enrich the dataset for improved model training, comprehensive analysis regarding the real-world application and validation of the synthesized ECG signals is absent. The study's comparison with traditional models lacks a detailed discussion on the convergence properties of the network involved.

Synthetic generation of ECG data using GANs

In synthetic ECG data generation, Kuntalp and Düzyel introduced a novel GAN-based approach[19]. Utilizing ECG signals from the MIT-BIH arrhythmia dataset, they employed the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique to analyse class-specific cluster structures. This method enhanced classification performance in scenarios where specific classes exhibited distinct clusters. A range of classifiers, including 1-D Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), One-vs-One Classifiers (Ovo), K-Nearest Neighbors (KNN), and Random Forests (RF), were employed in their experiments, highlighting the potential advantages of their proposed data augmentation technique. Nevertheless, while Kuntalp and Düzyel's method showed promise under certain conditions, it possessed limitations. The generalizability of their approach across varied datasets and signal types had yet to be established, suggesting a need for further empirical evaluation. In addition, their work considered data enhancement, and the feature extraction capabilities and applicability of their approach to other biomedical signals remain to be explored. Despite these limitations, their work offered a fresh perspective on enhancing GAN-based data augmentation methods and set the stage for subsequent research in this domain.

In an in-depth look at synthetic ECG signals, Bagga, Jeon, and Issokson proposed ECGNet[20], a GAN-based approach capable of synthesizing complete 12-lead ECG signals from single-lead inputs. Using a combination of a bi-directional Long Short-Term Memory (LSTM) generator and a CNN discriminator, the limitations of traditional GAN models were overcome when applied to cardiovascular disease (CVD) classification models. Through cross-correlation and autocorrelation analyses, the generated 12-lead ECG signals retained the features of CVD. Despite ECGNet's breakthrough in synthesizing ECG signals and achieving advanced accuracy in Fréchet distance scores, the study experienced key limitations. In particular, the study failed to detail how the model handled any unrealistic features presented in the generated ECG signals, which may affect its effectiveness in disease diagnosis.

Noise reduction in ECG signals through GANs

In ECG signal processing, an adversarial denoising convolutional neural network (ADnCNN)-based method for denoising ECG recordings, contaminated by noise in practical scenarios, was introduced by Hou et al.[21]. The authors developed an architecture that integrated a denoising convolutional neural network (DnCNN) with a GAN, enabling the estimation of residual signals within noisy data through an end-to-end learning framework. This approach facilitated effective noise reduction. The ADnCNN was empirically tested on diverse datasets such as the MIT-BIH Arrhythmia Database, the MIT-BIH Noise Stress Test Database, and the QT Database, with evaluation metrics including signal-to-noise ratio (SNR), root mean square error (RMSE), and percentage root mean square difference (PRD). Despite outperforming several existing models in denoising efficacy, the

ADnCNN's computational demand and slower convergence highlight areas for improvement, particularly in enhancing model interpretability.

Recently, Wang et al. proposed a deep convolutional generative adversarial network (dcGAN) combined with LSTM for denoising ECG signals subjected to multiple noise disturbances[22]. The method was tested on the MIT-BIH arrhythmia database. It showed excellent performance in removing single and mixed noise, obtaining average signal-to-noise ratio improvement (SNRimp), RMSE, and PRD metrics of 19.254 dB, 0.028, and 10.350, respectively. It outperformed existing dcGAN and LSTM-GAN methods. However, the performance of the model in processing the data was not fully discussed in the paper. In addition, the study did not fully consider the computational complexity of the model and the requirement of hardware resources.

Geometrical transformation and augmentation techniques for ECG

Traditional approaches to addressing ECG data scarcity have extensively utilized geometrical transformations and signal augmentation techniques as foundational methods for dataset expansion. Eriksson et al.[23] demonstrated the application of adversarial training methods that involve creating worst-case disturbed artificial samples through geometric perturbations, showing that such approaches can improve model transferability across different ECG datasets when applied to atrial fibrillation classification tasks. This study revealed that adversarially trained models achieved superior transferability performance (0.732 vs. 0.685 AUPRC) compared to conventionally trained models when fine-tuned on new datasets. Similarly, Wang et al.[24] explored various ECG transformation methods including amplitude scaling, baseline wander simulation, and controlled noise injection strategies to artificially expand ECG datasets for deep learning applications. These conventional augmentation techniques typically encompass time-domain manipulations such as scaling, shifting, stretching of ECG signals, and frequency-domain transformations and systematic noise addition protocols.

However, conventional augmentation techniques face several critical limitations that restrict their clinical applicability and effectiveness in generating realistic synthetic ECG data. Wang et al.[24] identified that existing augmentation methods are often not customized based on ECG noise characteristics and may inadvertently alter the morphological properties of signals or disrupt temporal sequences, which proves counterproductive for improving noise robustness in clinical applications. Furthermore, these methods are fundamentally constrained by the characteristics of the original signals and cannot generate truly novel ECG patterns representing unseen cardiac conditions or rare arrhythmias. Eriksson et al.[23] noted that while adversarial training through geometric transformations can improve model robustness, such approaches may inadvertently alter clinically significant features such as precise QT intervals, P-wave morphology, and ST-segment characteristics, which are crucial for accurate cardiac diagnosis. Additionally, excessive augmentation through geometric transformations can introduce unrealistic artifacts that do not reflect genuine physiological variations, potentially misleading machine learning models during training and reducing their clinical validity. The computational overhead associated with iterative adversarial training processes also presents practical challenges for large-scale clinical deployment, highlighting the need for more efficient generative approaches that can produce high-quality synthetic ECG data without compromising diagnostic accuracy.

Based on the comprehensive review of existing literature, several critical research gaps emerge in the field of synthetic ECG generation. While traditional geometrical transformation and augmentation techniques provide basic data expansion capabilities, they are fundamentally limited by their inability to generate truly novel cardiac patterns and their tendency to introduce unrealistic artifacts that may compromise clinical validity. Current GAN-based approaches, particularly WaveGAN, have demonstrated promise in generating synthetic ECG signals; however, they suffer from training instability and lack sophisticated feature extraction mechanisms that could enhance the quality and clinical relevance of the generated data. Moreover, most existing studies focus primarily on single-lead ECG synthesis, with limited exploration of comprehensive 12-lead ECG generation that is essential for clinical diagnostic applications. To address these limitations, our study introduces the CAE-WaveGAN framework, which combines the robust feature extraction capabilities of convolutional autoencoders with the generative power of WaveGAN. This approach aims to bridge the gap between traditional augmentation methods and advanced generative models, providing a more stable and clinically relevant solution for synthetic 12-lead ECG generation while preserving the complex morphological characteristics essential for accurate medical diagnosis.

Methodology

In our proposed method, the hybrid generator comprises a convolutional autoencoder[25], a WaveGAN generator, and a WaveGAN discriminator[26]. Traditional WaveGAN has been partially optimized and adapted for ECG data. Regarding our research study, the preprocessing steps for ECG signals (normalization) and the postprocessing steps (filtering and trimming) have been adjusted based on the characteristics of ECG data. Adaptations are required in the generator network, particularly in the final layer of output, to ensure that the length of the output signal matches that of the ECG data. To construct the generative model, a CAE is added as the foundational architecture for the generator component, while utilizing a CAE is also utilized for feature extraction. Notably, in our CAE design, pooling layers are omitted. This decision was informed by the fact that pooling layers can lead to a loss of detailed information, which is critical to preserving the medical significance of ECG signals. Instead, skip connections are used to facilitate the direct flow of gradients to the initial layers of the network, thereby aiding in the preservation of important signal details. This design enables the hybrid generator to maintain the spatial locality and neighbor relations of ECG signals within high-level feature representations via the CAE. WaveGAN, traditionally applied to original audio waveform generation, is adept at learning and synthesizing audio waveforms from numerous real examples. Its application to time series generation has demonstrated comparable efficacy. In addition, skip connections are incorporated to facilitate gradient flow during the training

process, enhancing the denoising performance. The role of skip connections and their underlying studies will be elaborated in subsequent sections. A more detailed discussion of the CAE-WaveGAN components will be presented.

WaveGAN

WaveGAN, originally designed for generating raw audio waveforms, demonstrates strong potential for synthesizing ECG signals due to its one-dimensional convolutional architecture, which is well-suited for processing time-series data. Unlike traditional DCGANs[27], WaveGAN employs 1D convolutions, extended filter lengths (25 units), and a quadrupled upsampling factor in the generator, with corresponding adaptations in the discriminator, such as elongated filters and a 4-step stride. These modifications enable the model to generate longer sequences, an essential feature for biomedical applications like ECG synthesis. Notably, WaveGAN omits batch normalization, diverging from standard DCGAN configurations, which enhances its performance in generating high-fidelity waveforms.

Evaluation of WaveGAN has shown that it excels in producing clear, realistic audio samples[28], indicating its suitability for waveform and spectral data generation[26]. In this research, the application of the WaveGAN architecture is extended to synthesis ECG signals. Given that ECG data are inherently time-series signals, the one-dimensional convolutional approach of WaveGAN is well-suited to this task. This represents a novel challenge and an innovative step forward in assessing the model's capacity for one-dimensional convolutional processing in the biomedical domain.

Convolutional autoencoder

An autoencoder is a self-supervised learning algorithm that employs a neural network architecture to facilitate the feature learning process[29]. This compression is achieved through a bottleneck layer, which forces the network to capture the essential features of the data. The structure of a typical autoencoder with N layers is depicted in Fig. 1. In this figure, the number of input channels for each layer has been annotated in the Encoder Network, and the number of output channels for each layer in the Decoder Network, to provide a clearer understanding of the model structure.

The autoencoder architecture comprises four key components: the Encoder Network, the Bottleneck Layer, the Decoder Network, and the Reconstruction Loss Function. The Encoder Network compresses input data, with the Bottleneck Layer (its final layer) outputting the encoded representation. In an N-layer Encoder Network, the Nth layer serves as the Bottleneck Layer. Each layer in the Encoder Network operates to progressively transform the input into the compressed domain. The operation of each layer within the Encoder Network is defined as follows.

$$X_{e_{k+1}} = f_{e_k}\left(W_{e_k}^T X_{e_k} + b_{e_k}\right) \forall k = 0, 1, 2, \ldots, N \tag{1}$$

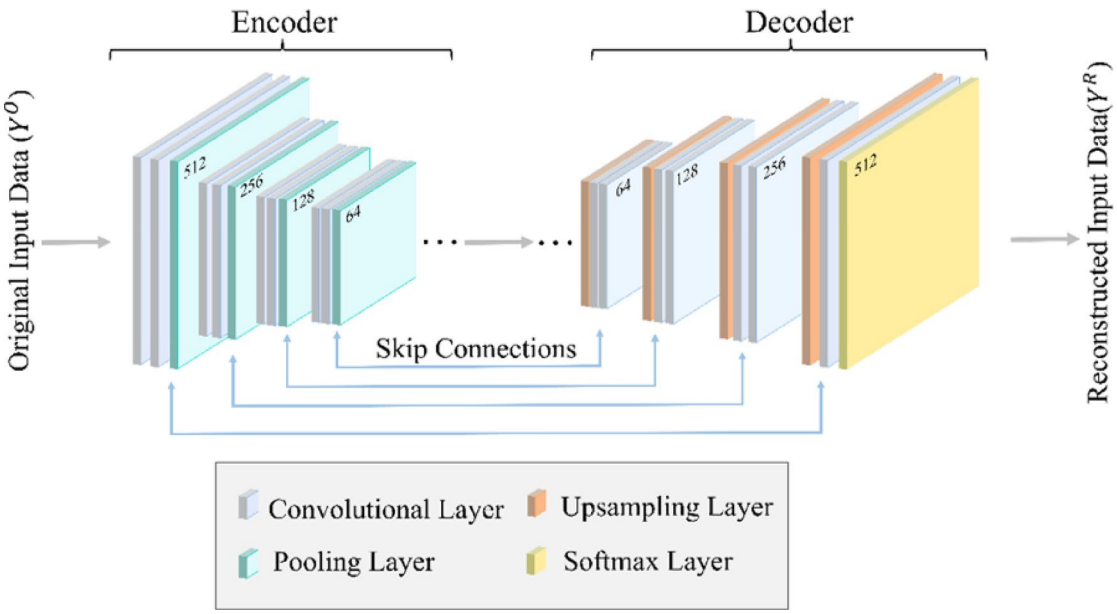

Fig. 1. Architecture of N-layer Autoencoder. The architecture integrates skip connections between the encoder and decoder to enhance feature representation and reconstruction fidelity. In addition, the number of input channels for each layer in the Encoder Network and the number of output channels for each layer in the Decoder Network are explicitly indicated.

where X_{e_k} is the input for the k^{th} layer of Encoder Network, $X_{e_{k+1}}$ is the output of the k^{th} layer of the Encoder Network, W_{e_k} is the weight vector for the k^{th} layer of the Encoder Network, b_{e_k} is the bias for the k^{th} layer of the Encoder Network, and f_{e_k} is the activation function for the k^{th} layer of the Encoder Network.

The Decoder Network, mirroring the structure of the Encoder Network in reverse, reconstructs the original data from the encoded representation in the Bottleneck Layer. Its final output approximates the original input, with some noise introduced during reconstruction. The functioning of each layer within the Decoder Network is presented as follows.

$$X_{dk+1} = f_{d_k}\left(W_{d_k}^T X_{d_k} + b_{d_k}\right) \forall k = 0, 1, 2, \cdots, N \tag{2}$$

where X_{d_k} is the input for the k^{th} layer of Decoder Network, $X_{d_{k+1}}$ is the output of the k^{th} layer of the Decoder Network, W_{d_k} is the weight vector for the k^{th} layer of the Decoder Network, b_{d_k} is the bias for the k^{th} layer of the Decoder Network, and f_{d_k} is the activation function for the k^{th} is the activation function for the k^{th} layer of the Decoder Network. The difference between the original data Y^{O} and the reconstructed data Y^{R} is known as Reconstruction Loss.

The autoencoder is trained using the Backpropagation Algorithm to minimize Reconstruction Loss, typically measured by the Mean Squared Error (MSE) or Binary Cross Entropy (BCE) loss functions, as defined in Eqs. (3) and (4), respectively.

$$MSE\left(Y^O,\ Y^R\right) = \frac{1}{D}\sum_{k=1}^{D}\left(Y_k^O - Y_k^R\right)^2 \tag{3}$$

$$BCE\left(Y^O, Y^R\right) = -\frac{1}{D}\sum\nolimits_{k=1}^{D}\left[Y_k^O \log\left(Y_k^R\right) + \left(1 - Y_k^O\right)\log\left(1 - Y_k^R\right)\right] \tag{4}$$

For Convolutional Autoencoders (CAE), convolutional and subsampling layers encode input images into compressed representations, while upsampling and convolutional layers reconstruct the original images. In this study, which focuses on ECG signals, the proposed hybrid model ensures that critical features of the ECG signals are preserved and not lost. These features are subsequently utilized to train the generator within the CAE-WaveGAN framework.

CAE-WaveGAN

In this study, a novel hybrid model has been designed for the autonomous generation of 12-lead ECG data. To the best of our knowledge, no existing work has proposed producing 12-lead ECG data based on a combination of CAE and WaveGAN technologies. The model harnesses two advanced deep learning techniques: CAE and WaveGAN.

Initially, a CAE network is trained to reduce the dimensionality of the input ECG data. This dimensionality reduction is meticulously performed to ensure that vital features of the ECG images are not lost—secured by applying an upper bound on the CAE reconstruction loss. Upon dimensionality reduction, the output from the CAE encoder network is fed into the WaveGAN generator. With the assistance of WaveGAN, the input ECG images are learned and used to generate ECG data.

Overall, our proposed CAE-WaveGAN model preserves critical information about the ECG signals through the encoder and decoder of the CAE and then employs WaveGAN to generate synthetic 12-lead ECG data. The impact of varying the number of layers in the CAE on the proposed model under the same dataset training conditions is also discussed, including model performance, average training time per epoch, and overall training stability. The training process is illustrated in Fig. 2.

In this research, a series of CAE-WaveGAN models with varying numbers of layers is established. Experiments on the ImageNet-R dataset and observed that the reconstruction accuracy improved with increased convolutional layers in the CAE, starting from three layers[30]. The accuracy peaked when the CAE comprised four layers, with the performance remaining relatively stable for $4 \leq N\_CAE \leq 6$. Considering that ECG data contains more noise and complex perturbations than ImageNet-R data and possesses distinct features and intrinsic structures, this work builds upon previous research by extending the range of CAE layers. Therefore, the number of CAE layers is set to $2 \leq N\_CAE \leq 8$ to investigate the stability of different CAE layer counts in generating ECG signals. For this reason, models composed of CAE with encoder and decoder structures ranging from 2 to 8 layers are configured, wherein the number of layers in the encoder and decoder is typically identical. Throughout the subsequent sections of this paper, a model with a particular number of layers is referred as CAE (x layers)-WaveGAN, where 'x' denotes the number of layers in either the encoder or decoder, not the total number of layers in the CAE. It is important to clarify that since the CAE includes both an encoder and a decoder, the total number of layers in the corresponding CAE would be twice the value of 'x'. During the training process, the impact of skip connections on our model is further investigated. The architecture of the generator in our proposed model, the skip connection methodology, and the ECG data generation process are depicted in Fig. 3.

Experiments and evaluation

Procedure

Performance evaluation aims to evaluate the performance of our proposed CAE-WaveGAN and investigate the impact of its architectural components. Specifically, the influence of different CAE layers on model performance is assessed and the role of jump connections within the CAE-WaveGAN framework is explored. While each

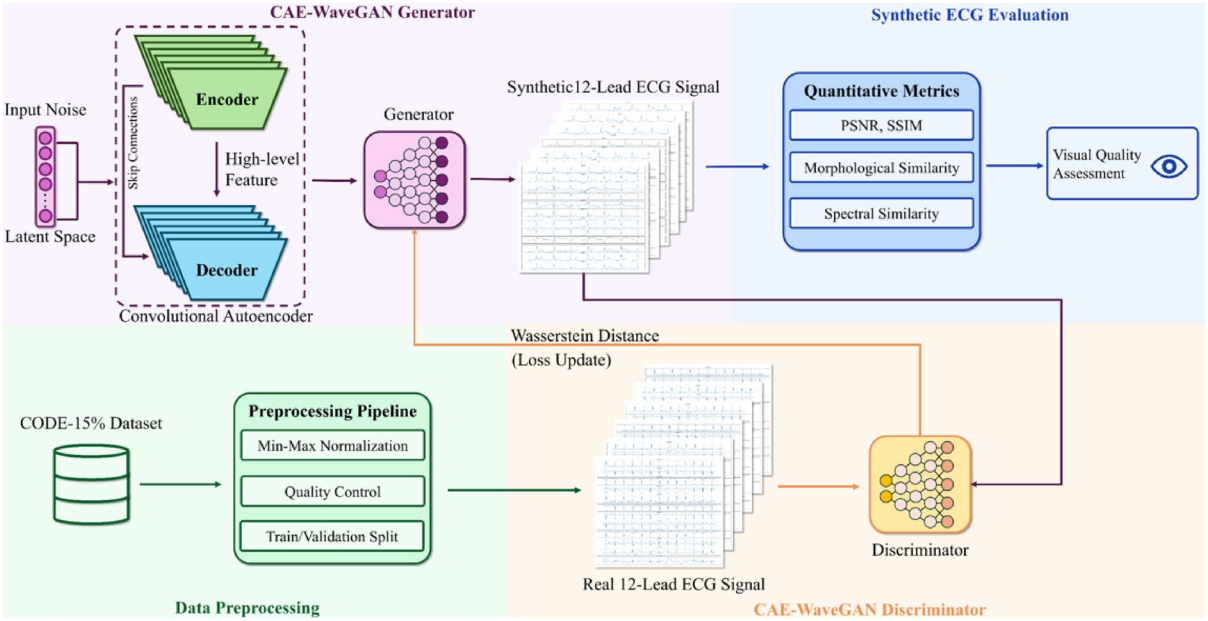

Fig. 2. Overview of the training process and architecture of CAE-WaveGAN for synthetic 12-Lead ECG signal generation.

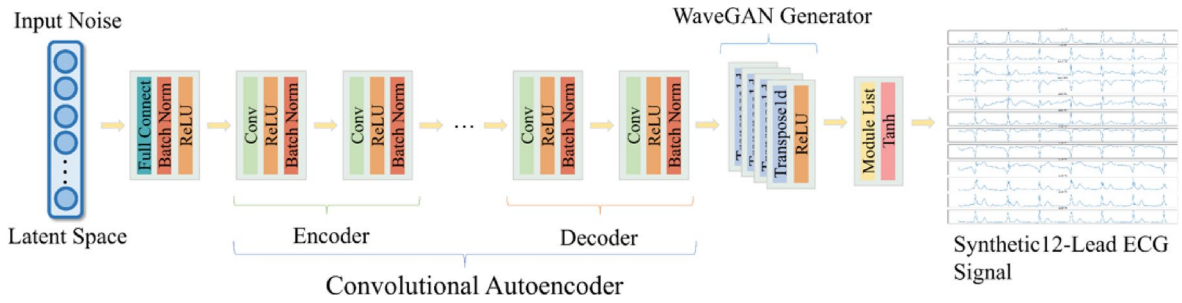

Fig. 3. The architecture of the generator, the skip connection methodology, and the ECG data generation process in CAE-WaveGAN.

experiment has its unique focus, they share a set of common procedural steps. These include the following steps: dataset splitting, data preprocessing, model definition, model training, and model evaluation:

Data splitting
In this research study, a 90% training and 10% testing split strategy was adopted. It formed the foundation of the 10-fold cross-validation using both the two-way split (train-test) and the conventional three-way split (train-validation-test). Through a comprehensive comparison, it is found that the two-way split consistently provided more stable and higher performance. This is because allocating a larger portion of data for training enables deep learning models to learn better and generalize complex data patterns, especially when the dataset size is limited. To ensure reproducibility, a consistent seed number for randomization was employed. Furthermore, the data was anonymised by removing labels to serve the specific objectives of the experiment.

Data preprocessing
Before commencing the training phase, ECG signals are preprocessed, comprising multiple files containing patient identifiers and their ECG readings. To enhance computational efficiency, the data is subsampled judiciously. Each ECG lead reading is normalized using a Min-Max Scaler transformation, scaling the values to a uniform range from −1 to 1. This approach was chosen because ECG signals naturally fluctuate around zero, and preserving this polarity is important for maintaining the physiological characteristics of the waveforms. Moreover, using a symmetric range centred at zero is beneficial for deep learning models that employ activation functions such as tanh, which can help improve convergence speed and training stability. This approach also helps reduce potential bias in the learning process by ensuring that the input data are centred. This step is crucial for reducing signal amplitude variations, facilitating a more consistent learning process for our correlation model. After normalization, the dataset is shuffled to ensure its random distribution. As a part of our quality

control procedures, a subset of 30 ECG samples from all leads is extracted for visual inspection by our team, ensuring data integrity before proceeding to CAE-WaveGAN model training.

Defining model

The subsequent phase involved establishing the architecture for the CAE-WaveGAN model. This process entails selecting the appropriate parameters and configurations to optimize model performance. Specific layers are designated to construct the Convolutional Autoencoders, which are central to our research. Additionally, the inclusion of jump connections was deliberated, assessing their potential to enhance model training by facilitating feature propagation.

Model training and evaluation

After defining the architecture of the model, training is using the designated training data set, which varies across experiments. Upon completing the training phase, the model's performance is evaluated using the test dataset. To ensure the credibility of the experimental results, extensive training is conducted, running each model for 500 epochs. During the training process, various loss functions are monitored, including the Wasserstein distance loss for the discriminator, the generator's loss, and the generator's performance on the experimental validation set. The implications of these losses will be discussed in a subsequent section.

Algorithm I delineates the training regimen for CAE-WaveGAN, over a pre-defined number of epochs (n_{epoch}), the algorithm iteratively refines a generator (G) and a discriminator (D) until they are optimally trained.

Initially, both G and D are set up alongside their respective optimizers, opt_{G} for G and opt_{D} for D. The number of iterations (N_{iter}) is computed as the product of n_{epoch} and the ceiling function applied to the ratio of sample size (N_{sample}) over the batch size. Furthermore, N_{iter_epoch}, indicating the iteration count per epoch, is determined. During the iterative training, which executes for N_{iter} iterations, the focus alternates between the discriminator and the generator. The discriminator's training precedes, spanning n_{critie} cycles within each iteration. Its objective is to differentiate authentic ECG signals from the synthetic signals produced by G. This step entails the computation of the discriminator's loss (D_{loss}) and the subsequent adjustment of D's parameters using opt_{D}.

Attention then shifts to the generator. With D's gradient updates disabled, G's gradients are enabled. Here, G crafts synthetic ECG signals from a noise vector (z_{noise}). The generator's loss (G_{loss}) is calculated as the negative mean discriminator output on these synthetic signals, prompting an update of G's weights via opt_{G}.

At every epoch's conclusion, identified when $\mathrm{iter} \bmod \mathrm{N}_{iter\_epoch}$ equals zero, the algorithm records the elapsed time, logs the losses for both G and D, and adjusts the learning rate for opt_{g} and opt_{d} if a learning rate decay is in place. Additionally, sample ECG signals are generated by G for inspection and storage. Upon completing N_{iter} iterations, the algorithm yields a proficiently trained G and D for deployment or further evaluation.

Algorithm 1: Training process of CAE-WaveGAN

Input: n_{epoch} (Number of training epochs), batch_size, N_{sample}
Output: Trained generator G and discriminator D

Initialize generator G and discriminator D;
Initialize optimizers opt_G and opt_D;
$N_{\text{iter}} \leftarrow n_{\text{epoch}} \times \lceil N_{\text{sample}}/\text{batch\_size} \rceil$;
$N_{\text{iter\_epoch}} \leftarrow \lceil N_{\text{sample}}/\text{batch\_size} \rceil$;
for $iter = 1$ **to** N_{iter} **do**
  Enable gradient updates for D only;
  for $n = 1$ **to** n_{critic} **do**
    real_ECG $\leftarrow$ Sample real ECG signals;
    $z_{\text{noise}} \leftarrow$ Sample noise from prior distribution;
    fake_ECG $\leftarrow G(z_{\text{noise}})$;
    $D_{\text{loss}} \leftarrow \text{CalculateLoss}(D, \text{real\_ECG}, \text{fake\_ECG})$;
    Update D using opt_D with D_{loss};
  Disable gradient updates for D;
  Enable gradient updates for G;
  $z_{\text{noise}} \leftarrow$ Sample noise from prior distribution;
  fake_ECG $\leftarrow G(z_{\text{noise}})$;
  $G_{\text{loss}} \leftarrow -\text{mean}(D(\text{fake\_ECG}))$;
  Update G using opt_G with G_{loss};
  if $iter \bmod N_{iter\_epoch} = 0$ **then**
    Record epoch timing;
    Store G_{loss} and D_{loss};
    if *decay_lr* **then**
      Update learning rates for opt_G and opt_D;
    Generate and save ECG samples using G;
return G, D

Tools

Our experiments require leveraging several existing tools. The following highlights the details:

Software

Programming Language Python programming language (Version 3.11.5, accessed on https://www.python.org/downloads/release/python-3115/) was used in all our experiments. Python provides many useful libraries for GAN tasks and deep learning experiments in general.

Libraries In these experiments, several publicly available software libraries are used as needed, including Numpy and Matplotlib. For deep learning algorithms, PyTorch, an open-source machine learning library widely used in applications such as computer vision and natural language processing, is utilized. In addition, structural similarity from the scikit-image library is used to assess image quality. Combining these tools allows us to implement and evaluate our model efficiently.

Hardware

Instead of depending on other computing resources, tests are run our tests on a locally available machine with powerful hardware, which is required for the high processing demands of our work. This system possesses a 12th Generation Intel (R) Core (R) i9-12900 @ 2.40 GHz processor, which provides a high-performance platform suitable for parallel processing and complex algorithms. This is paired with a robust NVIDIA RTX 3060 graphics card, which offers substantial GPU acceleration, a particularly beneficial feature for machine learning frameworks and graphically intensive applications.

Datasets

For a robust evaluation of our solution, the publicly available dataset CODE-15% dataset[31] is chosen, and describe its details below:

CODE-15% is a comprehensive dataset for ECG analysis featuring 12-lead ECG tracings. It comprises 345,779 exams from 233,770 unique patients. The Telehealth Network of Minas Gerais sourced the data between 2010 and 2016.

Each ECG exam entry includes patient ID, age, gender, predicted neural network age, presence of various cardiac conditions, a mortality flag, and follow-up time or time to death. The ECG tracings are stored in HDF5

files and are standardized to 4096 samples per lead to facilitate uniform analysis. In the original records, ECG tracings have two typical durations: some are 10 s long (4000 samples at 400 Hz), and others are 7 s long (2800 samples at 400 Hz). The minimum original signal length is 2800 samples, and the maximum is 4000 samples. To ensure a consistent input size for all exams, signals shorter than 4096 samples are symmetrically zero-padded at the beginning and end. Due to its large size, diversity, and comprehensive annotations, the CODE-15% dataset is particularly valuable for medical research. It is an indispensable resource for developing and testing advanced neural network models for ECG interpretation. CODE-15% facilitates in-depth studies into ECG-derived health predictions and mortality risk assessment.

Results and discussions

In this section, we evaluate the performance and stability of CAE-WaveGAN models with varying numbers of CAE layers and the incorporation of skip connections. Our analysis first examines the performance and stability differences between the original and our proposed models. We employ the standard deviation[32] as a typical comparative metric for the stability analysis. The formulation is given in Eq. 5, where σ denotes the standard deviation, N is the number of observations in the dataset, Y_x is the value of each observation, μ is the mean value of the dataset. The standard deviation quantifies the dispersion of dataset values, signifying the extent to which these values deviate from the mean (μ). In the context of our research, a higher standard deviation of the generator's loss during the model's training or validation phases suggests greater fluctuations. Such variability implies the model may exhibit insufficient stability throughout its training or validation process.

$$\sigma = \sqrt{\frac{1}{N}\sum_{x=1}^{N}(Y_x - \mu)^2} \tag{5}$$

To compare the performance of the models, we consider the loss in the training and validation phases since it indicates how well each model performs. In the next two subsections, we will explore the details of this analysis further.

Ablation study on batch size and learning rate of the proposed CAE-WaveGAN

We evaluated the impact of different batch sizes (32, 64, 128, and 256) and learning rates (1E-3, 1E-4, and 1E-5) on model performance and stability by comparing the standard deviation of the loss during training and validation phases. In addition, the average time per epoch and the standard deviation of the time spent per epoch for each combination were studied. All experiments were conducted on the same network architecture and dataset to ensure fair comparisons. Table 1 presents the results of the ablation study on batch size and learning rate for the proposed CAE-WaveGAN, including evaluation metrics: the standard deviation of the loss during training and validation phases, the average time per epoch, and the standard deviation of the time spent per epoch.

During the training process, we observed that when the learning rate was set to 1E-3, batch sizes of 32, 64, and 256 led to significant fluctuations in loss, and in some cases, the loss could not converge. In contrast, a batch size of 128 yielded relatively stable results, although it still underperformed compared to when the learning rates were decreased to 1E-4 and 1E-5. This indicated that a batch size of 128 gave the model a commendable capacity to resist fluctuations. Additionally, we noted that combining a batch size of 128 with a learning rate of 1E-4 and a batch size of 64 with a learning rate of 1E-5 resulted in similar standard deviations of loss during the training and validation phases. However, the former combination demonstrated higher efficiency and more stable training durations. Across three different learning rates, it was shown that an increase in the batch size typically resulted in a shorter average training duration during the training process. However, at a learning rate of 1E-5, the effect of batch size on the reduction of training time was less noticeable. Notably, the batch size 128 resulted in the least average training time at a learning rate 1E-4. We consider this to be indicative of superior

Learning rate	Batch size	Standard deviation		Epoch times (s)	
		Training loss of generator	Validation loss of generator	Standard Deviation	Average
1E-03	32	1205.5107	1230.7247	1.5033	413.5934
	64	1024.1884	1025.2282	2.9315	321.0210
	128	22.7918	26.9390	1.8830	236.7822
	256	1738.0792	1714.1352	3.7822	230.5701
1E-04	32	15.4738	15.3769	6.0432	406.3117
	64	15.9150	25.0364	4.1331	320.1466
	128	**10.7790**	**19.7521**	**0.8460**	**105.7031**
	256	14.2760	24.9031	1.7118	229.4406
1E-05	32	11.4577	28.9998	1.9272	321.0868
	64	10.0201	19.2200	1.4375	323.9666
	128	12.5774	27.7517	2.5086	307.4525
	256	18.6418	34.7183	3.9029	232.1860

Table 1. Standard deviations of generator training and validation Losses, and epoch times for original WaveGAN and CAE (6 layers)-WaveGAN. The best configuration is bolded.

Model	Standard deviation	
	Training loss of generator	Validation loss of generator
Original WaveGAN [26]	26.5180	45.6472
CAE (6 layers)-Wave GAN	10.7790	19.7521

Table 2. Standard deviations of generator training and validation losses for original WaveGAN and CAE (6 layers)-WaveGAN.

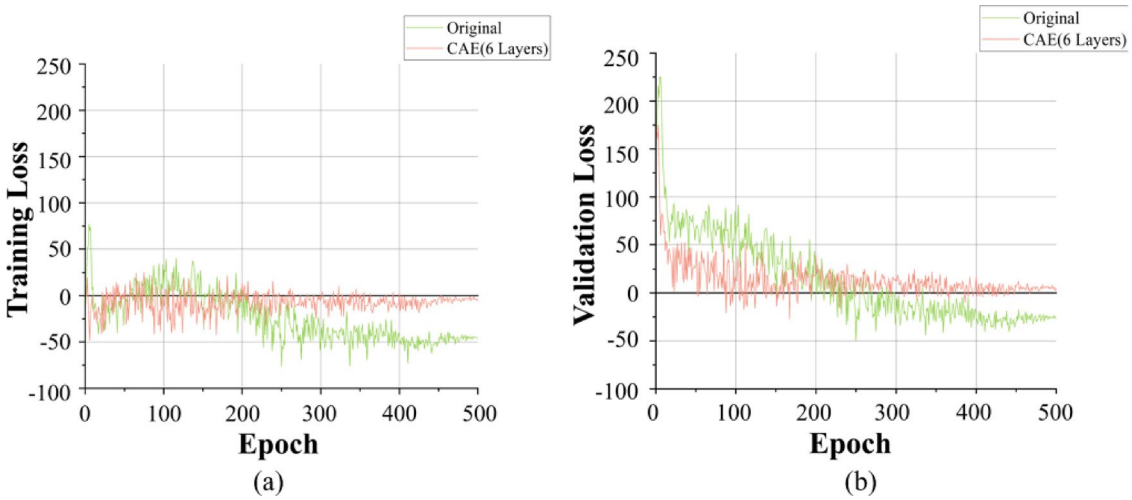

Fig. 4. Generator training (**a**) and validation (**b**) losses for original WaveGAN and CAE (6 layers)-WaveGAN.

performance. The results indicated that smaller batch sizes exhibit better performance on the validation set, particularly with a batch size of 128. Moreover, a learning rate setting 1E-4 demonstrated the best performance across all experiments, showing rapid convergence and stable performance. This suggested that the synergy of a batch size of 128 and a learning rate of 1E-4 could enhance the training efficiency and generalization ability of our proposed model while maintaining better stability during the training process.

Analysis of the performance of the proposed CAE-WaveGAN and original WaveGAN

As mentioned, we employed the same dataset to train the CAE-WaveGAN and the original WaveGAN for 500 epochs. The results revealed significant differences in stability between the two models. For the original WaveGAN, we used the best-performing hyperparameters, which were identified through ablation studies in prior research. Upon completion of its training, we obtained the results for this model. Furthermore, we compared the CAE-WaveGAN against the WaveGAN model by choosing the configuration that performed the best out of all the combinations. Table 2 presents the generator's training and validation loss standard deviation for both models. The stability performance of any combination of our proposed model outperformed the original WaveGAN model. The generator training loss standard deviation for the original WaveGAN model was 26.5180, with a validation loss standard deviation of 45.6472. In contrast, our proposed CAE (6 layers)-WaveGAN model achieved a generator training loss standard deviation of 10.7790 and a validation loss standard deviation of 19.7521. This indicates that, throughout the training process, our proposed model exhibited significantly enhanced stability compared to the original WaveGAN model.

Throughout 500 training epochs, we tracked the model generators' performance based on loss values for the two models at each stage of training and validation. The proposed CAE (6 layers)-Wave GAN demonstrated less overall fluctuation during the training stage, with a stable trend in the loss curve and a faster convergence rate, as illustrated in Fig. 4a. Our model demonstrated clear stability around the 300-epoch mark and achieved satisfactory convergence around 400 epochs, with the loss gradually approaching zero. The loss curve was smooth, without any significant signs of further overfitting. In contrast, the traditional WaveGAN's loss curve still experienced severe fluctuations between 200 and 300 epochs, with the loss remaining around − 50 in the rest. Notably, the loss values can be both positive and negative due to the subtraction involved in the calculation of the discriminator loss during adversarial training, which is a common characteristic in Wasserstein GAN frameworks. In this context, values closer to zero—regardless of sign—indicate better convergence and lower reconstruction error for both models. As depicted in Fig. 4b, during the validation phase, the loss of the original model remained higher than that of our proposed model at the onset. Similarly, our model performed remarkably well, showing clear stability by 200 epochs and stabilizing near zero loss at the end, consistent with the training phase performance, without any apparent overfitting. To summarize the results in Fig. 4, the CAE (6 layers) model exhibits lower and more stable loss in both training and validation compared to the original WaveGAN

Number of layers in CAE	Standard deviation		Epoch times (s)	
	Training loss of generator	Validation loss of generator	Standard deviation	Average
2	19.7596	37.0138	1.0251	100.6066
3	14.3179	23.0265	1.2467	101.4278
4	14.4196	27.5480	1.0512	101.8182
5	12.3213	24.0072	1.3816	103.9110
6	10.7790	19.7521	0.8460	105.7031
7	12.7710	21.6471	0.9638	102.7245
8	11.0726	24.2010	1.6524	376.6295

Table 3. Standard deviations of generator training and validation losses and epoch times with various CAE layer configurations utilization for CAE-WaveGAN.

Number of layers in CAE with (w) and without (w/o) skip connections	Standard deviation		Epoch times (s)	
	Training loss of generator	Validation loss of generator	Standard deviation	Average
2 (w/o)	19.7596	37.0138	1.0251	100.6066
2 (w)	14.9641	26.9956	0.7673	169.4583
3 (w/o)	14.3179	23.0265	1.2467	101.4278
3 (w)	14.8717	24.9862	1.7522	203.6549
4 (w/o)	14.4196	27.5480	1.0512	101.8182
4 (w)	18.7561	28.0162	0.9960	236.2445
5 (w/o)	12.3213	24.0072	1.3816	103.9110
5 (w)	17.9302	22.4937	1.3853	271.5365
6 (w/o)	10.7790	19.7521	0.8460	105.7031
6 (w)	11.2175	20.6439	2.1845	305.1484
7 (w/o)	12.7710	21.6471	0.9638	102.7245
7 (w)	10.3109	21.4862	2.1327	340.2040
8 (w/o)	11.0726	24.2010	1.6524	376.6295
8 (w)	12.1674	18.8121	1.6524	376.6295

Table 4. Standard deviations of generator training and validation losses and epoch times with various CAE layer configurations and skip connection utilization for CAE-WaveGAN.

model. This demonstrates that the CAE architecture is designed to enhance feature extraction and improve the model's ability to represent ECG signals, resulting in better convergence and less fluctuation.

Analysis of the number of CAE layers and skip connections in the generator

The impact of the number of layers in the CAE of the generator within our proposed model and the presence of skip connections on the overall model performance were investigated. Considering the CAE-WaveGAN without skip connections: We configured seven models composed of CAE with encoder and decoder structures ranging from 2 to 8 layers, wherein the number of layers in the encoder and decoder is identical. To maintain the uniformity and fairness of the training process, all models were subjected to identical parameter settings. Specifically, a learning rate of 1E-4 and a batch size of 128 were employed based on previously identified optimal conditions. This approach ensures that comparisons among different models are consistent and reliable, facilitating a controlled evaluation of model performance under standardized conditions. We calculated the standard deviation of the training and validation losses and epoch times for these seven models over 500 epochs, as illustrated in Table 3. Notably, the CAE (6 layers)-WaveGAN, featuring six layers in both the encoder and decoder, exhibited the lowest standard deviations, indicating that this configuration was the most stable during training and validation among all models tested.

In the models with encoder and decoder layers ranging from 2 to 6, the standard deviation decreased as the number of layers increased, reaching a minimum with the six-layer configuration. However, as we added more layers, with 7 and 8 layers in the encoder and decoder, the stability during training progressively worsened, and the standard deviations increased correspondingly. The results revealed that while a modest increase in CAE layers could significantly aid in stabilizing the training process when the CAE has fewer layers, training stability diminishes with further enlarged models. This instability might be attributed to the increased complexity of the model. Therefore, determining an appropriate number of layers for the CAE is of paramount importance.

Upon integrating skip connections into our model, we observed a decline in the training stability of the modified structures. Table 4 presents the standard deviations of the training and validation losses and epoch times for the generator in our proposed CAE-WaveGAN model, spanning different numbers of CAE layers with and without the introduction of skip connections. It was found that skip connections only enhanced the

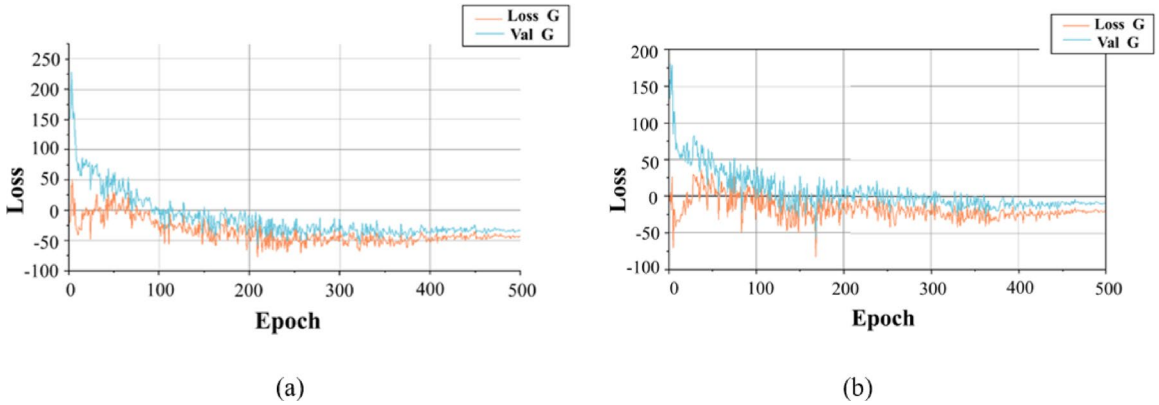

Fig. 5. Training and validation losses for CAE (2 layers)-WaveGAN generators: (**a**) standard and (**b**) with skip connections.

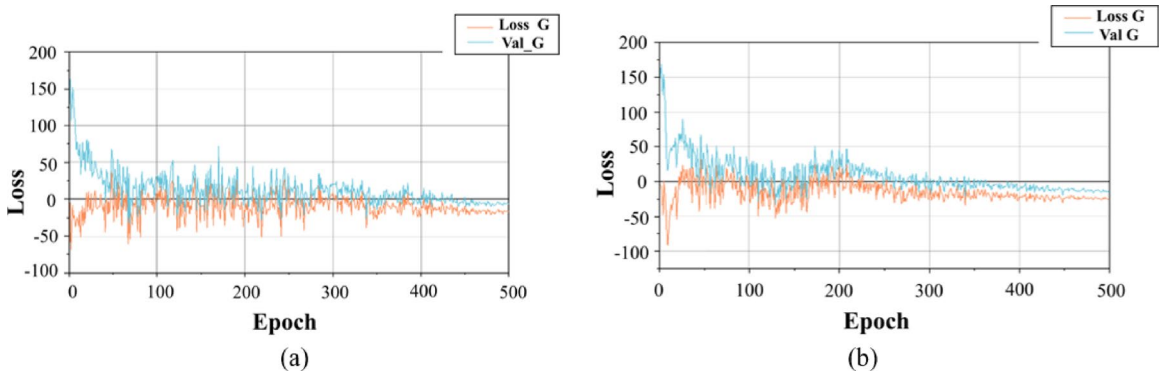

Fig. 6. Training and validation losses for CAE (3 layers)-WaveGAN generators: (**a**) standard and (**b**) with skip connections.

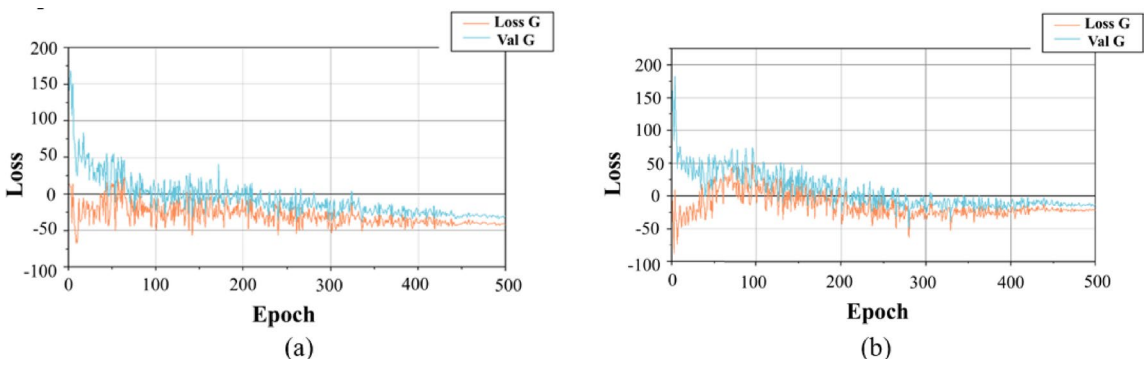

Fig. 7. Training and validation losses for CAE (4 layers)-WaveGAN generators: (**a**) standard and (**b**) with skip connections.

stability of models constituted by encoders and decoders with 2 and 7 layers, respectively. We contend that skip connections can help feature propagation by transmitting information from previous layers or by recycling features from previous layers when the model is too simple or too complicated, improving the model's stability and efficiency.

Despite the overall minimal contribution of skip connections to the enhancement of model stability, we cannot dismiss the assistance afforded by skip connections during the training process. Figures 5, 6, 7, 8, 9, 10 and 11 comprehensively illustrates the training and validation losses of the generators across seven distinct configurations of our proposed CAE-WaveGAN. In Figs. 6 and 8, the models incorporating skip connections demonstrated marginally lower training stability than their counterparts without such structures. Despite this,

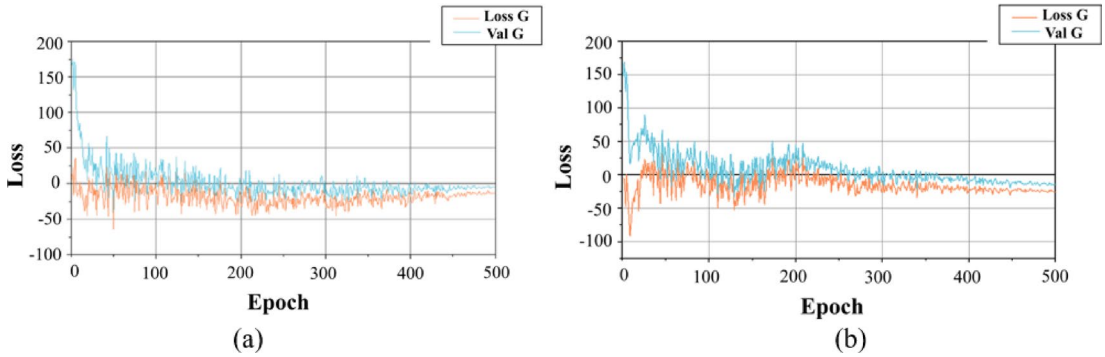

Fig. 8. Training and validation losses for CAE (5 layers)-WaveGAN generators: (**a**) standard and (**b**) with skip connections.

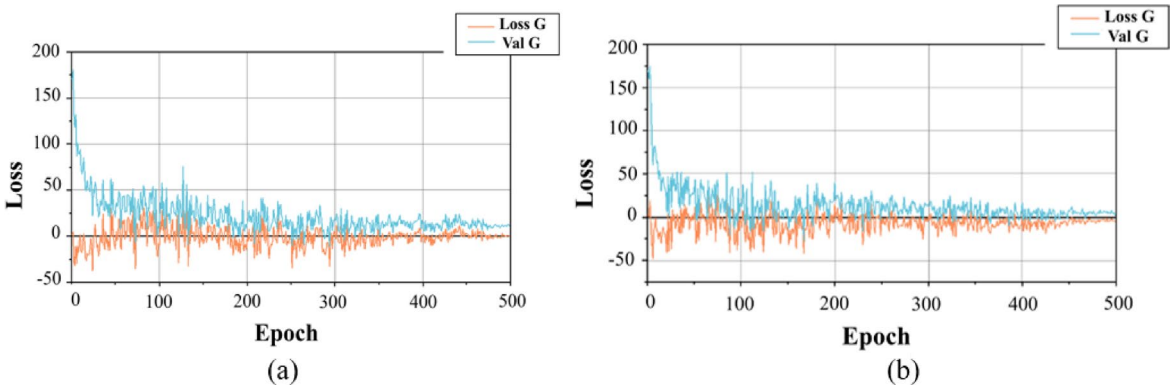

Fig. 9. Training and Validation Losses for CAE (6 layers)-WaveGAN Generators: (**a**) Standard and (**b**) with Skip Connections.

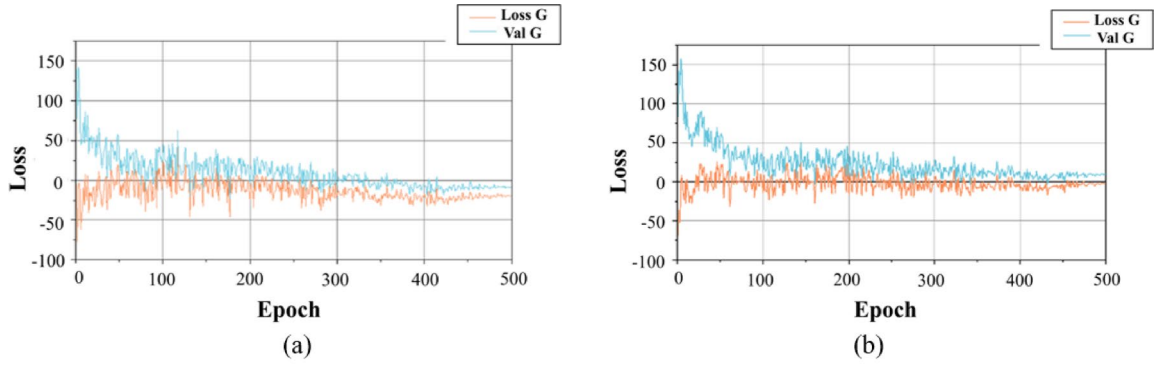

Fig. 10. Training and validation losses for CAE (7 layers)-WaveGAN generators: (**a**) standard and (**b**) with skip connections.

the disparity in performance was relatively small, suggesting that for CAE with 3 and 5 layers, the addition of skip connections does not contribute significantly to enhancing model performance. This observation may imply that for some simpler or shallower network architectures, the inherent complexity added by skip connections does not translate into substantial benefits, possibly due to the adequate learning capacity of the network without necessitating additional pathways for gradient flow or feature reuse.

Contrastingly, as depicted in Figs. 9, 10 and 11, while models with skip connections continued to experience slightly reduced stability during the training phase compared to models devoid of these connections, they exhibited markedly improved stability during the validation phase. This enhancement in performance during validation suggests that skip connections confer a significant advantage in scenarios involving more complex or

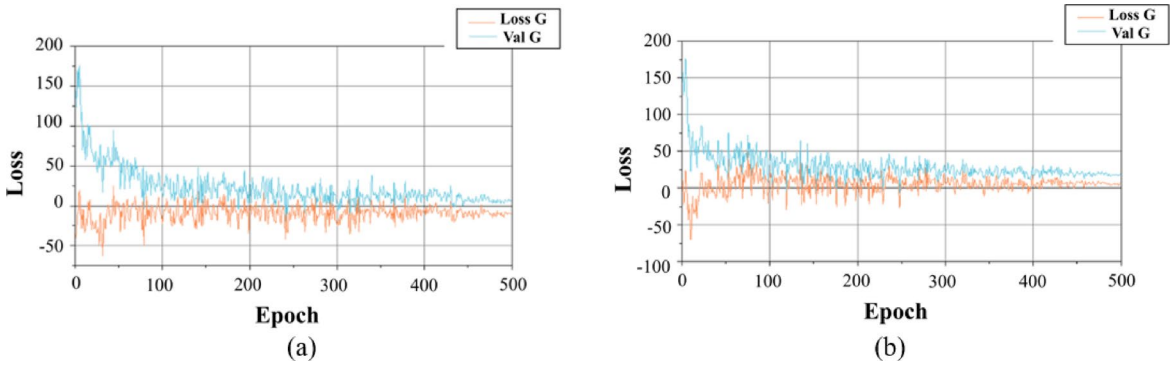

Fig. 11. Training and Validation Losses for CAE (8 layers)-WaveGAN generators: (**a**) standard and (**b**) with skip connections.

Number of layers in CAE	Standard deviation		Epoch times (s)	
	Training loss of generator	Validation loss of generator	Standard Deviation	Average
2 layers	3.3402	2.9784	0.9713	100.5650
2 layers with skip connections	3.5849	3.0023	0.9417	169.5778
3 layers	4.4823	4.8252	1.0609	101.0722
3 layers with skip connections	2.4744	3.0239	0.9205	205.8229
4 layers	3.1174	3.3462	0.9301	101.6271
4 layers with skip connections	3.1293	3.7106	0.9939	236.0755
5 layers	3.8363	3.2427	0.8293	104.5830
5 layers with skip connections	3.0331	2.9771	1.0869	272.0150
6 layers	4.0582	3.8862	0.8417	105.6567
6 layers with skip connections	3.9294	4.0854	1.2931	306.7158
7 layers	3.7371	3.4304	1.0581	102.6527
7 layers with skip connections	3.6994	3.2303	1.3968	341.7712
8 layers	4.7231	5.1779	1.6961	104.5222
8 layers with skip connections	3.2786	3.5719	1.6333	376.6644

Table 5. Standard deviations of generator training and validation losses between 400 and 499 epochs with various CAE layer configurations and skip connection utilization for CAE-WaveGAN.

deeper network architectures. In particular, for CAE with a more significant number of layers, skip connections play a crucial role in mitigating the effects of vanishing gradients, thereby maintaining effective training over extended layers.

Figures. 5, 7, and 10 notably show that models with skip connections exhibit end-of-training losses closer to 0 compared to the standard CAE-WaveGAN, signifying superior model performance and better generalization capabilities. The lower training losses in models with skip connections suggest that these architectures are more efficient at capturing and preserving essential information throughout the network layers. Skip connections, by allowing direct pathways for data flow between non-adjacent layers, help mitigate the degradation of information that typically occurs in deeper networks. This is particularly crucial in generative models like WaveGAN, where output quality relies heavily on the nuanced and precise representation of input data features.

As our research progressed, we observed that skip connections significantly enhance the stability of training and validation in the later stages. All training and validation processes extracted data from epochs 400 to 499. Similarly, we computed the standard deviation for all models during this training phase. Refer to Table 5, models equipped with skip connections demonstrated smaller standard deviations in the last 100 epochs than the standard models. This suggests that skip connections may mitigate the vanishing gradient problem as training progresses, thereby stabilizing gradient updates. Such stabilization is immensely beneficial during the advanced stages of model training.

Incorporating skip connections serves as a stabilizing force within the network, allowing for sustained learning and adaptation without the typical degradation that can occur with deep architectures as they converge. This finding aligns with the literature that posits skip connections as a facilitator of continued learning capacity, especially in deep neural networks where information persistence is crucial. The empirical evidence of reduced variance in loss values reinforces the theoretical understanding that skip connections can help maintain the flow of gradients throughout the layers of the network, thus playing a crucial role in the training efficacy of deep learning models.

Model	Best MorphSim	Best Epoch	Best SpecSim	Best Epoch
Proposed CAE-WaveGAN	0.549336	273	0.830069	471
dcGAN[33]	0.261403	34	0.75162	4
ECGGAN[34]	0.511233	201	0.80686	386
Original WaveGAN[26]	0.530098	317	0.824969	347
WGAN[35]	0.458608	124	0.809738	371

Table 6. Morphological and spectral similarity performance Comparison.

Model	Train PSNR	Val PSNR	SSIM
Proposed CAE-WaveGAN	3.995 ± 0.182	3.664 ± 0.184	0.18 ± 0.02
dcGAN [33]	2.977 ± 0.174	2.978 ± 0.182	0.004 ± 0.002
ECGGAN [34]	3.24 ± 0.435	3.259 ± 0.452	0.106 ± 0.029
Original WaveGAN [26]	3.334 ± 0.336	3.325 ± 0.306	0.113 ± 0.162
WGAN [35]	3.149 ± 0.117	3.151 ± 0.115	0.104 ± 0.01

Table 7. Morphological and spectral similarity performance Comparison.

Evaluation metrics and comparative performance

To comprehensively assess the quality and clinical relevance of generated ECG signals, we employed a suite of domain-specific evaluation metrics that capture multiple aspects of signal fidelity. Peak Signal-to-Noise Ratio (PSNR) measures the similarity between generated and real ECG signals in terms of signal reconstruction quality, providing a quantitative assessment widely recognized in the signal processing community. The Structural Similarity Index Measure (SSIM) evaluates the structural correspondence between synthetic and authentic ECG waveforms, effectively capturing the essential clinical features that are crucial for accurate diagnostic interpretation. Morphology Similarity (MorphSim) focuses specifically on waveform shape characteristics, assessing the morphological resemblance that is fundamental to ECG pattern recognition and clinical analysis. Spectral Similarity (SpecSim) compares the frequency-domain characteristics of signals, ensuring that generated ECGs preserve the spectral patterns essential for accurate clinical interpretation and automated analysis systems.

The comparative analysis included several established generative models to provide comprehensive benchmarking. The Deep Convolutional GAN (DCGAN)[33] represents a foundational approach in generative modeling, while ECGGAN[34] is a specialized variant designed specifically for ECG synthesis. The Wasserstein GAN (WGAN)[35] incorporates improved training stability through the Wasserstein distance metric, and the original WaveGAN serves as the baseline architecture specifically designed for raw audio and time-series generation. Each model was trained under identical conditions and evaluated using the same metrics to ensure fair comparison.

The experimental results demonstrate that our proposed CAE-WaveGAN achieves superior performance across all evaluation metrics, establishing new benchmarks for ECG signal generation quality. As shown in Table 6, CAE-WaveGAN achieved a MorphSim score of 0.549, representing a 3.6% improvement over the best baseline method (original WaveGAN, which achieved a score of 0.530). This improvement is particularly significant as morphological accuracy is critical for clinical applications where waveform shape directly impacts diagnostic interpretation. The spectral similarity results are even more compelling, with CAE-WaveGAN reaching 0.830 in SpecSim, substantially outperforming all comparative methods and demonstrating superior preservation of frequency-domain characteristics essential for ECG analysis.

The signal reconstruction quality metrics presented in Table 7 further underscore the effectiveness of our approach. CAE-WaveGAN achieved a training PSNR of 3.995 ± 0.182 and a validation PSNR of 3.664 ± 0.184, representing a remarkable 19.8% improvement over the second-best-performing method. The SSIM results are particularly noteworthy, with CAE-WaveGAN achieving 0.18 ± 0.02, which represents a 59.3% improvement over the next best method and demonstrates significantly enhanced structural preservation. In contrast, DCGAN showed the poorest performance across all metrics, achieving only 0.261 in MorphSim and 0.004 ± 0.002 in SSIM, highlighting the challenges of applying general-purpose GANs to specialized medical time-series data. ECGGAN and WGAN demonstrated moderate performance, with ECGGAN achieving competitive spectral similarity but falling short in morphological and structural metrics.

The multi-faceted evaluation approach confirms that CAE-WaveGAN not only generates visually realistic ECG signals but also preserves the critical clinical and physiological properties necessary for reliable medical applications, establishing a new paradigm for high-fidelity ECG synthesis.

To ensure clinical viability and filter out low-quality synthetic signals, a rigorous threshold-based quality selection system was implemented across all evaluation metrics. Specifically, an SSIM threshold was established for high-quality sample selection, complemented by threshold criteria of MorphSim ≥ 0.50 and SpecSim ≥ 0.80. This quality filtering mechanism automatically identifies and separates high-quality synthetic ECG signals that preserve essential physiological and morphological characteristics required for clinical interpretation. The visual comparisons and analysis presented in subsequent sections exclusively showcase ECG signals that have successfully passed these stringent quality thresholds.

Visual comparison of ECG signal quality

To provide a qualitative assessment of synthetic ECG fidelity and demonstrate the superior performance of the CAE-WaveGAN approach, side-by-side visual comparisons between real and synthetic ECG signals across all evaluated models are presented. Given the extensive nature of 12-lead ECG data and the need for clear visual interpretation, three representative leads are strategically selected for detailed comparison: Lead I, Lead aVR, and Lead V1. All synthetic ECG signals used in the comparison are generated from the final epoch of each model's training process, ensuring fair and consistent evaluation across all approaches.

The selection of these specific leads is based on their distinct electrophysiological characteristics and diagnostic significance. Lead I represent the standard limb leads and provides fundamental information about cardiac electrical activity in the frontal plane, making it essential for assessing basic rhythm and conduction patterns. Lead aVR offers a unique perspective from the right shoulder, typically displaying inverted waveforms that serve as an important reference for detecting certain cardiac abnormalities and validating signal authenticity. Lead V1, as a precordial lead positioned over the right ventricle, captures critical information about ventricular depolarization and is particularly sensitive to morphological variations in QRS complexes and T-wave patterns. The selected leads collectively demonstrate the models' performance in generating physiologically plausible P-waves, QRS complexes, and T-waves across different anatomical perspectives, providing comprehensive insight into synthetic ECG quality without overwhelming visual presentation.

Figures 12, 13 and 14 presents a comprehensive visual comparison of Lead I, Lead aVR and Lead V1 ECG signals across all evaluated generative models against real ECG data as the reference standard. The real ECG signal exhibits characteristic physiological features with well-defined P-waves, QRS complexes, and T-waves, etc., maintaining regular rhythm with consistent R-R intervals and stable baseline throughout the recording.

The CAE-WaveGAN demonstrates exceptional fidelity to the real ECG signal, closely replicating authentic morphological characteristics across all waveform components. The generated signals preserve precise P-wave timing, accurate QRS complex shapes, and physiologically plausible T-wave patterns while maintaining appropriate rhythm consistency without artificial artifacts.

The original WaveGAN shows reasonable overall performance but reveals notable quality limitations compared to the proposed method. While QRS complexes remain recognizable, the model exhibits morphological distortions, inconsistent amplitude scaling, and loss of fine-grained waveform details crucial for clinical interpretation.

WGAN exhibits significant quality degradation with severe morphological distortions throughout the generated signals. The output contains high-frequency noise components that are physiologically unrealistic, leading to loss of clear wave definition and inconsistent amplitude scaling, which renders the signals clinically unusable.

ECGGAN demonstrates moderate performance with basic ECG structure preservation and recognizable major wave components. However, the model shows reduced amplitude consistency and less preserved fine morphological details compared to CAE-WaveGAN, with noticeable rhythm variability that deviates from physiological norms.

DCGAN shows complete failure in ECG generation, producing output that resembles pure noise rather than physiological signals. The complete absence of discernible ECG waveform patterns and cardiac rhythm structure indicates a fundamental training failure and an inability to capture the temporal dependencies inherent in ECG signals.

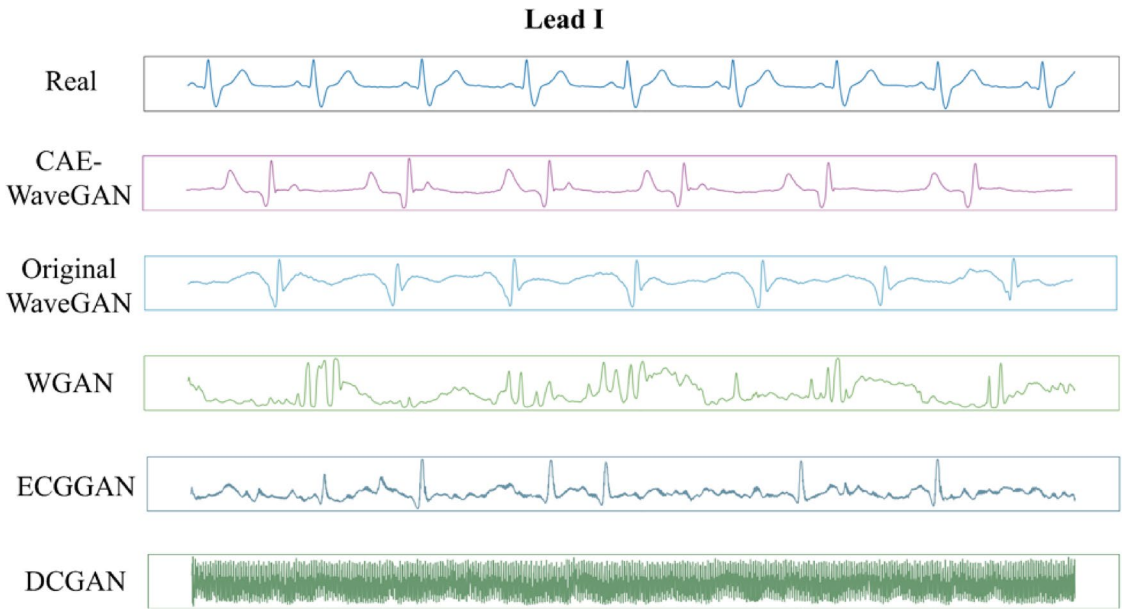

Fig. 12. Visual comparison of lead I ECG signals across different generative models.

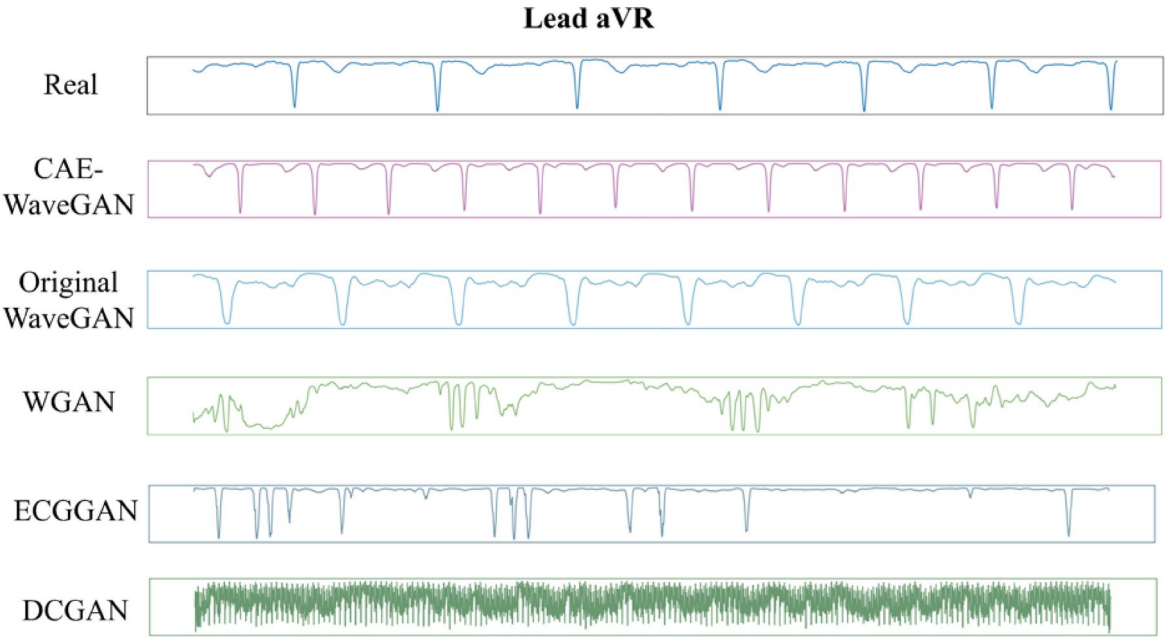

Fig. 13. Visual comparison of lead aVR ECG signals across different generative models.

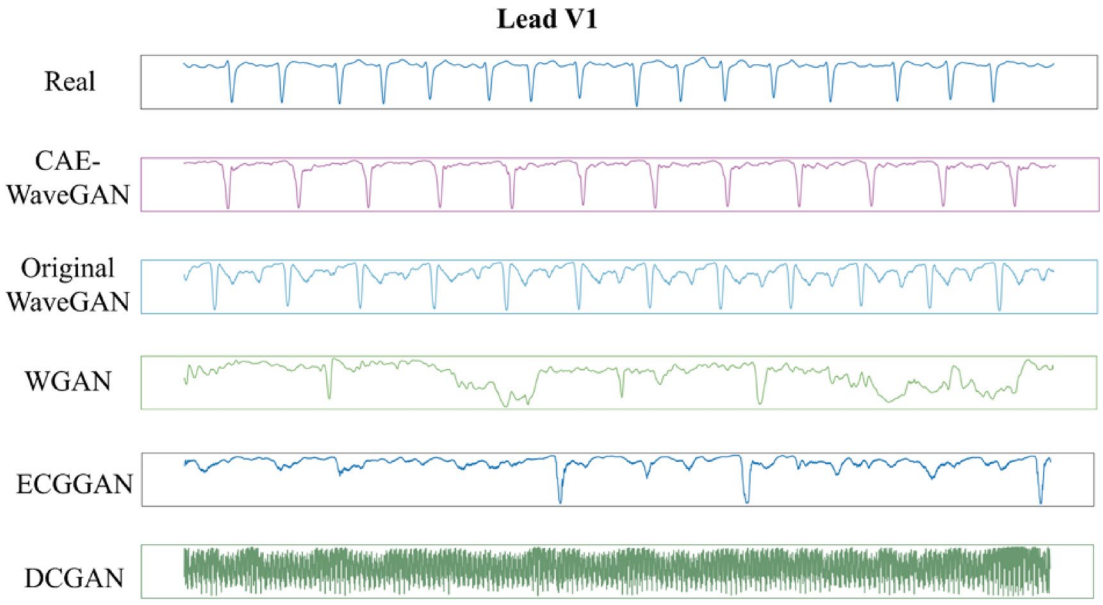

Fig. 14. Visual comparison of lead aVR ECG signals across different generative models.

Overall, the comparative analysis clearly demonstrates the superior performance of CAE-WaveGAN in preserving both macro and micro-level ECG characteristics.

Cross-Dataset performance evaluation and model generalizability

To further validate the robustness and generalizability of our proposed CAE-WaveGAN architecture, additional experiments were conducted on the PTB-XL dataset[36–38], which represents one of the largest publicly available clinical ECG databases containing over 21,000 10-second 12-lead ECG records from clinical routine. This cross-dataset evaluation provides crucial insights into the proposed model's adaptability across different data distributions and clinical populations, demonstrating its potential for broader clinical applications beyond the Code-15% benchmark dataset.

The cross-dataset evaluation results presented in Tables 8 and 9 demonstrate the remarkable adaptability and consistent performance of our CAE-WaveGAN architecture across diverse ECG datasets, as measured by MorphSim and SpecSim metrics, and PSN and SSIM metrics, respectively. When trained on the PTB-XL dataset,

Dataset	Best MorphSim	Best Epoch	Best SpecSim	Best Epoch
Code-15%	0.529336	273	0.830069	471
PTB-XL	0.593221	483	0.921263	166

Table 8. Morphological and spectral similarity performance comparison of the proposed model.

Dataset	Train PSNR	Val PSNR	SSIM
Code-15%	3.299 ± 0.182	3.266 ± 0.184	0.15 ± 0.02
PTB-XL	6.377 ± 0.316	6.327 ± 0.321	0.323 ± 0.027

Table 9. Morphological and spectral similarity performance comparison of the proposed model.

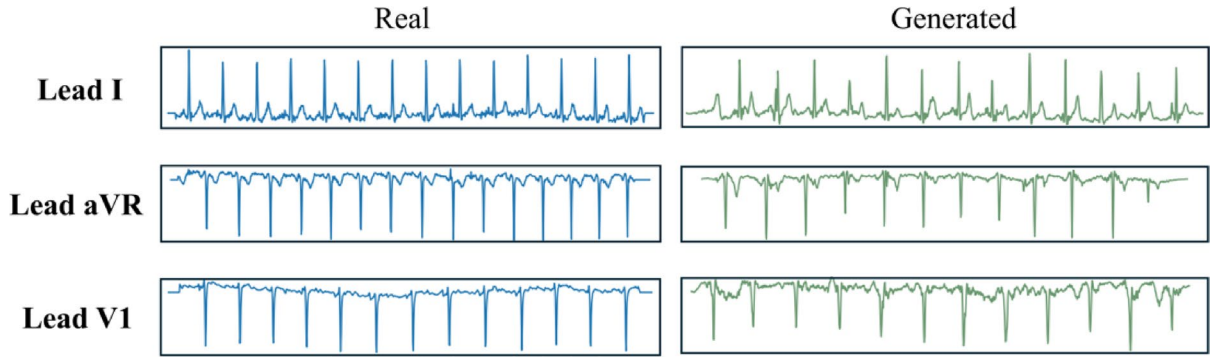

Fig. 15. Visual comparison of real vs. generated ECG signals on PTB-XL dataset (Leads I, aVR, and V1).

the proposed model achieved a MorphSim score of 0.593, representing a 12.1% improvement over the Code-15% baseline, while the SpecSim metric reached an impressive 0.921, indicating exceptional preservation of frequency-domain characteristics. The signal reconstruction quality metrics show even more pronounced improvements, with PSNR values reaching 6.377 ± 0.316 for training and 6.327 ± 0.321 for validation, representing substantial enhancements in signal fidelity. The SSIM metric achieved 0.323 ± 0.027 on PTB-XL, more than doubling the performance observed on Code-15%, highlighting the model's enhanced structural similarity preservation capabilities across different data distributions.

The qualitative assessment presented in Fig. 15 provides visual confirmation of these quantitative improvements, demonstrating that ECG signals generated from the PTB-XL-trained model exhibit exceptional morphological fidelity across the three representative leads (I, aVR, and V1). The generated signals closely replicate the characteristic waveform patterns, temporal dynamics, and amplitude relationships observed in real ECG data, validating the numerical metrics through direct visual inspection. These results underscore several critical aspects of our proposed architecture's design philosophy and practical utility. The consistent high-quality performance across both datasets validates the universal applicability of the CAE-WaveGAN framework, demonstrating that the convolutional autoencoder's feature extraction capabilities effectively adapt to different ECG data characteristics without requiring fundamental architectural modifications.

Conclusion

In conclusion, our research presents a significant advance in generating synthetic 12-lead ECG data using a CAE-WaveGAN model, which has been meticulously engineered to produce high-quality images that could be utilized for clinical training and diagnostic algorithm enhancement. Integrating a Convolutional Autoencoder within the WaveGAN framework has proven to be a robust approach, providing the dual benefits of detailed feature extraction and efficient image synthesis. Our extensive experiments and comparative studies on the CODE-15% dataset have demonstrated that the optimal configuration of our model, the CAE (6 layers)-WaveGAN, offers superior performance in terms of training stability and validation loss, hence ensuring the generation of realistic ECG images.

The merits of our proposed CAE-WaveGAN approach are multifaceted and significant for the field of synthetic medical data generation. First, our method demonstrates superior performance across all evaluation metrics, achieving a 19.8% improvement in PSNR and a 59.3% enhancement in SSIM compared to baseline methods, thereby establishing new benchmarks for ECG synthesis quality. Second, the integration of a convolutional autoencoder with WaveGAN provides enhanced training stability, as evidenced by a significantly reduced loss standard deviation (10.7790 vs. 26.5180 for the original WaveGAN), addressing a critical limitation of existing generative models. Third, our approach successfully preserves complex inter-lead relationships essential for

12-lead ECG clinical validity, demonstrated through comprehensive morphological and spectral similarity assessments. Fourth, the method offers practical clinical applicability by generating synthetic ECG data that maintains physiological authenticity while addressing privacy concerns inherent in medical data sharing.

Moreover, the analysis of the impact of skip connections within the CAE-WaveGAN architecture has revealed that while they contribute positively to the performance of certain model configurations, their role is not universally beneficial and requires careful consideration based on model complexity. This nuanced understanding of skip connections in CAE-WaveGAN reinforces the need for a tailored approach to model architecture design.

The significance of this research is not limited to the generation of synthetic ECG data. Still, it also opens new avenues of research on generative models for other types of medical data where privacy concerns and data scarcity are prevalent issues. The proposed CAE-WaveGAN model can be a versatile tool for researchers and clinicians to extend medical datasets safely and ethically. Future work may explore the application of this model to different types of biomedical markers.ls and images, further enhancing its utility and impact on the medical community.

The success of the CAE-WaveGAN model also demonstrates that it has applications in real-world clinical scenarios, such as augmenting the dataset used to train machine learning models when detecting and diagnosing cardiac abnormalities. As we continue to refine the model and its applications, we hope this work will significantly advance the use of AI in medicine, helping to develop more accurate and reliable diagnostic tools that will ultimately improve patient diagnostic efficiency and treatment results.

Despite the promising results of our CAE-WaveGAN approach, several limitations should be acknowledged. First, the computational requirements of the proposed model, particularly for configurations with more CAE layers and skip connections, may limit its practical deployment in resource-constrained environments. Additionally, it is planned to conduct comprehensive evaluations in future research to assess whether models trained on the synthetic data achieve comparable performance to those trained on real data when applied to clinical diagnostic tasks. Finally, this study concentrated on 12-lead ECG synthesis, and the applicability of the proposed approach to other biomedical signal types or ECG configurations remains to be explored. These limitations highlight important directions for future research: (i) clinical validation studies should be conducted to assess whether diagnostic algorithms trained on our synthetic ECG data achieve comparable performance to those trained exclusively on real data when applied to actual clinical diagnostic tasks; (ii) optimization of computational efficiency through model compression techniques, knowledge distillation, or progressive training strategies could enable deployment in resource-limited clinical settings; (iii) extending the framework to generate other biomedical signals, such as EEG, EMG, or multi-modal physiological data could broaden the impact of our approach; (iv) developing adaptive architectures that can automatically adjust complexity based on input data characteristics could improve generalizability across diverse clinical populations and recording conditions; (v) incorporating domain-specific medical knowledge through constrained generation or guided synthesis could ensure that rare cardiac conditions are adequately represented in synthetic datasets; and (vi) investigating federated learning approaches that leverage our synthetic data generation capabilities could enable collaborative model development while maintaining patient privacy across multiple healthcare institutions.

Data availability

The datasets generated and/or analysed during the current study are available in the CODE-15% dataset repository, https://zenodo.org/records/4916206.

Received: 30 April 2025; Accepted: 15 September 2025
Published online: 17 October 2025

References

1. Ketu, S. & Mishra, P. K. An intelligent hybrid classification model for heart disease detection using imbalanced electrocardiogram signals. *J Supercomput 80*. 4286–4308. https://doi.org/10.1007/s11227-023-05583-8 (2024).
2. Sampath, A. & Sumithira, T. R. Sparse based recurrent neural network long short term memory (RNN-LSTM) model for the classification of ECG signals. *Appl. Artif. Intell.* 36, 2018183 (2022). https://doi.org/10.1080/08839514.2021.2018183
3. Singh, S. et al. Meta-analysis of the performance of AI-driven ECG interpretation in the diagnosis of valvular heart diseases. *Am. J. Cardiol.* **213**, 126–131. https://doi.org/10.1016/j.amjcard.2023.12.015 (2024).
4. Chen, J. et al. Congenital heart disease detection by pediatric electrocardiogram based deep learning integrated with human concepts. *Nat Commun. 15*. **44930** https://doi.org/10.1038/s41467-024-44930-y (2024).
5. Ramkumar, G. et al. IoT-based patient monitoring system for predicting heart disease using deep learning. *Measurement 218*. 113235. https://doi.org/10.1016/j.measurement.2023.113235 (2023).
6. Alzubaidi, L. et al. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J Big Data 10*. **66** https://doi.org/10.1186/s40537-023-00727-2 (2023).
7. Piccialli, F. et al. A survey on deep learning in medicine: why, how and when? *Inf. Fusion*. **66**, 111–137. https://doi.org/10.1016/j.inffus.2020.09.006 (2021).
8. Monachino, G. et al. Deep generative models: the winning key for large and easily accessible ECG datasets? *Comput. Biol. Med.* **167**, 107655. https://doi.org/10.1016/j.compbiomed.2023.107655 (2023).
9. Sajeeda, A. & Hossain, B. M. M. Exploring generative adversarial networks and adversarial training. *Int. J. Cogn. Comput. Eng.* **3**, 78–89. https://doi.org/10.1016/j.ijcce.2022.03.002 (2022).
10. Skandarani, Y., Jodoin, P. M. & Lalande, A. GANs for medical image synthesis: an empirical study. *J Imaging 9*. **69** https://doi.org/10.3390/jimaging9030069 (2023).
11. Wulan, N. et al. Generating electrocardiogram signals by deep learning. *Neurocomputing 404*. 122–136. https://doi.org/10.1016/j.neucom.2020.04.076 (2020).
12. Li, W. et al. SLC-GAN: an automated myocardial infarction detection model based on generative adversarial networks and convolutional neural networks with single-lead electrocardiogram synthesis. *Inf Sci. 589*. 738–750. https://doi.org/10.1016/j.ins.2021.12.083 (2022).

13. Seo, H. C. et al. Multiple electrocardiogram generator with single-lead electrocardiogram. *Comput Methods Programs Biomed. 221* . **106858** https://doi.org/10.1016/j.cmpb.2022.106858 (2022).
14. Wang, Z., Stavrakis, S. & Yao, B. Hierarchical deep learning with generative adversarial network for automatic cardiac diagnosis from ECG signals. *Comput. Biol. Med.* **155**, 106641. https://doi.org/10.1016/j.compbiomed.2023.106641 (2023).
15. Zhang, C. et al. A 12-lead ECG correlation network model exploring the inter-lead relationships. *Europhys Lett. 140*. **31001** https://doi.org/10.1209/0295-5075/ac9b89 (2022).
16. Chen, M. et al. Deep feature learning for medical image analysis with convolutional autoencoder neural network. *IEEE Trans. Big Data*. **7**, 750–758. https://doi.org/10.1109/TBDATA.2017.2717439 (2021).
17. Antczak, K. A generative adversarial approach to ECG synthesis and denoising. Preprint at https://arxiv.org/abs/2009.02700 (2020). https://doi.org/10.48550/arXiv.2009.02700
18. Shaik, J. & Bhavanam, S. N. Arrhythmia detection using ECG-based classification with prioritized feature subset vector-associated generative adversarial network. *SN Comput. Sci. 4*. **519** https://doi.org/10.1007/s42979-023-01970-3 (2023).
19. Kuntalp, M. & Düzyel, O. A new method for GAN-based data augmentation for classes with distinct clusters. *Expert Syst. Appl. 235* . **121199** https://doi.org/10.1016/j.eswa.2023.121199 (2024).
20. Bagga, M., Jeon, H. & Issokson, A. ECGNet: a generative adversarial network (GAN) approach to the synthesis of 12-lead ECG signals from single lead inputs. Preprint at https://arxiv.org/abs/2310.03753 (2023). https://doi.org/10.48550/arXiv.2310.03753
21. Hou, Y. et al. An ECG denoising method based on adversarial denoising convolutional neural network. *Biomed. Signal. Process. Control.* **84**, 104964. https://doi.org/10.1016/j.bspc.2023.104964 (2023).
22. Wang, H. et al. Deep convolutional generative adversarial network with LSTM for ECG denoising. *Comput. Math. Methods Med.* **2023** (6737102). https://doi.org/10.1155/2023/6737102 (2023).
23. Eriksson, A., Schön, T. & Ribeiro, A. H. Transferability and adversarial training in automatic classification of the electrocardiogram with deep learning. In Proc. Computing in Cardiology (2024). https://doi.org/10.22489/CinC.2024.096
24. Wang, N. et al. Adversarial Spatiotemporal contrastive learning for electrocardiogram signals. *IEEE Trans. Neural Netw. Learn. Syst.* **35**, 13845–13859. https://doi.org/10.1109/TNNLS.2023.3272153 (2024).
25. Zhang, Y. A better autoencoder for image: convolutional autoencoder. *ArXiv Abs*. **1803**, 00045 (2018).
26. Donahue, C., McAuley, J. & Puckette, M. Adversarial audio synthesis. In *International Conference on Learning Representations* (2018). https://doi.org/10.48550/arXiv.1802.04208
27. Radford, A., Metz, L. & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *ArXiv Abs*. https://doi.org/10.48550/arXiv.1511.06434 (2015). /1511.06434.
28. Drysdale, J., Tomczak, M. & Hockman, J. Adversarial synthesis of drum sounds. In *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx)* 167–172 (2020).) 167–172 (2020). (2020).
29. Li, P., Pei, Y. & Li, J. A comprehensive survey on design and application of autoencoder in deep learning. *Appl Soft Comput. 138*. **110176** https://doi.org/10.1016/j.asoc.2023.110176 (2023).
30. Li, Y., Angelov, P. & Suri, N. Rethinking self-supervised learning for cross-domain adversarial sample recovery. *ArXiv Abs* (2024). /2401.12345.
31. Ribeiro, A. H. et al. CODE-15%: a large scale annotated dataset of 12-lead ECGs. *Zenodo* (2021). https://doi.org/10.5281/zenodo.4916206
32. Khushi, M. et al. A comparative performance analysis of data resampling methods on imbalance medical data. *IEEE Access 9*. 109960–109975. https://doi.org/10.1109/ACCESS.2021.3102397 (2021).
33. Radford, A., Metz, L. & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. Preprint at https://arxiv.org/abs/1511.06434 (2016). https://doi.org/10.48550/arXiv.1511.06434
34. Wang, H., Luo, Z., Yip, J. W. L., Ye, C. & Zhang, M. ECGGAN: A framework for effective and interpretable electrocardiogram anomaly detection. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining 5071–5081 (2023). https://doi.org/10.1145/3580305.3599812
35. Arjovsky, M., Chintala, S., Bottou, L. & Wasserstein, G. A. N. Preprint at https://arxiv.org/abs/1701.07875 (2017). https://doi.org/10.48550/arXiv.1701.07875
36. Wagner, P., Strodthoff, N., Bousseljot, R., Samek, W. & Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). PhysioNet (2022). https://doi.org/10.13026/kfzx-aw45
37. Wagner, P. et al. PTB-XL: A large publicly available ECG dataset. *Sci. Data* **7**, 154 https://doi.org/10.1038/s41597-020-0495-6. (2020).
38. Goldberger, A. et al. PhysioToolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215 (2000).

Acknowledgements

The work described in this paper was fully supported by a grant from Hong Kong Metropolitan University (RIF/2021/05).

Author contributions

J.L. designed and implemented the experiments, analysed the data, and drafted the manuscript. K.T.C. conceived the study, supervised the project, and revised the manuscript. L.K.L. assisted in data collection and analysis. K.L. contributed to the interpretation of the results and manuscript revision. A.Y. provided clinical expertise and supervised medical aspects. E.K.S.F. provided project oversight and contributed to manuscript editing. All authors reviewed and approved the manuscript for submission.

Funding

The work described in this paper was fully supported by a grant from Hong Kong Metropolitan University (RIF/2021/05).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.T.C.