공학박사 학위논문

딥러닝을 활용한 심혈관계 디지털 헬스케어 연구

Deep Learning approaches for digital healthcare in cardiology

울 산 대 학 교 대 학 원

의 과 학 과

서 효 창

Deep Learning approaches for digital healthcare in cardiology

지 도 교 수 주 세 경

이 논문을 공학박사학위 논문으로 제출함

2022년 8월

울 산 대 학 교 대 학 원

의 과 학 과

서 효 창

서효창의 공학박사학위 논문을 인준함

심사위원 남 기 병 인

심사위원 주 세 경 인

심사위원 김 영 학 인

심사위원 진 은 선 인

심사위원 신 수 용 인

울 산 대 학 교 대 학 원

2022년 8월

Abstract

본 연구는 인공지능을 활용한 다양한 디지털 헬스케어 환경에서의 애플리케이션들을 개발 및 적용한다. 생체신호는 디지털 헬스케어 환경에서 특히 중요한 부분을 자지하고 있으며 실시간으로 오랜 시간 감시가 가능한 장점에 반해 비정형적, 비표준적, 불규칙성 등을 가지며 노이즈에 매우 취약하여 지금까지 주목받지 못했으나 최근 컴퓨팅파워와 인공지능 분야의 연구가 활발해지며 장시간 데이터를 인력소모 없이 적용이 가능해지며 주목을 받고 있다. 그러므로, 인공지능을 기반으로 생체신호를 활용한 다양한 애플리케이션을 개발하였으며 이를 소개하고자 한다.

첫째로, 딥러닝은 머신러닝 중 하나의 분야로 인공신경망을 활용한 모델링을 통해 학습하는 모델을 말한다. 최근 의료분야에서도 활발하게 활용되고 있으며, 떠오르는 문제로는 데이터 편향문제가 있다. 일반적인 데이터는 쉽게 구할 수 있으며, 데이터 증강(augmentation)이 가능하지만 의료데이터에서는 신뢰성을 잃을 수 있다. 이러한 문제로 다양한 데이터베이스의 자료를 통합하여 사용하는 메타분석(meta analysis)이 효과적이지만 각 데이터베이스에 따른 측정장비의 AC resolution과 sampling rate이 달라 같은 측정방식에도 불구하고 데이터의 형태가 약간씩 달라진다. 이러한 상황에서 동일한 형태의 데이터로 구성해 준다면 과연 데이터베이스에 따른 편향성이 사라지는지 확인해보았다. ResNet 모델을 활용하여 3개의 공공데이터베이스에서 수집된 심전도를 활용하여 AF와 non-AF로 각각의 데이터베이스별로 학습하였을 경우 다른 데이터베이스에서 테스트를 하였다. 결과는 자신의 데이터베이스 외 다른 데이터베이스에서는 정확도가 떨어지는 것을 확인하였으며 이를 통해 학습데이터에 따른 편향이 존재함을 입증하였다.

다음으로, 혈압은 주요한 생체신호 지표 중 하나이다. 비침습적으로 연속혈압을 측정하기 위한 rule base기반의 다양한 시도들이 있어왔다. ECG와 PPG를 기반으로 혈압을 추정

하는 맥파전달시간(pulse transit time, PTT)과 맥파도달시간(pulse arrival time, PAT)이 대표적이다. 하지만 정확한 혈압추정에는 도달하지 못하였으며 최근 딥러닝을 활용한 시도가 있다. 많은 PAT와 PTT연구와 같이 ECG와 PPG는 혈압과의 상관관계를 딥러닝을 활용하여 추정해보았다. 결과적으로 연속혈압추정이 가능함을 입증하였고 국제표준의 성능을 확인하였다. 또한 심방세동 환자를 대상으로 심방세동 중에도 혈압추정이 가능함을 통해 질환이 있는 환자에서도 추정이 가능함을 확인하였다.

마지막으로, ECG는 벡터 값이므로 하나의 전극으로부터 얻은 심전도를 활용하여 다른 전극의 심전도를 추정할 수 있을 것이라는 가설을 입증하고자 하였다. 적대적 생성신경망(Generative Adversarial Networks, GAN)을 활용하여 단일전극 lead I을 입력하여 나머지 다중전극 11개를 생생하는 딥러닝 모델을 학습하여 임상현장에서 활용이 가능한지 검증하였다. 개발된 모델을 활용하여 precordial ECG 기반의 질환을 선정하여 임상의가 생성된 ECG를 임상현장에서 활용가능 여부를 판단하였다. 그 결과 단일전극으로 생성한 다중전극 심전도를 활용하기에 충분함을 입증하였으며 디지털 헬스케어로써 웨어러블 시스템에 탑재하여 활용 가능할 것으로 기대한다.

다양한 심장관련 디지털 헬스케어 연구들을 통해 딥러닝 기반의 애플리케이션들이 임상현장에서 충분히 활용 가능함을 확인하였으며 추후 더 많은 데이터들을 활용하여 유효성을 입증한다면 의료기기로써 자리매김할 수 있을 것으로 기대한다.

Contents

Figures

Tables

Introduction

Digital healthcare is a multidisciplinary concept that incorporates ideas from an intersection between technology and healthcare. Among other technologies, digital healthcare includes mobile health (mHealth) applications, electronic health records (EHRs), electronic medical records (EMRs), wearable devices, telehealth, telemedicine, personalized medicine, and electronic intensive care units (eICUs) [1].

Healthcare data are classified into image, text, and signal data. [2] These data are usually atypical, non-standardized, and irregular. Here, we focus on signal data. In the medical field, biosignals are used to monitor a patient's condition and inform decision making. A biosignal is defined as any signal in humans that can be measured and monitored. Biosignals include biopotential signals, such as electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG), and electrooculography (EOG); non-electrical signals, such as heart sound, photoplethysmogram (PPG), and saturation of partial pressure oxygen (SpO_2); as well as other signals, such as gait cycle, gait speed, stride, and ballistocardiogram (BCG). Table 1 shows biosignal types.

Biopotentials (wave)	**Vital signs (numeric)**	**Other biosignals**
Electroencephalogram, EEG	Body temperature, BT	Photoplethysmogram, PPG
Electrocardiogram, ECG	Blood pressure, BP	Saturation of partial pressure oxygen, SpO_2
Electromyogram, EMG	Pulse or heart rate, HR	Ballistocardiogram, BCG
Electrooculogram, EOG	Respiratory rate, RR	Gait cycle
Magnetoencephalogram, MEG		Gait speed
Heart sound		stride

Table 1. Type of biosignal data

To analyze biosignals, this study uses machine learning (ML) techniques, particularly deep learning (DL), a subfield of ML, rather than rule-based methods, which are more commonly used in medical research. In the artificial intelligence (AI) field, rule-based systems rely on the proficiency of experts in addition to an inference engine to deduce new information, whereas DL methods learn directly from raw data, extract features, and adapt without explicit instructions (Figure 1). DL techniques have an advantage in that they can detect patterns not observed by humans. However, the lack of interpretability of many ML and DL models is a major limitation, the so-called black box problem. In addition, unlike humans, current DL models require manual work for cleaning and normalizing the input data (data preprocessing).

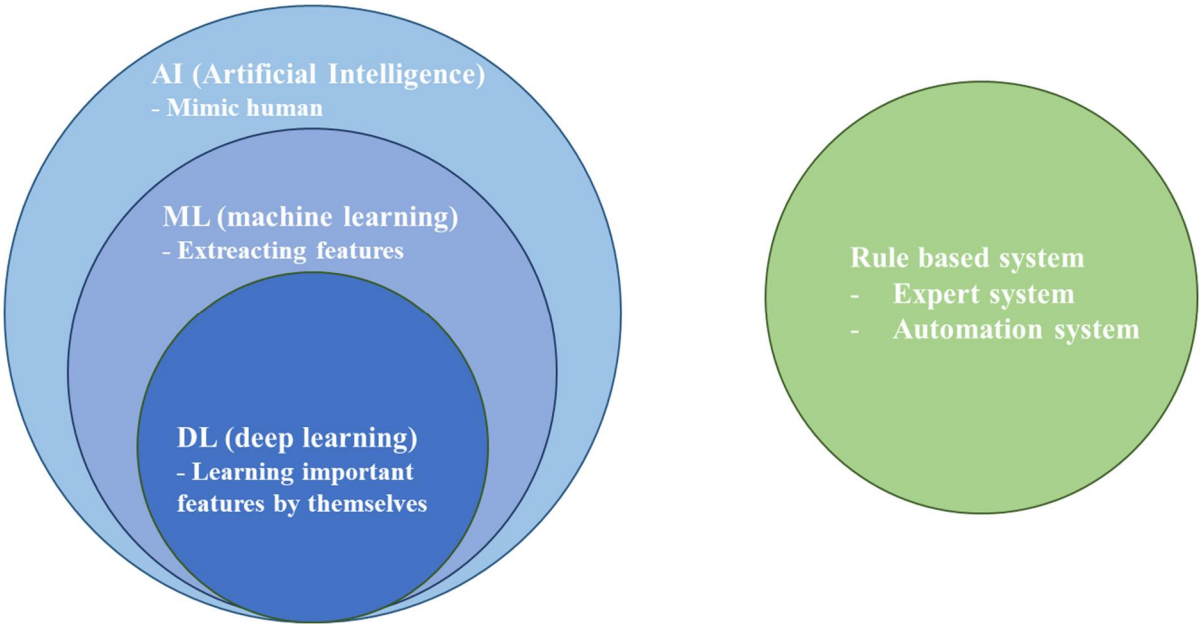


Figure 1. Schematic diagrams of AI tools: Machine learning and deep learning vs. rule-based systems

In this study, first, we address the problem of medical data dependency caused by the differences in AD resolution and sampling rate of the measuring equipment. Specifically, we develop a DL-based model for AF detection (binary classification into AF and non-AF patients) using ambulatory (Holter) ECG data obtained from three datasets. Second, we develop a DL-based model for real-time indirect biosignal estimation. Finally, we construct a generative adversarial network (GAN) model to synthesize multi-lead ECG from single-lead ECGs, and verify its clinical applicability for diagnosis.

CHAPTER 1. ECG data dependency for atrial fibrillation detection based on residual networks

Atrial fibrillation (AF) is the most common type of cardiac arrhythmia. It occurs when disordered or rapid electrical signals cause the atria to contract too quickly and chaotically. AF is associated with five-fold increase in stroke risk and two-fold increased risk of death. Estimates of the prevalence of AF in the United States range from approximately 2.7 to 6.1 million and the number is expected to rise to 12.1 million in 2030 [3, 4]. Worldwide, the estimated number of individuals with AF in 2010 was 33.5 million (0.5% of the world's population). Particularly, approximately 9% of people aged 65 and older had AF, as compared to only 2% of people younger than 65 years [5]. In this study, we developed a DL model for AF detection in patients (binary classification into AF and non-AF patients) using ambulatory (Holter) ECG data obtained from three long-term databases.

Data acquisition is an important issue in medical research using DL techniques. That is because it is not easily acquired and a huge cost can be incurred for data collection [6]. Additionally, if data augmentation is applied to reduce cost, there is a limit to the reliability of data. Therefore, another method is to collect data from different sources, such as hospitals and/or equipment, as metadata. However, when data from various public databases are collected and used, we need to ensure that the DL model is not biased by learning the characteristics of each dataset. Therefore, in our study, we test the data dependency of the DL model for AF detection using three datasets [7].

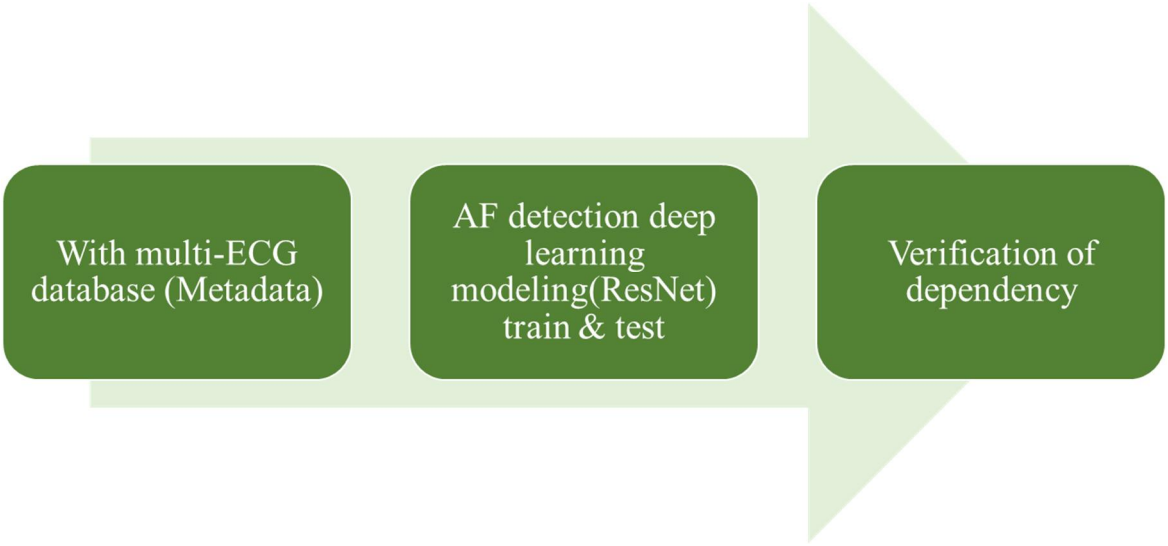


Figure 2. Verifying ECG data dependency for atrial fibrillation detection using residual networks

In this study, we used a residual neural network (ResNet) model, developed by Kaiming He, which won the 2015 ImageNet Challenge [8]. It attempted to improve learning of deep plain neural networks by overcoming the vanishing gradient problem in models with many layers. Specifically, it has skip connections that act as gradient superhighways, allowing gradients to propagate to deep layers before they are attenuated to small or zero values. . For comparison, the architectures of VGG, plain, and residual neural networks are shown in Figure 3. In addition, Figure 4 compares the performance of different plain neural networks and ResNet architectures with the increase in layers. From the figure, we find that the accuracy of plain neural networks decreases as the number of layers increases. Conversely, the performance of ResNet architectures is improved even in networks more than 100 layers. Thus, we can ensure that ResNet architectures overcome the gradient vanishing problem.

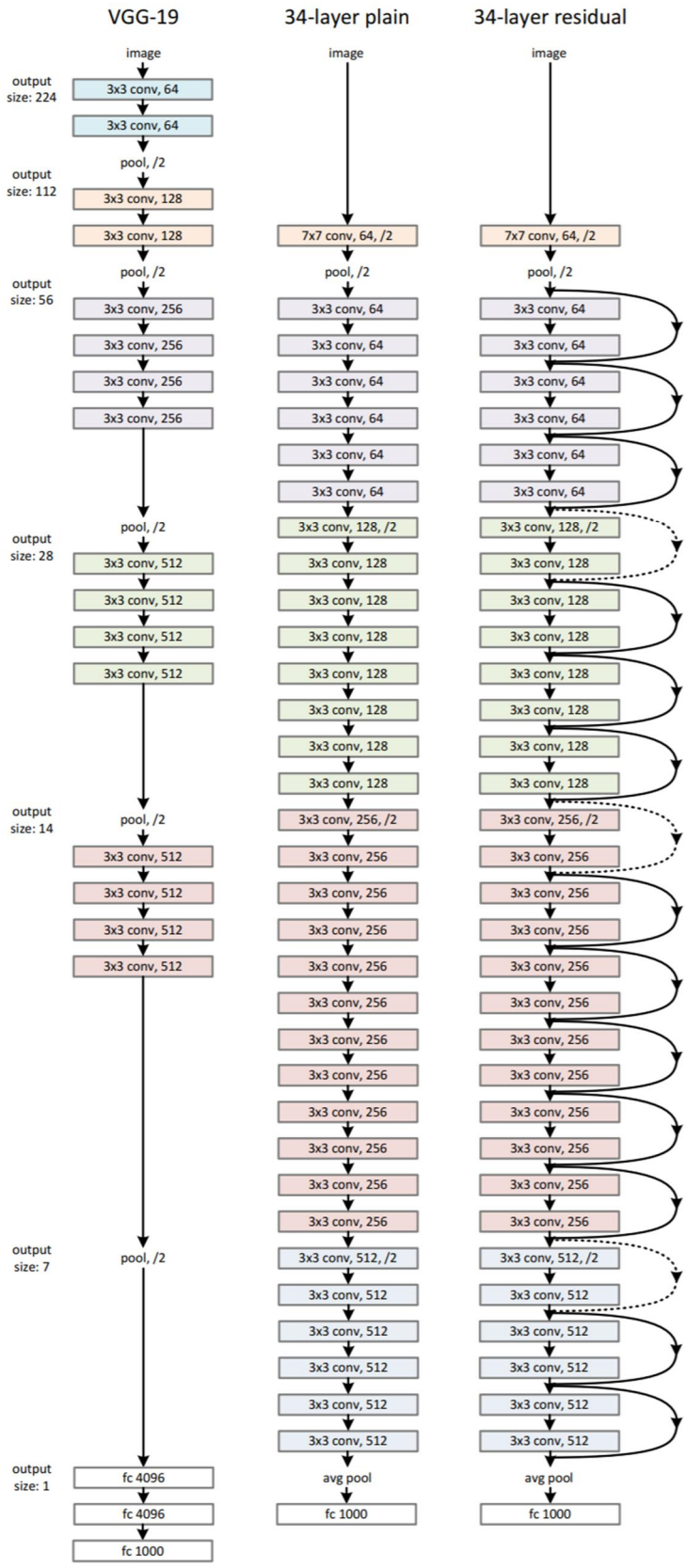


Figure 3. Architectures of VGG-19, 34-layer plain, and 34-layer residual networks [8]

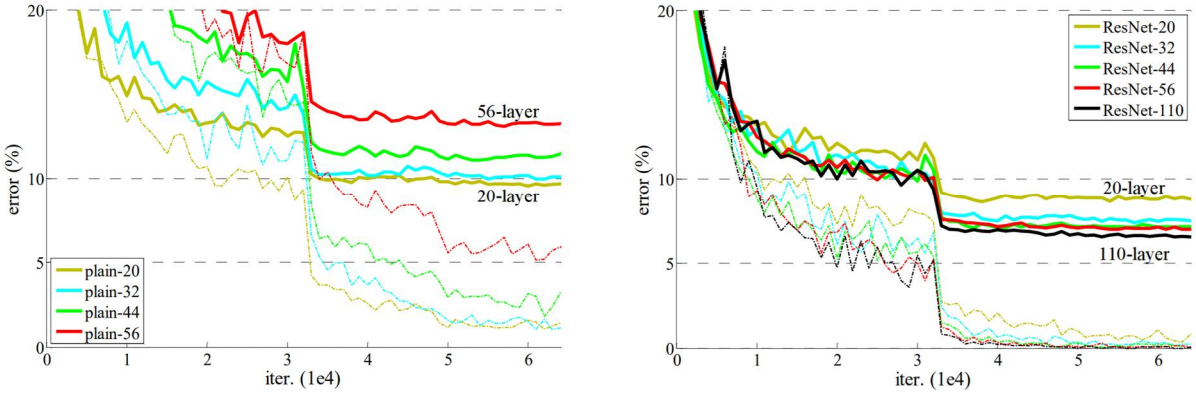


Figure 4. Performance of plain neural network (left) and residual network (right) with respect to the number of layers: Dashed lines denote training error, and bold lines denote testing error. [8]

In this study, we used a ResNet model for AF detection using data from the long-term Atrial Fibrillation database (LTAFDB) [9], MIT-BIH Atrial Fibrillation database (AFDB) [10], and MIT-BIH Arrhythmia database (MITDB) [11]. Table 2 shows the characteristics of each database with different measurement periods, sampling rates, and analog to digital (AD) resolution.

	LTAFDB	AFDB	MITDB
No. of records	84 records	25 records	48 records
Channel	2 Ch	2 Ch	2 Ch
Duration	24 to 25 hours	10 hours	Half-hour
Sampling rate	128 Hz	250 Hz	360 Hz
Resolution	12 bit	12 bit	11 bit
Voltage range	20 mV	±10 mV	10 mV
Acquisition Location	Not reported	Boston's Beth Israel Hospital	Boston's Beth Israel Hospital

Table 2. Characteristic of the three databases: Long-Term Atrial Fibrillation database(LTAFDB), MIT-BIH Atrial Fibrillation database (AFDB), MIT-BIH Arrhythmia database (MITDB)

To obtain our training data, the raw ECG data of all databases were cut in units of 10 seconds, as shown in Table 3 Next, we downsampled all data to 128 Hz to ensure the same data format. Additionally, we used z-score normalization. For our model, one ECG database was used for training and the other two databases were used for testing. Figure 5 shows an overview of the binary classification process.

	Training dataset		Test dataset		Total dataset	
	Non-AF	**AF**	**Non-AF**	**AF**	**Non-AF**	**AF**
LTAFDB	460,740	588,990	115,615	146,816	576,355	735,806
AFDB	79,960	53,535	19,944	13,428	99,904	66,963
MITDB	9,696	1,213	2,435	291	12,131	1,504

Table 3. Number of cropped 10-second ECG units

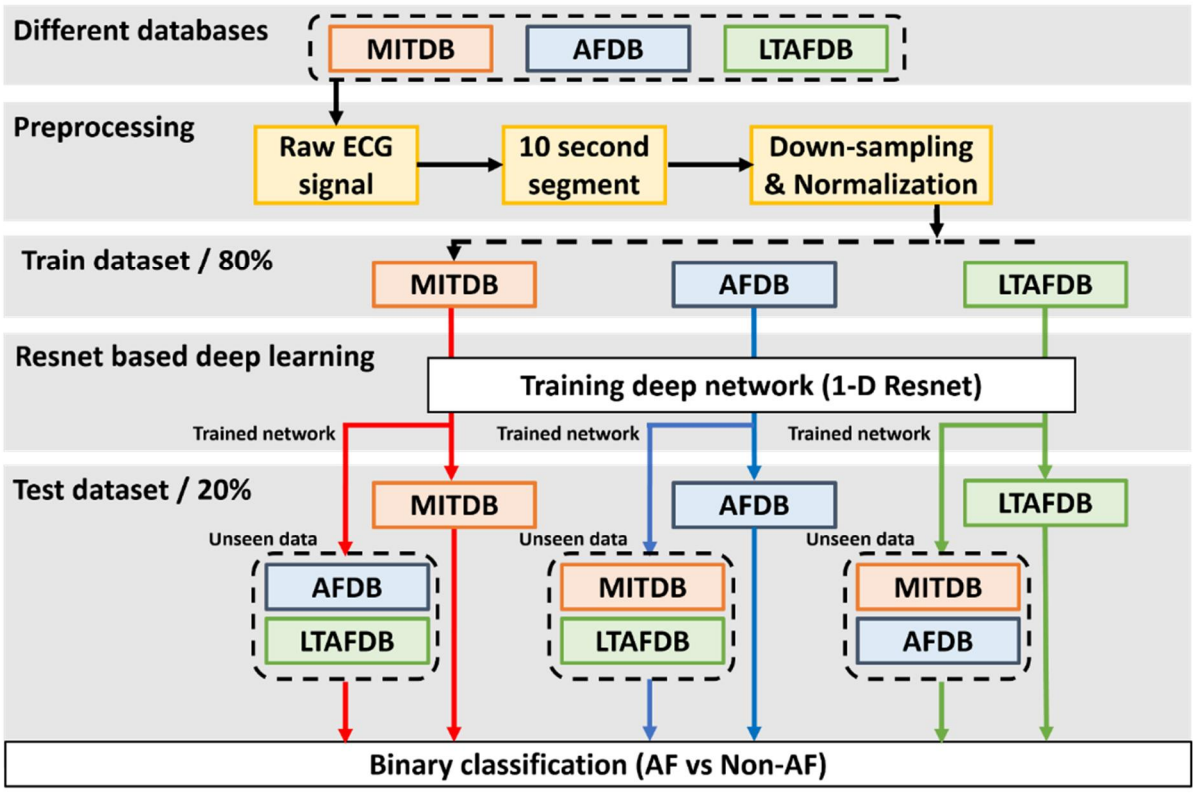


Figure 5. Overview of the binary classification process

In this study, we employed different ResNet architectures (i.e., ResNet 18, ResNet 34, ResNet 50, and ResNet 152) and evaluated their performance. In all models, two types of residual blocks were used. When the previous and present layers had the same dimensions, residual block type 1 was applied, whereas when the previous and present layers had different dimensions, residual block type 2 was applied. The architecture of ResNet 50 is shown in Figure 6. The rest of the models are similarly configured except for the number of layers.

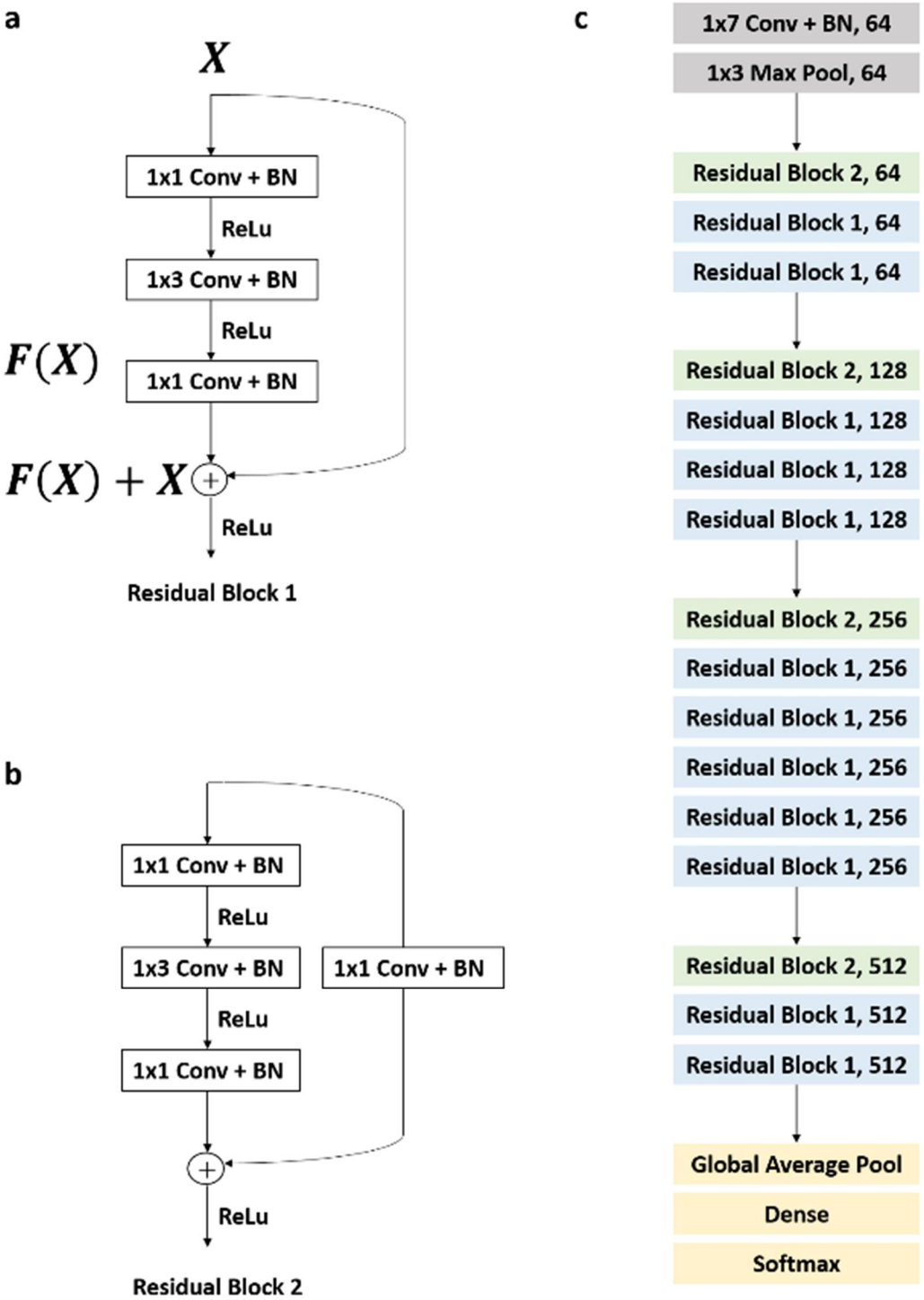


Figure 6. ResNet 50 architecture

Results for the trained models are shown in Table 4. In general, the accuracy of the trained models on MITDB was the lowest, as compared to their performance on other databases due to the small size of the MITDB database. As shown in Table 3, the total number of instances in MITDB was 13,635 (non-AF = 12,131, AF = 1,504), which was very small. Additionally, AF cases were approximately 8 times higher than non-AF instances, indicating a severe data imbalance. Meanwhile LTAFDB contained a total of 1,312,161 segment (non-AF = 735,806, AF = 576,355) and the ratio of AF to non-AF cases was approximately 0.78, indicating the suitability of the dataset for training. The AFDB dataset had 166,867 segment (non-AF=99,904, AF=66,963) which was less than that of LTAFDB, but higher than MITDB. Additionally, the AFDB dataset had a ratio of non-AF to AF of 1.49, indicating that the dataset was not biased and thus, it was adequate for training. The confusion matrix for the performance of each model on each test dataset is shown in Figure 7, and the receiver operating characteristic (ROC) curve is shown in Figure 8.

Trained model		Test datasets		
Model	**Training dataset**	**LTAFDB**	**AFDB**	**MITDB**
ResNet 18	LTAFDB	99.04	92.03	82.10
	AFDB	85.01	99.41	86.68
	MITDB	74.48	70.98	99.89
ResNet 34	LTAFDB	98.70	92.19	81.14
	AFDB	84.94	99.27	78.14
	MITDB	63.65	65.55	99.78
ResNet 50	LTAFDB	98.66	92.13	83.90
	AFDB	84.35	99.20	80.48
	MITDB	68.79	65.61	99.82
ResNet 152	LTAFDB	98.53	92.00	84.67
	AFDB	83.87	99.21	81.14
	MITDB	66.23	64.41	99.56

Table 4. Accuracy of each model on each test dataset

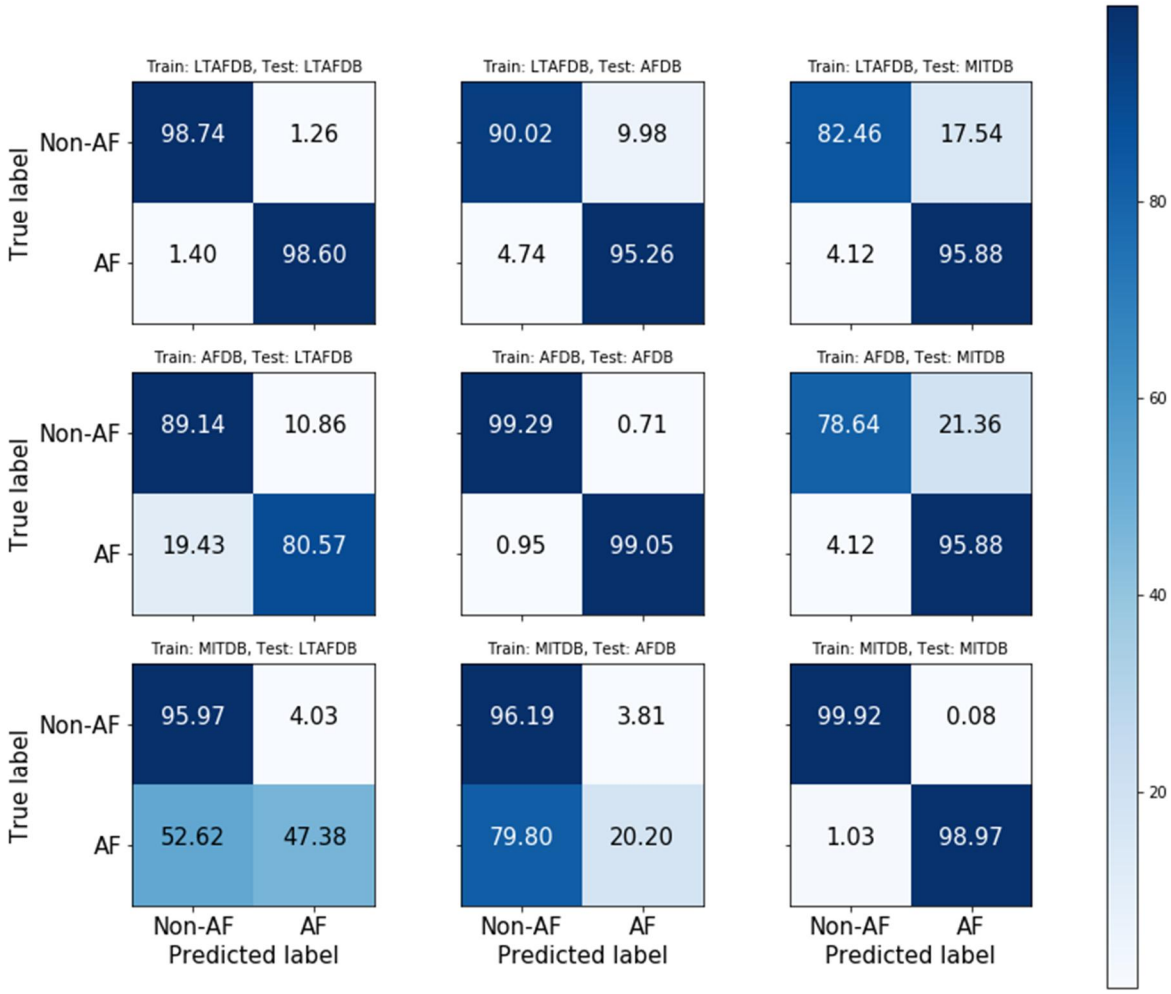


Figure 7. Confusion matrix for the performance of each model on each test dataset

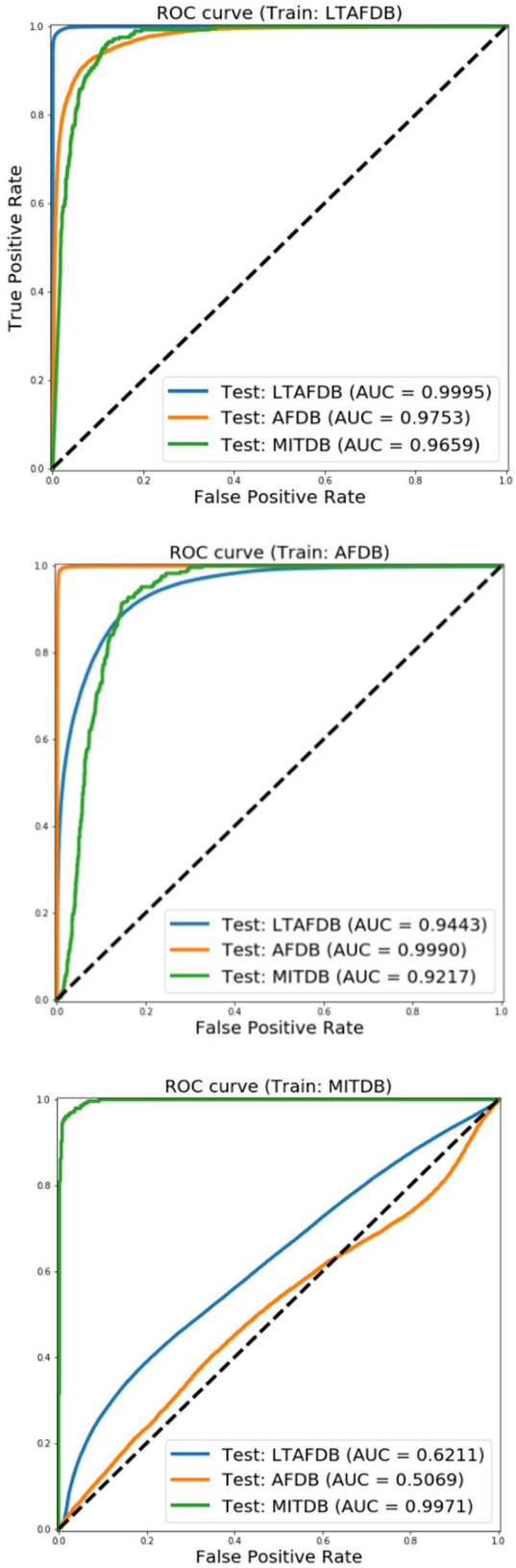


Figure 8. ROC curve for each dataset

We experimentally investigated the data dependency of DL-based AF classification using ResNet architectures and raw ECG signals. As indicated in Table 4, the highest accuracy for each trained model (ResNet 18, 34, 50, and 152) was obtained by evaluating the models on the test dataset extracted from the same source that contained the training dataset. In contrast, the model accuracy decreased when they were evaluated on different datasets, other than the one containing the training dataset. In general, all ResNet architectures showed good performance without the problems of exploding and vanishing gradients, even in networks with many layers. However, the data dependency occurred regardless of the depth of the ResNet architecture. Therefore, a deeper network could not resolve the data dependency problem. On unseen data, when the true positive rate increased, the false positive rate tended to rise as well. Further, the true negative and false negative rates showed the same trend. Unlike the evaluation results on own database, if the trained model showed a high sensitivity for external data, specificity was low. Similarly, high specificity for external data led to low sensitivity in trained models. These results imply that the predictions of the trained models on external data may be biased.

Although data dependency is a widely known phenomenon in deep learning, especially in medical signal processing, this study verifies this phenomenon for the first time for 1-D signals, such as biosignals. Additionally, although we employed ResNet architectures in our study, which is the most widely used ECG classification algorithm, we could not overcome the data dependency problem. Further, we used various normalization methods; however, the dependency problem could not be completely resolved. For example, all trained ResNet 50 models trained on LTAFDB, AFDB, and MITDB showed

specificity of more than 95% and a false positive rate of approximately 2–4% when evaluated on external data extracted from NSRDB

The data dependency problem in DL models can be caused by several aspects. Data imbalance is one of the most common causes. If a database has AF events far less than normal rhythms, which is common in medical databases, the trained model can be biased, resulting in reduction in performance when tested on external datasets. Another problem is the noise in ECG signals caused by motion artifacts or other reasons. The physical movement of patients when measuring ECG signals can cause wandering of the ECG baseline or unwanted noises. These noises can be minimized by digital filtering or other signal processing techniques. However, the original ECG signal could be distorted or lose its characteristic waveform due to the use of preprocessing methods, resulting in lower performance of DL models. Another problem is the discrepancy in the measurement values. As there are several companies making devices for measuring ECG signals, these devices have different hardware settings (e.g., amplifier configuration, filters, and gain) and software settings (e.g., sampling frequency and resolution). However, waveforms from approved ECG measurement devices do not differ largely. Additionally, resampling and normalization techniques can reduce such problems; however, they cannot eliminate them entirely.

It can be concluded that it is necessary to validate the DL-based AF detection algorithm using various external databases to overcome the problem of data dependency.

CHAPTER 2. Real-time indirect continuous arterial blood pressure measurements from ECG and PPG waveforms using deep learning

Blood pressure (BP) is defined as the force applied by circulating blood against any unit area of blood vessels as it passes through them. It is one of the major vital medical signs. With regard to hemodynamics, accurate measurements of BP can enable optimal intervention in unstable patients and play an important role in determining long-term perfusion. An abnormal BP (e.g., high or low BP) constitutes a major risk factor for cardiovascular disease (CVD), leading to heart attacks and can even cause blindness or cerebral apoplexy. [12, 13]. The range of blood pressure is divided into normal, low, and high blood pressure. The main indicators are the mean arterial pressure (MAP), systolic BP (SBP) and diastolic BP (DBP). MAP is calculated by SBP and DBP and is used as an indicator to check the trend of blood pressure. (eq. 1)

MAP = DBP + 1/3 Pulse pressure (eq. 1)

BP measurement methods are classified into non-invasive blood pressure (NIBP) and invasive blood pressure (IBP) methods. NIBP measurement methods can be used without pain, whereas IBP methods involve pain, infection, inflammation, and bleeding, but it can be measured continuously [14]. Therefore, NIBP methods are usually used in wards or for health checkups, whereas IBP methods are used in critical situations such as ICU environments. There are two types of NIBP measurement methods: Korotkoff sound-

based and oscillometry-based.

The two methods are divided depending on whether the measurement is performed by sound or the intensity of pressure. The Korotkoff sound-based BP measurement is usually used a ground truth, while the oscillometry-based method is mainly used in the case of an automatic blood pressure monitor.

However, owing to the need of a cuff, the two methods are limited as measurement is difficult if there is a problem in the arm; moreover, it is impossible to undertake continuous measurement. Therefore, for decades, numerous studies have been performed to measure blood pressure using biosignals [15-19]. Biosignal-based methods used the pulse arrival time (PAT) and pulse transit time (PTT) for BP measurement. The two parameters of ECG and PPG signals are used to construct an equation for BP estimation by utilizing the difference between the time the heart contracts and the time the pulsation of the contracted BP reaches the fingertips and toes. However, although it is possible to estimate the level of the trend of the actual SBP, it is difficult to estimate the absolute value. In particular, there is a limitation to applying it to all people because of the possibility of generating a few errors while comparing the normal and patient groups. In this study, we developed a DL model for real-time indirect continuous arterial BP measurements using ECG and PPG waveforms.

We obtained our data from the Medical Intensive Care Unit (MICU) of the Seoul Asan Medical Center Hospital for 980 patients between April 01, 2018 and May 31, 2019. The raw ECG, PPG, and ABP data were collected from a patient monitor (GE B650, General Electric, USA) in MICU. This study was approved by the Institutional Review Board of the Seoul Asan Medical Center Hospital (IRB No. 2021-0833). To collect stable states of ECG

and PPG data for the 980 patients, we collected data on the state of the patient 30 minutes and up to 3 hours after entering the ICU per patient. Missing or corrupted data caused by motion artifacts were excluded from our dataset. Additionally, patients with SBP of less than 50 or more than 200 mm Hg were excluded, as abnormal. Moreover, cases where ICU hospitalization time was less than 30 minutes were excluded. Finally, a total of 284 patients were selected. Information on patient screening is shown in Figure 9. For long-term data, both ECG and PPG data were cropped for 2 seconds, and then the ECG data were downsampled from 240 Hz to 60 Hz. The downsampling was carried out despite the disappearance of electrophysiological features of the data because it was more important to ensure that the information was on the same time axis along the correlation between PPG and ECG signals. In addition, we selected a segment of 2 seconds to ensure that the output data were as close to real-time data as possible, since SBP and DBP are output once every 2 seconds. The details of the data structure are presented in Figure 10.

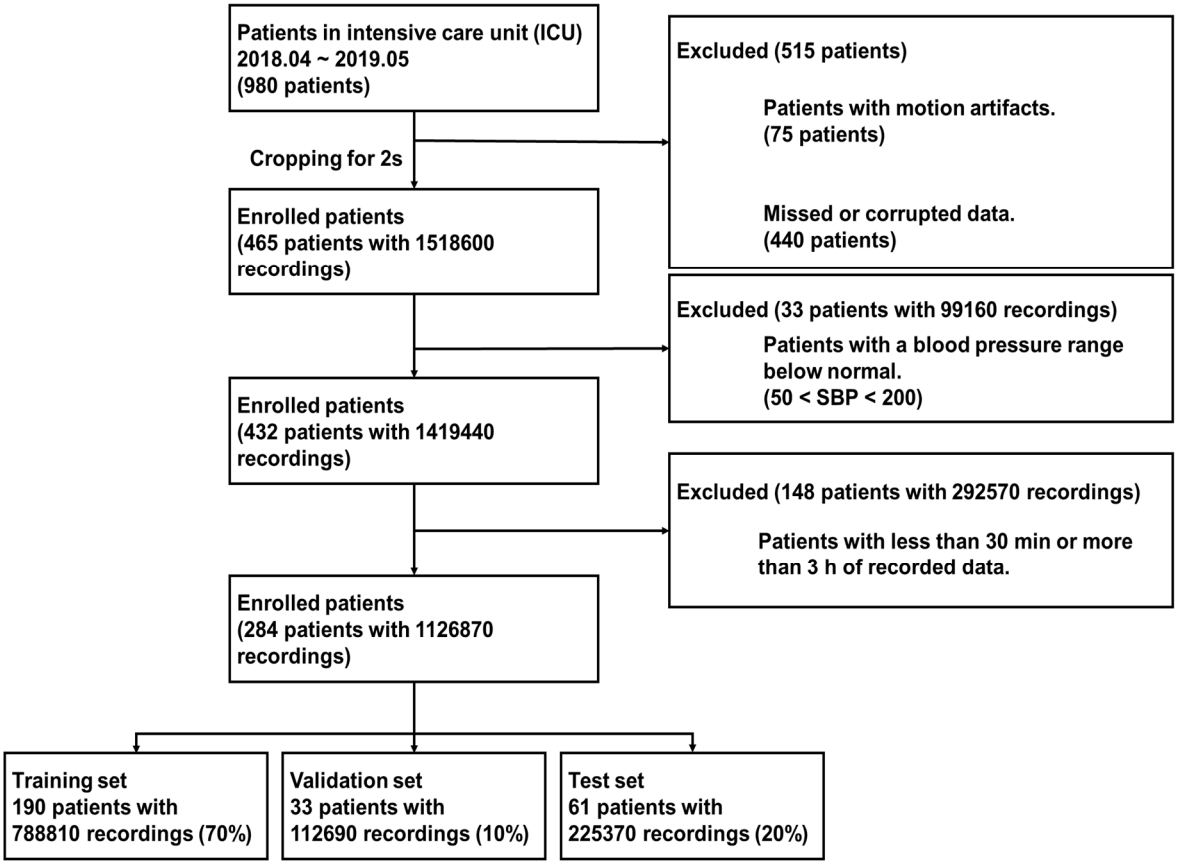


Figure 9. Patient data screening and composition of the training and test datasets

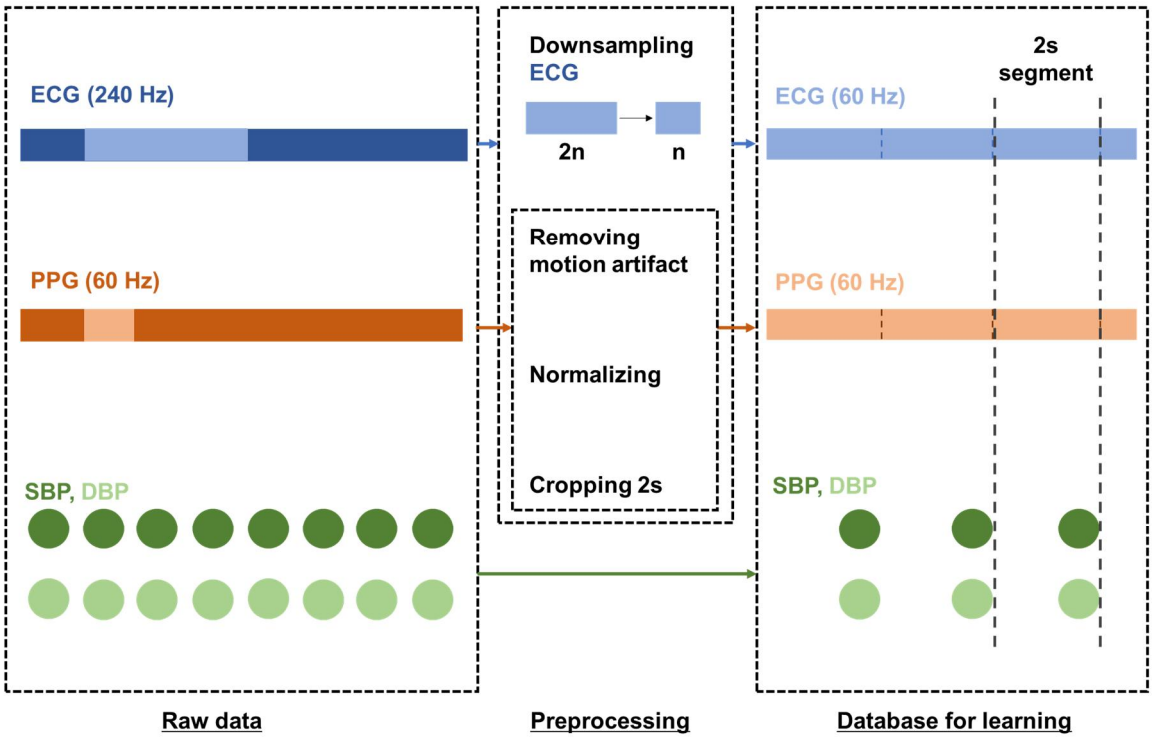


Figure 10. Data preprocessing and configuration for training

The DL model was constructed based on a convolutional recurrent neural network (C-RNN) by combining a ResNet architecture and the long short-term memory (LSTM) model. ECG and PPG were used as the input data, 2 X 60, and SBP and DBP were used for labeling data. We described the ResNet architecture in detail in Chapter 1. Regarding LSTM, as a type of recurrent neural networks (RNNs), it stores sequential features and additionally learns the correlations between previous and current data. Hence, it plays a role of additionally delivering continuity of time-domain data for ECG and PPG waveforms. The architecture of our model is shown in Figure 11.

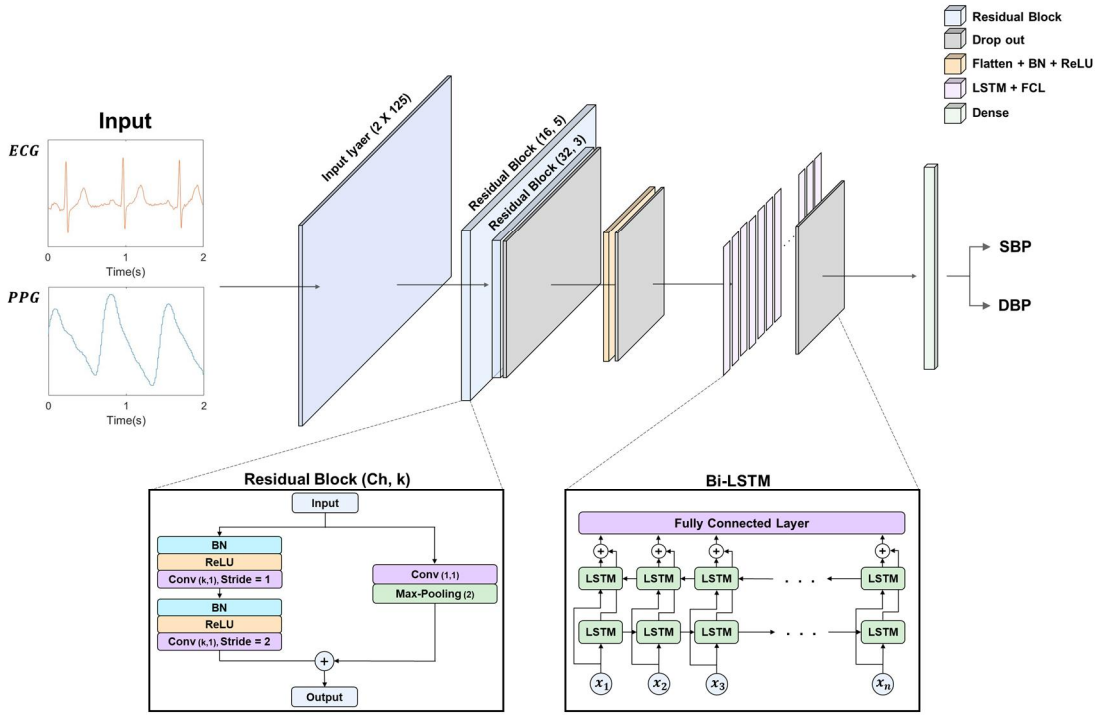


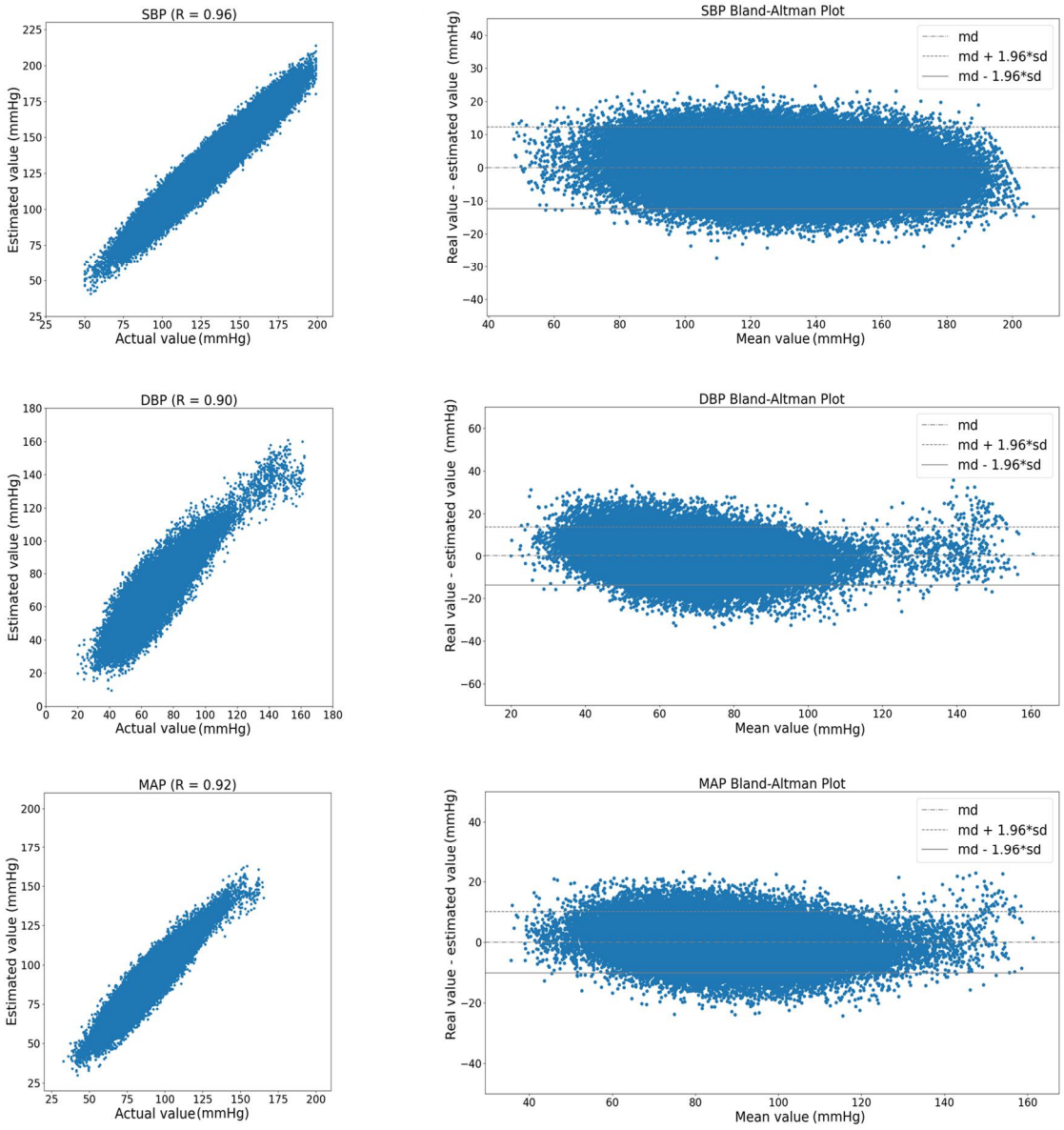
Figure 11. Architecture of the blood pressure estimation model

To evaluate the performance of the proposed blood pressure estimation model, we used the measures of mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). Table 5 shows the performance evaluation results. The model showed a relatively high accuracy for SBP estimation, and less accuracy for DBP estimation.

	SBP	DBP	MAP
Correlation Coefficient, R	0.96	0.90	0.92
MAE (mmHg)	5.09	5.55	4.07
RMSE (mmHg)	6.33	7.07	5.16
MAPE (%)	4.31	7.93	4.54

Table 5. Performance evaluation for the blood pressure estimation model

Figure 12 visually presents correlation plots and Bland–Altman plots for SBP, DBP, and MAP, comparing the ground truth to the estimated results. The dotted line represents the 95% confidence interval. The correlation plots verify the linearity by matching the estimated BP and the ground truth to the X and Y coordinates. Meanwhile, the Bland–Altman plots verify the normal distribution and bias of the model performance by fitting the difference between the estimated BP and the ground truth to the Y axis, and the mean between both to the X-axis. Additionally, as we verified the results only at a specific point in time, we fitted the continuous estimated BP and ground truth to verify whether the BP was estimated as a continuous trend. Hence, we verified the existence of a similar trend, to a certain degree, even in rapidly changing periods (Figure 13).



R: Pearson correlation coefficient

md: mean of the difference; sd: standard deviation of the difference

Figure 12. Correlation plots and Bland–Altman plots for SBP, DBP, and MAP

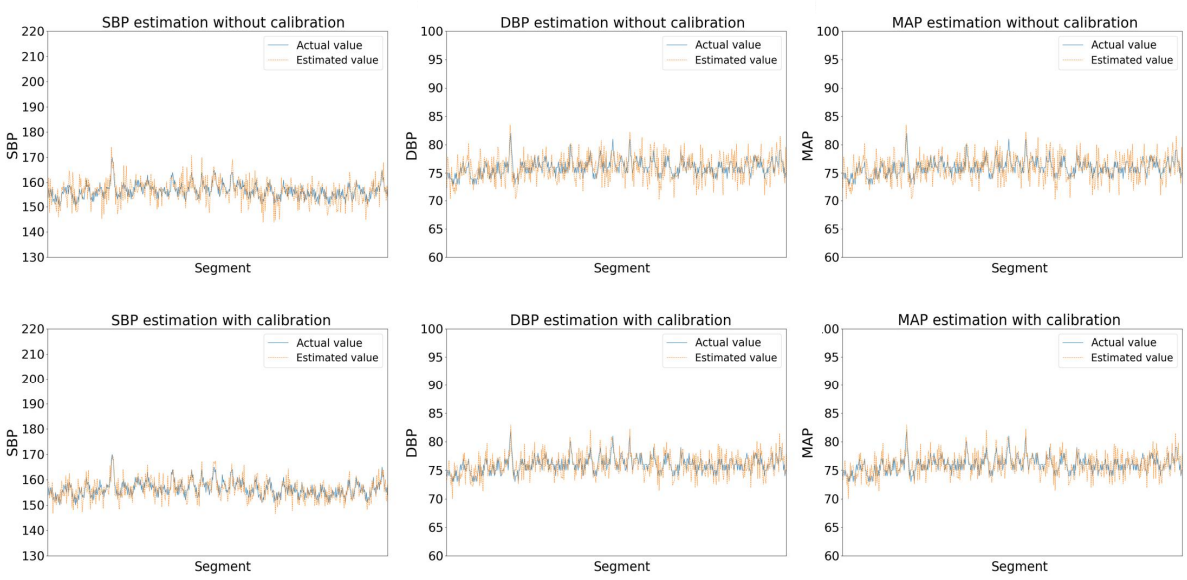


Figure 13. Evaluation of continuous trend for estimated blood pressure and ground truth

Further, we tested whether the proposed BP estimation model adheres to the international standards of BP measurement. First, we confirmed the compliance of the estimation model with the BP monitor standards suggested by the British Hypertension Society (BHS). The standards have grades A, B, C. If a BP monitor has grade B or higher, its performance is confirmed to comply with the BHS standards. The proposed model received B, B, and A grades for SBP, DBP, and MAP, respectively, in accordance with the BHS standards. In addition, the Association for the Advancement of Medical Instrumentation (AAMI) states that the mean difference and standard deviation should be less than ± 5mmHg and ± 8mmHg, respectively. The performance of all BP monitors has to adhere to these standards when the number of subjects is 85 or more. In this study, the proposed model satisfied the aforementioned international standards (both BHS and AAMI) as shown in Tables 6 and 7.

		Absolute difference			Grade
		≤ 5	≤ 10	≤ 15	
Proposed model	SBP	56.5	88.6	98.5	B
	DBP	53.6	85.1	96.4	B
	MAP	67.9	94.7	99.3	A
BHS	Grade A	60%	85%	95%	
	Grade B	50%	75%	90%	
	Grade C	40%	65%	85%	

Table 6. Compliance of the proposed model with standards for blood pressure monitors as suggested by the British Hypertension Society (BHS)

		MD	SD
Proposed	**SBP**	−0.04	6.31
model	**DBP**	0.01	7.07
AAMI	**SBP / DBP**	≤ 5	≤ 8

Table 7. Compliance of the proposed model with sphygmomanometer standards as presented by Association for the Advancement of Medical Instrumentation (AAMI)

Table 8 presents a comparison of the proposed method with previous DL techniques for estimating BP using biosignals. Previous studies performed training based on continuous BP using MAE and SD as evaluation metrics for the BP prediction accuracy.

Miao et al [20]., Eom et al [21]., and Li et al [22]. estimated BP using at least two biosignals and reported performance results comparable to those of the proposed model. Unlike other studies, however, the proposed method can analyze the data output at intervals of 2 seconds in real time. Hence, it is expected to be more useful in urgent situations, such as in clinical surgeries. Other studies have also reported BP monitoring based on NIBP methods using cuffs indirectly and discontinuously. However, in this study, the BP at the arteries was directly measured via IBP monitoring, which was considered as the ground truth. Although the accuracy of the method by Eom et al. was high, ensuring credibility was difficult because the data used were limited to only 15 subjects. Hence, the BP results of our method are more accurate and have greater applicability. In addition, using clinical data instead of public data, as in our study, can help increase reliability, as the training involves more diverse patient groups. Furthermore, the proposed method utilizes shorter learning periods because processes such as feature extraction are not included.

Author	Model	Input data	Dataset	SBP		DBP		MAP	
				MAE	SD	MAE	SD	MAE	SD
Ours	Resnet + Bi-LSTM	Raw ECG, PPG	284 subjects and 23 AF patients	5.06	6.31	5.55	7.07	4.07	5.17
Ours in AF				6.32	8.11	7.21	7.98	5.59	6.94
Fen Miao et al.	Resnet + LSTM	Raw ECG	30 arrhythmia patients	7.10	9.99	4.61	6.29	4.66	6.36
H Eom et al.	CNN + Bi-GRU	Raw ECG, PPG, BCG	15 subjects	4.06	4.04	3.33	3.42	-	-
YH Li et al.	Resnet + Bi-LSTM	Featured ECG, PPG	315 subjects	6.73	14.51	2.52	6.44	-	-

Table 8. Comparison of the performance of the proposed model with that of other selected studies

Moreover, hypertension is likely one of the major causes of AF [24]. Therefore, accurate measurement of BP in AF patients is important, especially, for the elderly, as hypertension with sudden elevations in SBP or SBP exceeding 120 mmHg is associated with an increased risk of AF accidents [25-27]. The ECGs of the AF and normal patients exhibit visual differences in terms of the waveforms. That is because the hearts of AF patients may not produce sufficient BP due to atrial problems. Therefore, we estimated the BP of AF patients using DL models that can learn such features. In this study, the AF dataset was added as part of transfer learning because the probability of AF occurrence in an ICU environment is approximately 5%. That is, our dataset included approximately 14.2 patients with AF from 284 patients in ICU. We added the AF dataset with 23 AF patients. Among them, two patients were excluded because they had motion artifacts or their hospitalization time was less than 30 min. Among the remaining 21 AF patients, the data of 14, 3, and 4 patients were used as the training, validation, and test datasets, respectively, to evaluate the performance of the proposed model with respect to estimating BP in AF patients. The total number of ECG recordings in the test dataset was 18,540. The test results are presented in Fig. 13. We evaluated the estimates of BP in AF patients using existing performance evaluation methods. The MAE and SD values of the SBP, DBP, and MAP were 6.32 mmHg and 8.11 mmHg, 7.21 mmHg and 7.98 mmHg, and 5.59 mmHg and 6.94 mmHg, respectively.

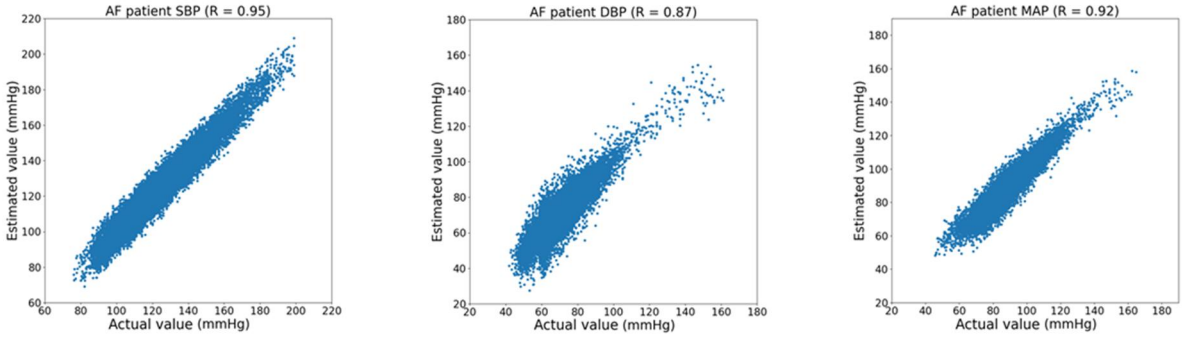


Figure 14. Estimation of BP in AF patients

CHAPTER 3. Multiple Electrocardiogram Generator with Single-Lead Electrocardiogram

An ECG is a test used to measure the rhythm and electrical activity of the heart. The limb-lead and chest-lead methods are the standard ECG measurement methods. In the limb-lead method, electrodes are attached to the left and right arms as well as the legs, and we can observe the formation of the Einthoven's triangle. In the chest-lead method, electrodes are attached to the surface of the chest near the heart. Limb leads are represented as I, II, III, aVR, aVL, and aVF, whereas chest leads are represented as V1, V2, V3, V4, V5, and V6. The heart is a three-dimensional structure; the limb leads measure the frontal plane of the heart, while the chest leads measure the transverse plane of the heart.

An ECG is used to diagnose and monitor the heart condition. The resting ECG, Holter ECG, and treadmill method are the conventional ECG measurement methods. In resting ECG, electrodes are attached to the patient's body, and the ECG signals are measured for 10 seconds while the patient is lying down. A Holter ECG monitor is a portable device used for monitoring cardiac activity. In this method, three to eight electrodes are attached to observe heart activity over long periods to detect arrhythmias. In treadmill ECG, electrodes are attached to the body of the patient, and ECG signals are measured while the patient is exercising to determine the possibility of heart disease. Generally, in conventional methods, the use of multiple electrodes is limited.

Unlike conventional methods, currently, ECG signals can be estimated using a variety of devices such as smart wearable devices. Smartwatches have evolved from being a fashion accessory to a comprehensive health and fitness device. One of the biggest health features of current smartwatches is their ability to measure a single-lead ECG signal. However, although wearable devices are less complex, they provide fewer details owing to single-lead measurement. In contrast, a 12-lead ECG measured in hospitals can obtain more detailed readings of heart signals. Nevertheless, the devices used in hospitals are complex and expensive, as several small sensors called electrodes are attached to the arms, legs, and chest of the patient on one end and are wired to ECG recording machines on the other end. Therefore, to overcome these limitations, ECG lead conversion has been widely studied.

Several studies have attempted to reduce the number of electrodes used in multi-lead ECGs. The vector cardiography (VCG) method was developed by displaying the spatial locations of ECG waveforms dependent on time. To perform this method, three orthogonal leads (i.e., X, Y, and Z) are required, wherein X and Z constitute the transverse plane, while Y and Z constitute the sagittal plane. Then, the Dower transformation is used to calculate the coefficients to reconstruct the 12-lead ECG data from the X, Y, and Z signals, based on Frank's lead system [23, 24]. Field et al. presented a dipole-based hypothesis of VCG based on Frank E, A, I, and S, where E, A, I, and S are the ECGs recorded at four sites [25]. Dawson et al. presented an affine transformation based on the previous Dower transformation [26]. However, these methods require complex mathematical calculations and parameter changes, and the results vary depending on the

measurement position.

Therefore, various DL methods, such as CNNs, RNNs, LSTM, and GANs were applied For ECG signal classification and prediction [27, 28]. GANs have demonstrated excellent performance in the field of image synthesis [29-31]. However, in the field of signal synthesis, GANs have not matured yet and are mainly used for data augmentation. In particular, the SynSigGAN model can generate biomedical signals, which are preprocessed by discrete wavelet transformation, with a generator that uses Bi-Grid LSTM layers [32]. However, inputs are treated as latent variables during the synthesis process, and therefore their use is limited to data augmentation. Shin et al. proposed a method for generating PPG signals from ECG signals using an LSTM generator and a CNN discriminator [33]. Although the evaluation scores between the reference and generated signals were high, the dicrotic notches of PPG signals were not properly generated. Wulan et al. used WaveNet, spectroGAN, and wavelet GAN models to generate ECG signals [34]. However, they managed to only generate one beat signal from lead Ⅱ. Additionally, for the WaveNet model, the resolution was low, and the learning time was long. Lee et al. proposed the R-peak alignment and time sequence embedding method to transform one-dimensional time-series data into two-dimensional owing to the better performance of GANs on two-dimensional time series; the model inputs are lead Ⅱ [35]. The proposed R-peak alignment method was effective, as chest leads were converted from limb leads with high accuracy. However, the intervals between P-Q-R-S-T in the ECG signal were not equal, and choosing R-peak as the median value resulted in more than one beat while preprocessing the data.

Studies on generating biosignal (e.g., ECG) synthetic data using GANs have achieved promising results. However, using limb lead I as an input to generate or be converted into other leads has not been investigated. Moreover, most prior studies have generated only one-beat signals.

Furthermore, although lead Ⅱ contains more information than lead I, the electrodes must be attached to the patient's legs. [36] However, using wearable devices, such as smartwatches, ECG signals can be obtained from the arm corresponding to lead I. Different from previous studies, this study uses lead I, obtained from wearable devices, as the input for the learning and generating processes. Then, using our proposed method, 12-lead ECGs can be generated from single-lead ECGs for further diagnosis purposes [37].

We trained our model on two independent datasets and one combined dataset. For Experiment 1, the PTB-XL dataset was used as the training dataset, and the China dataset was used as the test dataset. For Experiment 2, the China dataset was used as the training dataset, while the PTB-XL dataset was used as the test dataset, as depicted in Figure 14. For Experiment 3, both the PTB-XL and China datasets were used as the training dataset, as shown in Figure 15. In this study, two datasets were combined to obtain more records for learning. Moreover, the evaluation scores of Experiments 1, 2, and 3 were compared. Figure 16 depicts the overall structure of the proposed model. After learning, the signals generated by the optimized GAN model were compared to the reference ECG signals using the Fréchet distance (FD) score and MSE. Furthermore, after reconstruction, these signals were plotted in the ECG paper format.

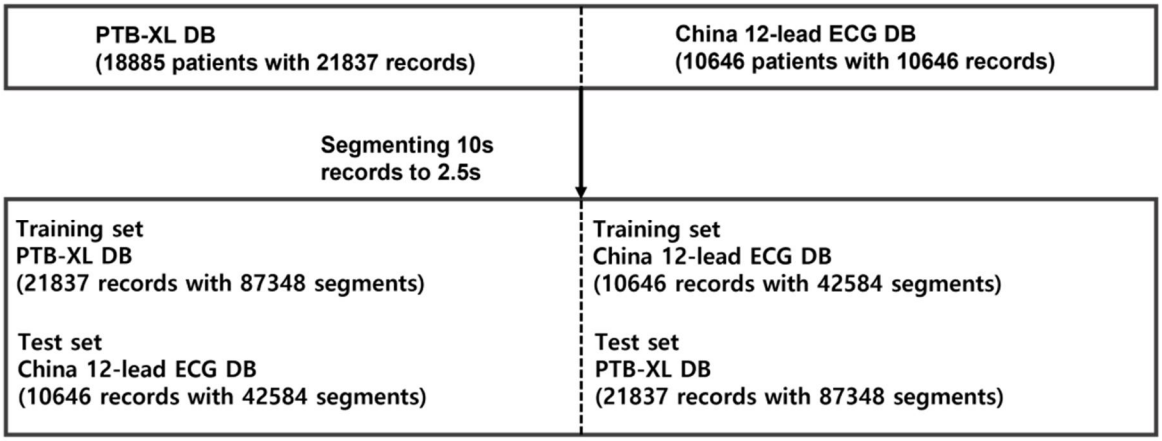


Figure 15. Data selection and segmentation for Experiments 1 and 2.

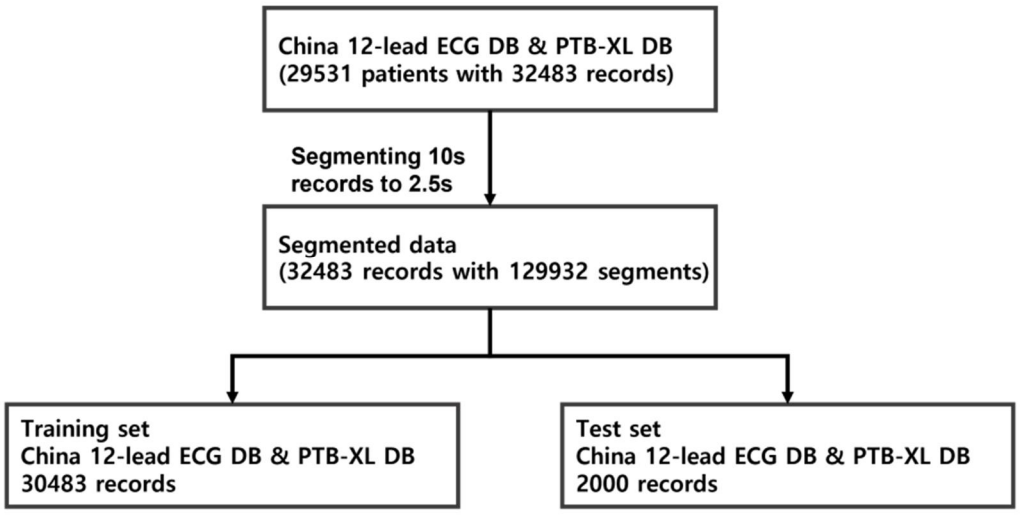


Figure 16. Data selection and segmentation for Experiment 3.

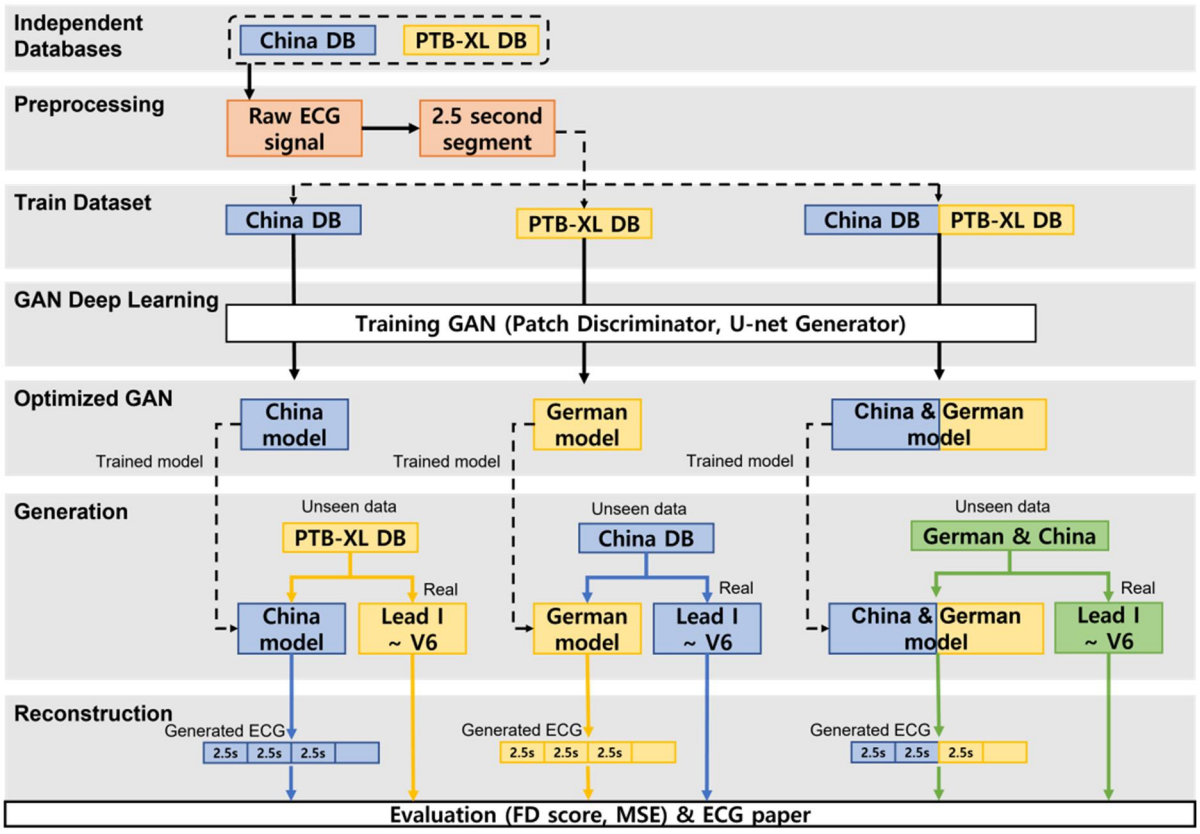


Figure 17. Overall structure of the proposed model

The 12-lead ECG data used in this study were obtained from the PTB-XL and China datasets, which are available on PhysioNet [38, 39]. Table 9 lists the clinical and demographic features of each dataset. The PTB-XL dataset contains 21,837 records obtained from 18,885 patients. The China dataset contains 10,646 records obtained from 10,646 patients. The duration of the records in the PTB-XL and China datasets is 10 s. The sampling rate of both datasets is 500 Hz, and the sampling points are 5,000 in each record. All 12-lead ECG signals were used in this study. Lead I was used as the input for our model, and the remaining leads were the reference signals.

Parameters	PTB-XL database	China database	Description
Age (years)	59.83 ± 16.95	59.18 ± 19.03	-
Gender (male: female)	11379: 10458	5956: 4690	-
Weight (Kg)	70.99 ± 15.97	-	-
Height (cm)	166.71 ± 10.86	-	-
Number of patients	**18885**	**10646**	
Diagnostic statement			
NORM	9528	5419	Normal ECG
STTC	5788	-	ST/T-Change
HYP	2819	-	Hypertrophy
MI	6886	-	Myocardial Infarction
CD	5772	-	Conduction Disturbance
Rhythm statement			
SR	16782	1826	Sinus Rhythm
AFIB	1514	1780	Atrial Fibrillation
ST	826	1568	Sinus Tachycardia
SA	772	-	Sinus Arrhythmia
SB	637	3889	Sinus Bradycardia
PACE	296	-	Normal Functioning Artificial Pacemaker
SVA	157	-	Supraventricular Arrhythmia
BIGU	82	-	Bigeminal Pattern
AF	73	445	Atrial Flutter
SVT	27	587	Supraventricular Tachycardia

PSVT	24	-	Paroxysmal Supraventricular Tachycardia
TRIGU	20	-	Trigeminal Pattern
SI	-	399	Sinus Irregularity
AT	-	121	Atrial Tachycardia
AVNRT	-	16	Atrioventricular Node Reentrant Tachycardia
AVRT	-	8	Atrioventricular Reentrant Tachycardia
SAAWR	-	7	Sinus Atrium to Atrial Wandering Rhythm

Table 9. Evaluation Score for Experiment 1

The 10 s recordings obtained from the PTB-XL and China datasets were divided into four segments, the duration of each was 2.5 s. Thus, we obtained 87,348 and 42,584 segments for the PTB-XL and China datasets, respectively.

The GAN architecture consists of two main networks: the generator and discriminator [40]. It is based on a minimax game between the generator and discriminator networks. In this study, the generator used lead I as the input and synthesized the remaining leads, while the discriminator distinguished the generated signals from real-world ones. Figure 17 depicts the overall architecture of the proposed method. Our model follows the main objective of conditional GANs, which learn to map according to signal x and random noise vector z to y [41, 42]. It is expressed as:

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[logD(x,y)] + \mathbb{E}_{x,z}[\log\big(1 - D\big(x, G(x,z)\big)\big)], \quad (2)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|], \quad (3)$$

where G attempts to minimize the objective of GAN (2) against D that attempts to maximize it. We used the L1 loss function. Our final objective of GAN can be represented as:

$$G^* = arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G). \quad (3)$$

The waveform of each lead is similar. Moreover, there are two to four beats in a 2.5 s ECG signal. Therefore, the generator of the proposed method is composed of U-net based encoder-decoder. The U-net generator is depicted in Figure 18. To avoid the bottleneck skipped connection, all channels were concatenated at corresponding layers [43]. The encoder part consisted of seven convolution layers. Batch normalization and the Leaky ReLU activation function were used in all layers, except for the first layer. The decoder consisted of seven upsampling convolution stacks. In general, data are usually normalized to [-1 1] or [0 1]; however, unlike images, ECG signals do not have minimum and maximum resolution values. In this case, the min-max normalization method is preferred. However, normalizing for each patient using minimum and maximum values could result in an output range of [-1 1], very different from the real-world ECG signals. To overcome this limitation, at the last layer of the decoder part, Leaky ReLU was used as an activation function instead of the hyperbolic tangent. The slope for all Leaky ReLU functions was 0.2, while the kernel size and stride length were 4 and 2, respectively.

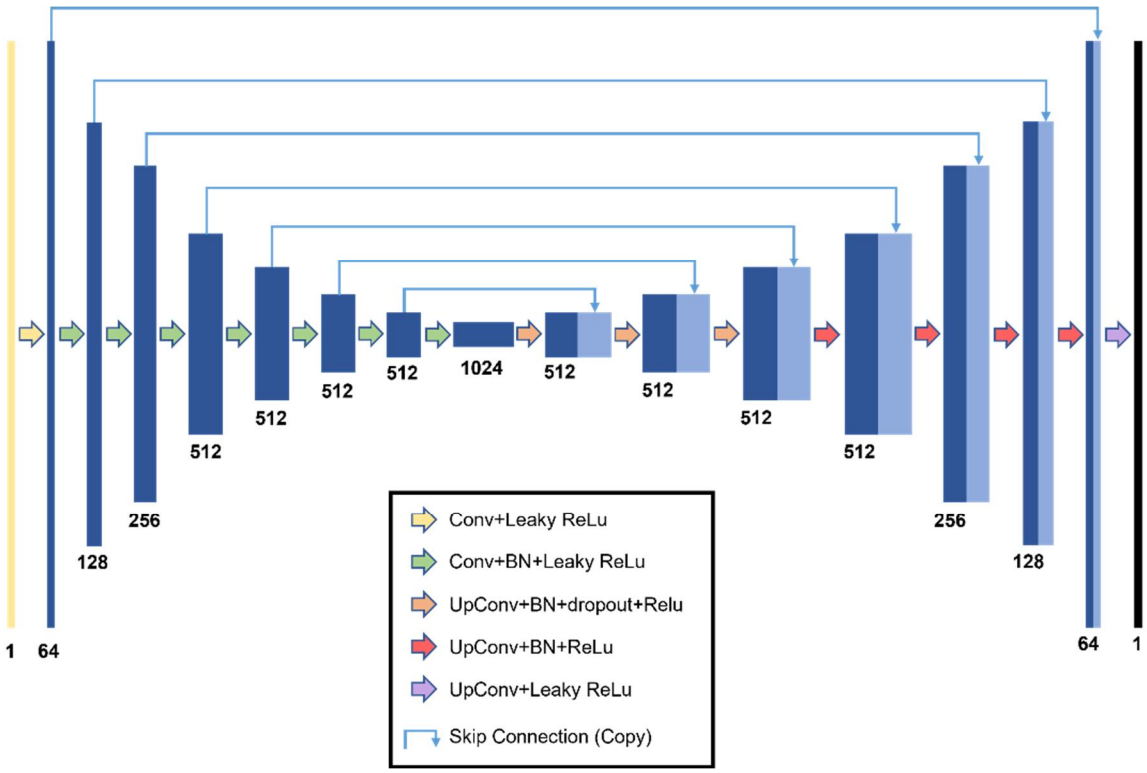


Figure 18. Architecture of the U-net generator

The discriminator depicted in Fig. 18 classifies each N patch in a signal as real or fake by labeling it as one for real and zero for fake. N is smaller than the full-size signal and is of high quality. The PatchGAN discriminator has an advantage in that it has few parameters and can be applied to longer signals. It contains five convolution layers with batch normalization and Leaky ReLU activation functions. After the last layer, a convolution layer is added to map one-dimensional output, followed by a Sigmoid function, further resulting in a discriminator's receptive field of 142. In this study, a slope of 0.2 for all Leaky ReLU functions, a kernel size of four, and a stride length of two were used. The learning rates were set to 0.0005 and 0.0001 for the generator and discriminator, respectively. In addition, the Adam algorithm was used for hyperparameter optimization, and the batch size was set as 32. A total of 11 models were trained for generating 11 leads. The complete process of the proposed method is depicted in Figure 20.

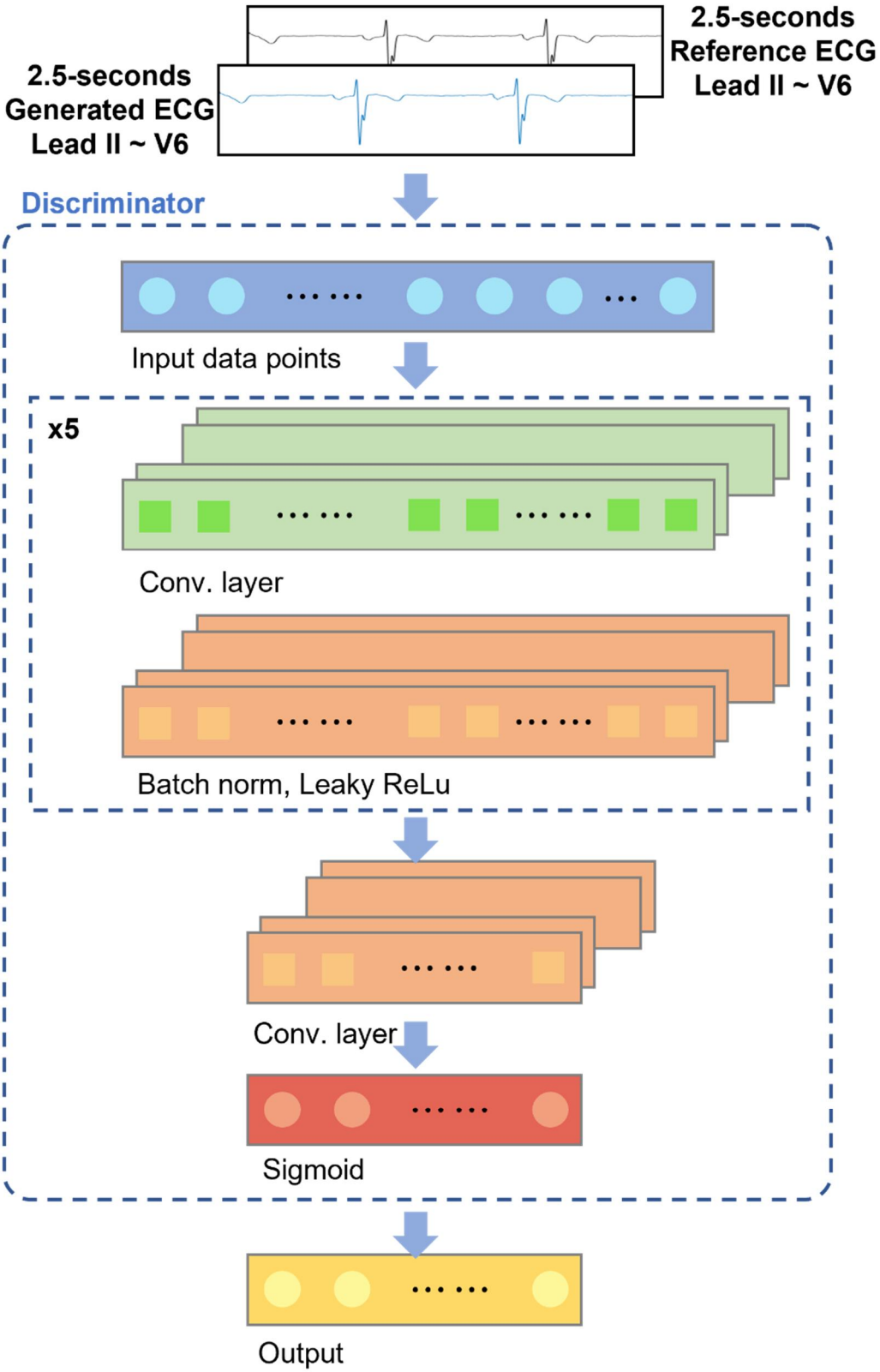


Figure 19. Architecture of the discriminator based on PatchGAN

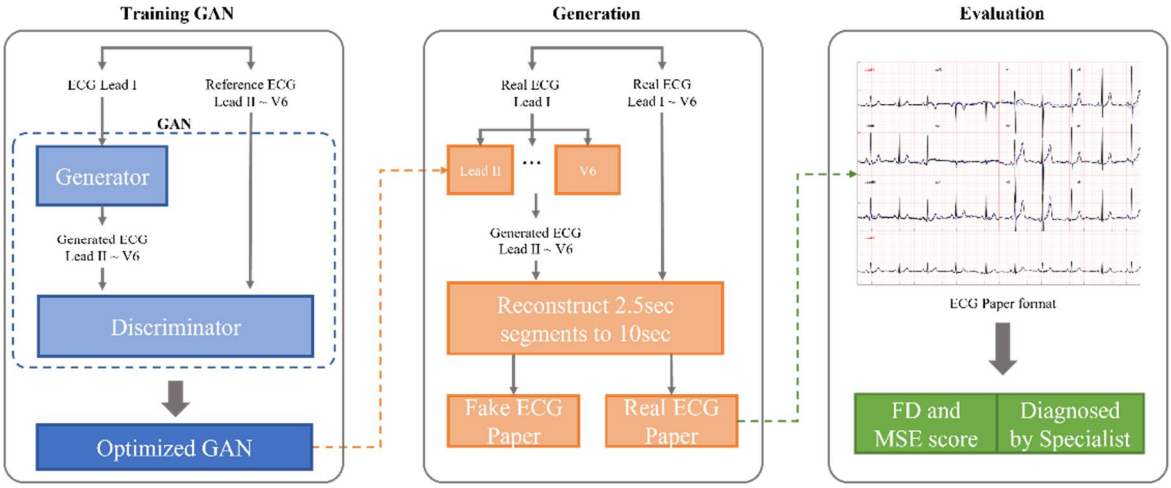


Figure 20. The proposed method: GAN training, generation, and evaluation

As depicted in Figure 21, the 12-lead ECG signals generated from the optimized GAN and corresponding signals were reconstructed. Then, the reconstructed 12-lead ECG signals with a duration of 10 s were evaluated using the FD and MSE scores. FD scores are used to find the similarity between curves, while MSE scores calculate the average difference between two signals. In this study, the FD score validated the P-Q-R-S-T sync between generated and reference signals, and the MSE score verified the difference in amplitude between signals. Moreover, for specific validation of ECG signals, the Bland-Altman plot and difference values between segments, such as RR interval, QRS duration, QT interval, and ST segments, were calculated.

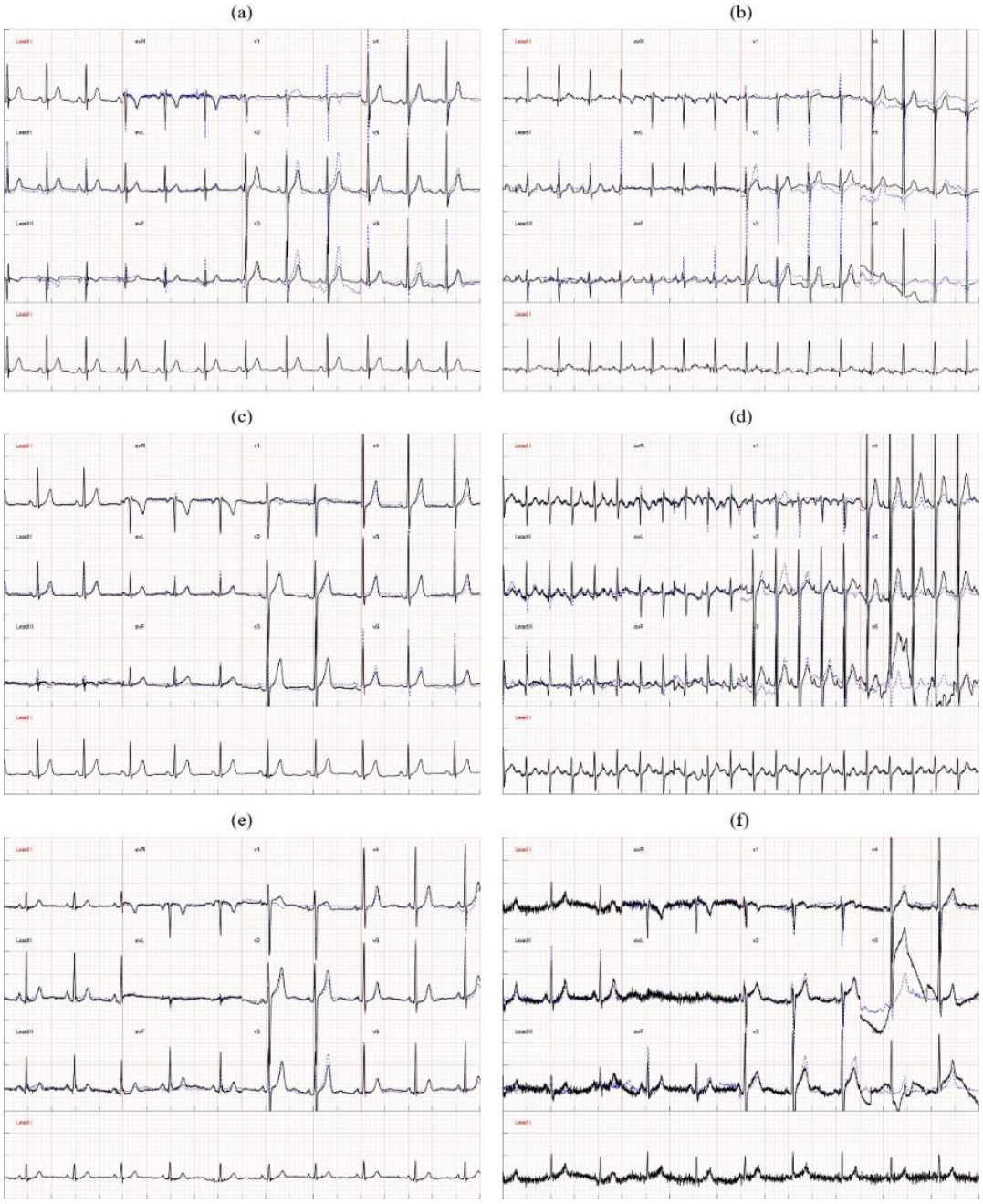


Figure 21. Generated samples by the proposed method. Dark lines represent the reference signals, and blue dashed lines are the generated signals. Here, (a) and (b) are samples of low and high evaluation results of Experiment 1, (c) and (d) are samples of low and high evaluation results of Experiment 2, and (e) and (f) are samples of low and high evaluation results of Experiment 3.

To train our model, two different activation functions were used: hyperbolic tangent and leaky ReLU. The hyperbolic tangent function was expected to perform better than leaky ReLU; however, the evaluation score of the former was higher than that of the latter. The generated signals trained using the hyperbolic tangent activation function failed mostly on transverse planes. Meanwhile, the FD scores of leaky ReLU were lower than those of hyperbolic tangent by 0.6 and 46 on frontal and transverse planes, respectively. The evaluation scores are shown in Table 10.

	Tanh		Leaky ReLU	
	FD score	**MSE**	**FD score**	**MSE**
Frontal Plane	4.086	0.01	3.429	0.01
Transverse Plane	59.249	0.076	13.757	0.042
All Planes	34.17	0.046	9.062	0.027

Table 10. Evaluation score of the two activation functions

In our study, FD and MSE scores were used for the evaluation. A total of 180 records of diagnostic data were evaluated in segments. The RR interval, QT interval, QRS duration, and ST interval were selected for further investigation. The differences were calculated for each segment. The results are shown in Table 11. The Bland-Altman plots are depicted in Figures 21-26. The results showed that segments of the generated and reference signals were similar.

	ΔRR interval	ΔQT interval	ΔQRS duration	ΔST interval
Lead Ⅱ	0.839	0.856	0.712	0.046
Lead Ⅲ	0.708	0.45	0.168	0.059
Lead avR	0.904	0.585	0.337	0.068
Lead avL	0.855	0.666	0.469	0.051
Lead avF	0.688	0.384	0.027	0.052
V1	0.905	0.469	0.176	0.059
V2	0.73	0.553	0.276	0.049
V3	0.872	0.622	0.251	0.055
V4	0.802	0.571	0.309	0.048
V5	0.682	0.523	0.311	0.047
V6	0.833	0.602	0.2888	0.049
Mean	0.801	0.571	0.302	0.053

Table 11. Calculated differences for each segment

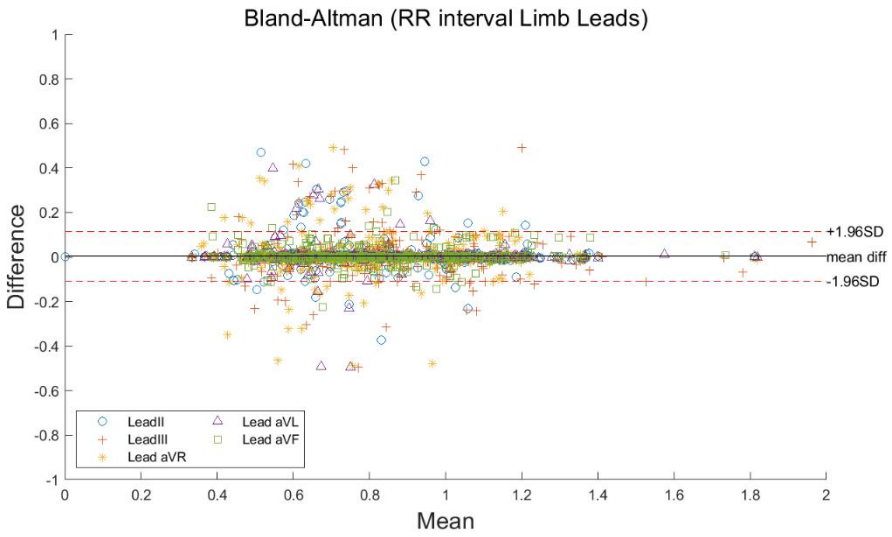


Figure 22. Bland-Altman plot of RR interval (limb lead) for 180 records of diagnosable data. Results are in the range of ±1.96 SD, representing the similarity between generated and reference signals

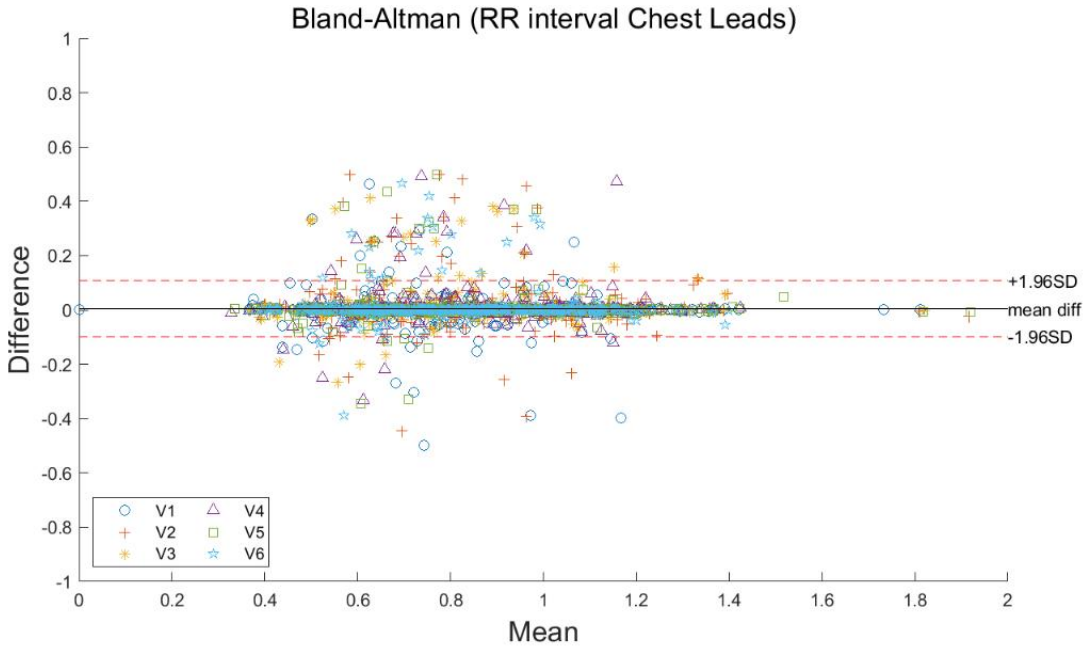


Figure 23. Bland-Altman plot of RR interval (chest lead) for 180 records of diagnosable data. Results are in the range of ±1.96 SD, representing the similarity between generated and reference signals.

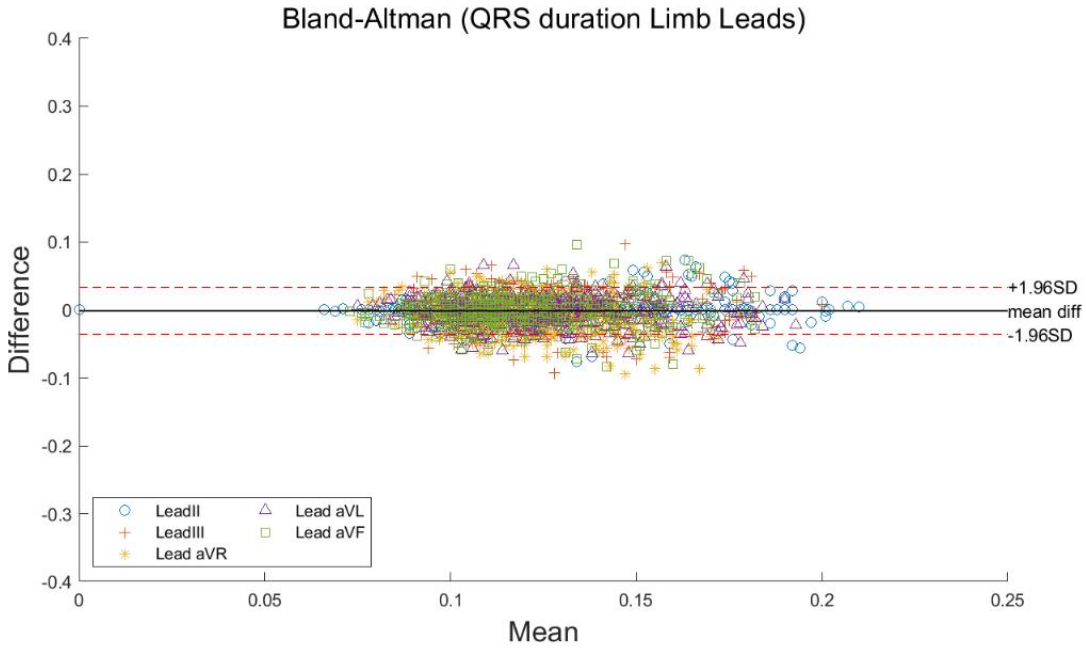


Figure 24. Bland-Altman plot of QRS duration (limb lead) for 180 records of diagnosable data. Results are in the range of ±1.96 SD, representing the similarity between generated and reference signals.

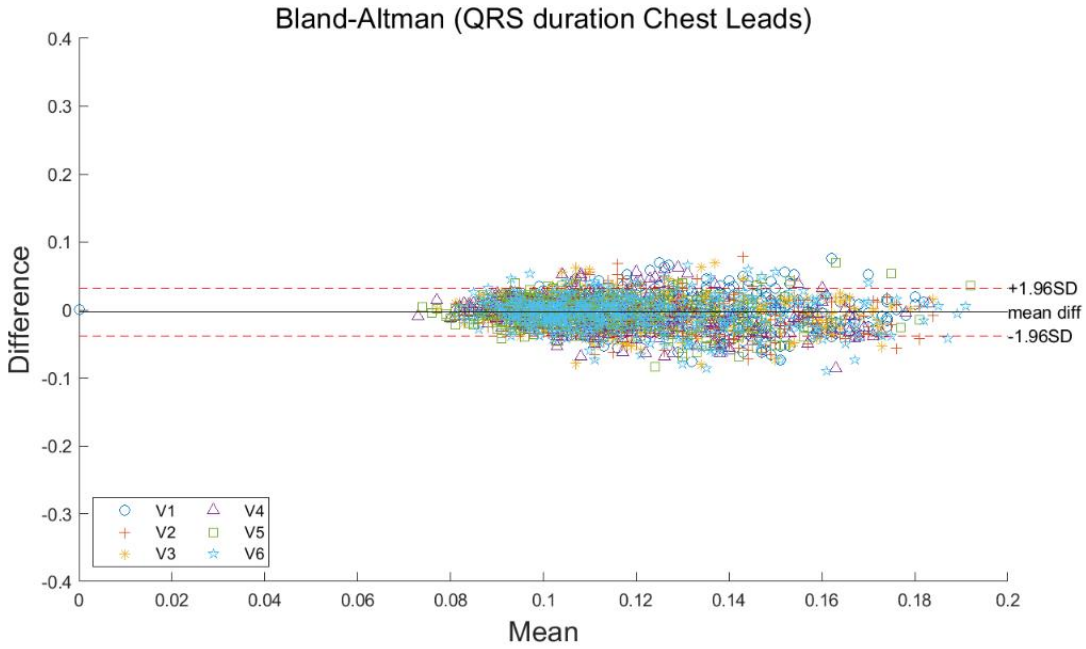


Figure 25. Bland-Altman plot of QRS duration (chest lead) for 180 records of diagnosable data. Results are in the range of ±1.96 SD, representing the similarity between generated and reference signals.

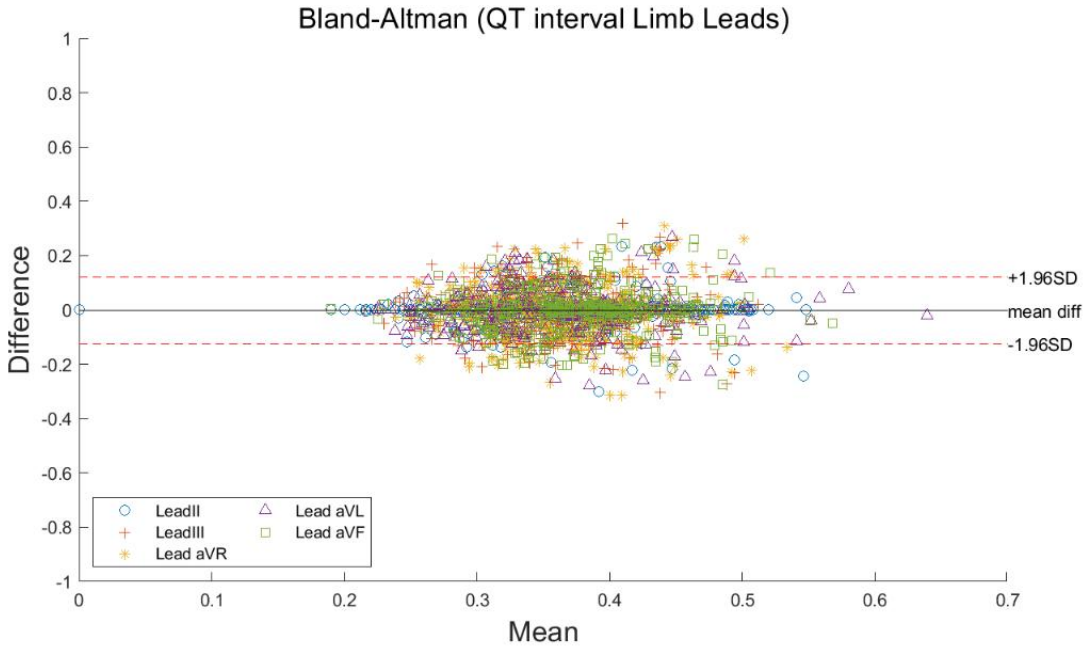


Figure 26. Bland-Altman plot of the QT interval (limb lead) for 180 records of diagnosable data. Results are in the range of ±1.96 SD, representing the similarity between generated and reference signals.

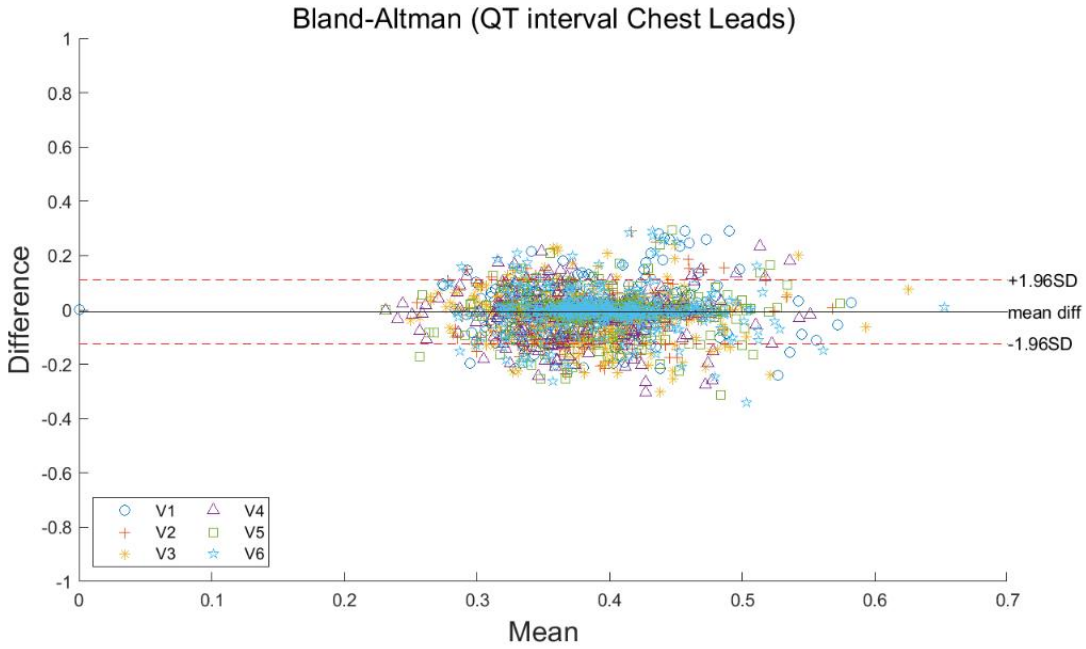


Figure 27. Bland-Altman plot of the QT interval (chest lead) for 180 records of diagnosable data. Results are in the range of ±1.96 SD, representing the similarity between generated and reference signals.

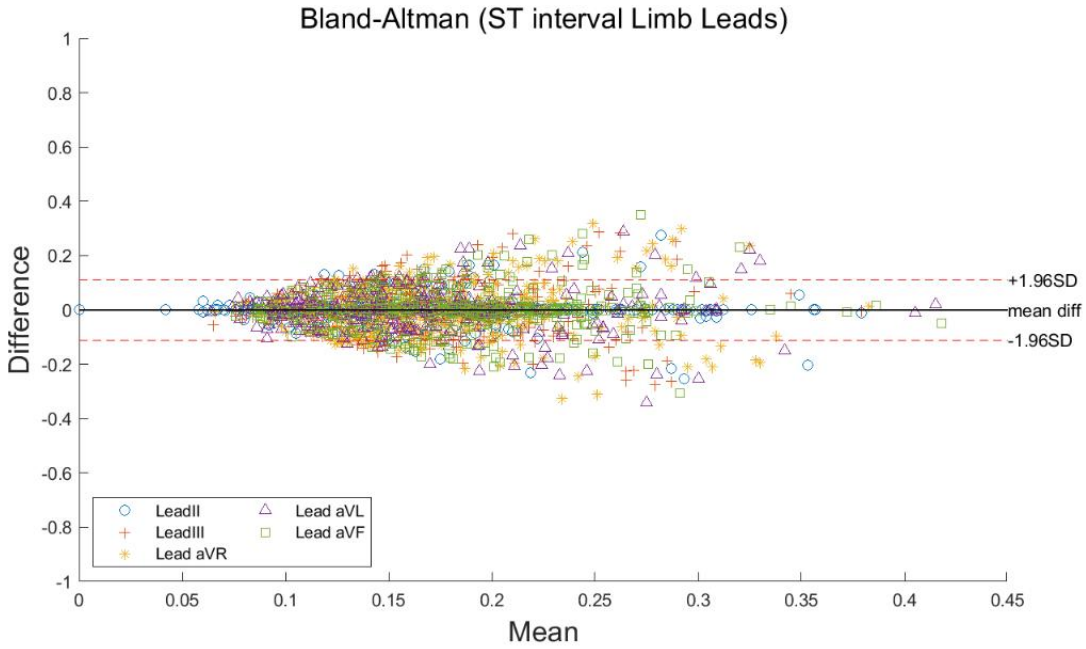


Figure 28. Bland-Altman plot of the ST interval (chest lead) for 180 records of diagnosable data. Results are in the range of ±1.96 SD, representing the similarity between generated and reference signals.

MSE measures the average squared difference between real-world (X) and generated signals (Y) as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2 \quad (4)$$

FD is used to find the similarity between the curves of the generated and corresponding real-world signals. If O_X=a_1,a_2,a_3,…,a_X is the ordered points of real-world signal curves and O_Y=b_1,b_2,b_3,…,b_Y is the ordered points of the generated signal curves, then the length of ||l|| can be calculated as follows:

$$\|l\| = \max_{i=1,\ldots,n} l(a_i, b_i). \quad (5)$$

Therefore, FD can be calculated as follows:

$$FD(X, Y) = min\|l\| \quad (6)$$

Thus far, the generated signals learned from two independent datasets and one combined dataset were presented and compared with the reference 12-lead ECG signals. Figure 20 depicts two sections of the proposed method: GAN training and evaluation.

The generated and reference ECG signals were plotted in an actual ECG paper to compare and verify if they match. Finally, the evaluation scores of the three experiments were compared. Evaluation scores of all three experiments are listed in Tables 12 and 13. The lowest evaluation scores are shown in bold font.

In Experiment 1, the China dataset was used to train the model, and the PTB-XL dataset was used to evaluate the optimized GAN model. From Tables 12 and 13, the mean values of the FD and MSE scores were 9.062 and 0.027, respectively. For the limb lead generated signals, the FD and MSE scores were lower, that is, 3.429 and 0.01, respectively. By contrast, evaluation scores for the transverse plane were higher than those of the frontal plane. In Experiment 2, the model was trained on the PTB dataset. The mean FD and MSE scores were 8.124 and 0.027, respectively. The FD and MSE scores for the frontal plane were lower than those of the transverse plane by 7 and 0.02, respectively. The evaluation scores for the two models were similar. In Experiment 3, the PTB-XL and China datasets were combined so that our model learns with more data. The FD and MSE scores of Experiment 3 were 6.7 and 0.017, respectively, which was less than the scores obtained in Experiments 1 and 2. In addition, the evaluation scores of the frontal and transverse plane were lower compared to those of Experiments 1 and 2. The results of Experiment 3 indicated that combining the PTB-XL and China datasets to acquire more data improved the model evaluation scores.

Figure 21 depicts the generated 12-lead ECG signals and the corresponding real-world 12-lead ECG signals from Experiments 1, 2, and 3. In the presented ECG paper format, one large box signal of P-Q-R-S-T segments followed the reference signals, and the

negative-positive values for each P-Q-R-S-T peaks matched. The results of the reconstructed 10 s ECG paper are depicted in Figure 21 (a); the sync of not only the R-peak, but also the remaining peaks matched with the input lead I ECG signals. For Experiment 2, the generated and reference 12-lead ECG signals are depicted in Figure 21 (c). Similar to Experiment 1, for generated signals, the sync between leads and P-Q-R-S-T segments followed the reference signals. Finally, the generated signals from Experiment 3 are depicted in Figure 21 (e).

Figure 21 depicts the results of both high and low FD score samples. Figures 21 (a), (c), and (e) illustrate samples with low FD and MSE scores, whereas Figure 21 (b), (d), and (f) show samples with high FD and MSE scores. As mentioned in Table 12, although the overall FD and MSE scores were low, a few samples depicted in Figure 21 had high FD and MSE scores.

	Experiment 1		Experiment 2		Experiment 3	
	FD score (mV)	MSE (mV)	FD score (mV)	MSE (mV)	FD score (mV)	MSE (mV)
Lead Ⅱ	4.551	0.014	2.189	0.011	2.561	0.008
Lead Ⅲ	4.927	0.014	1.596	0.012	1.974	0.008
Lead aVR	0.936	0.004	0.504	0.003	0.359	0.009
Lead aVL	1.483	0.004	0.690	0.004	0.219	0.002
Lead aVF	5.249	0.014	16.460	0.042	2.272	0.008
V1	13.637	0.029	6.101	0.021	8.032	0.016
V2	10.847	0.058	10.837	0.050	9.382	0.029
V3	17.639	0.058	10.128	0.051	9.373	0.031
V4	15.265	0.046	12.581	0.043	7.929	0.024
V5	13.602	0.033	15.931	0.034	12.153	0.025
V6	11.550	0.028	12.347	0.026	19.453	0.030
Mean value	9.062	0.027	8.124	0.027	6.701	0.017

Table 12. FD and MSE scores for each electrode

	Experiment 1		Experiment 2		Experiment 3	
	FD score (mV)	MSE (mV)	FD score (mV)	MSE (mV)	FD score (mV)	MSE (mV)
Frontal Plane	3.429	0.010	4.288	0.0144	1.477	0.007
Transverse Plane	13.757	0.042	11.321	0.038	11.054	0.026
All Planes	9.062	0.027	8.124	0.027	6.701	0.017

Table 13. FD and MSE scores for each plane

Next, we present a novel method of generating more than one-beat 12-lead ECG signal from lead I. The lead Ⅱ method of measuring ECG signals may be more efficient. However, wearable devices such as smartwatches are attached to the arm; therefore, the conversion of ECG signals from lead I is important. Minimizing the preprocessing procedure allows the utilization of all raw data. Moreover, to generate more than one-beat signal, the U-net generator and patch discriminator were used. The proposed method showed excellent performance by following the P-Q-R-S-T pattern. As depicted in Figure 21 (a), (c), and (e), the sync between signals matched. Low FD and MSE scores were obtained because of combining the PTB-XL and China datasets. However, the amplitude of generated signals was not equivalent to that of reference signals. As mentioned previously, similar to images, ECG signals do not have absolute maximum and minimum values, as they vary between patients. Additionally, since the PTB-XL and China datasets had 18,885 and 10,646 patients, variations in patient data resulted in uneven amplitude results. In Experiment 3, the overall FD and MSE scores were low, indicating that the model performed better after being trained with combined data.

Figures 21 (b), (d), and (f) show samples of high FD and MSE scores. The FD scores of lead V6 were 3222, 2669, and 2240. The MSE scores were 2.52, 2.08, and 1.74. This indicates that the generated and reference signals do not correlate. However, in Figure 21 (b) and (d), the reference V6 signal shows baseline wander, causing high evaluation scores and noise in lead I. As shown in Figure 21 (f), artifacts in lead V5 caused high evaluation scores. Therefore, signal generation with baseline wander and artifacts was limited.

Additionally, 180 records of abnormal cases were generated and evaluated. The evaluation scores of abnormal cases are listed in Tables 14 and 15. The mean FD and MSE scores were 23.645 and 0.065, respectively. The evaluation scores were higher than previous results. However, the evaluation scores of limb leads was lower by 30 mV compared to chest leads.

Lead	FD score (mV)	MSE (mV)
Lead Ⅱ	9.796	0.032
Lead Ⅲ	11.391	0.034
Lead aVR	1.689	0.009
Lead aVL	2.248	0.008
Lead aVF	10.544	0.035
V1	13.483	0.053
V2	57.493	0.143
V3	53.925	0.156
V4	45.807	0.123
V5	26.472	0.067
V6	27.248	0.054
Mean value	23.645	0.065

Table 14. FD and MSE scores for abnormal cases in each electrode

Plane	FD score (mV)	MSE (mV)
All Plane	23.645	0.065
Frontal Plane	7.134	0.024
Transverse Plane	37.405	0.099

Table 15. FD and MSE scores for abnormal cases in each plane

The generated abnormal signals were annotated by a cardiologist with 20+ years of experience. The abnormal signals learned in this study were left bundle branch block (LBBB), right bundle branch block (RBBB), left ventricular hypertrophy (LVH), right ventricular hypertrophy (RVH), myocardial infarction (MI), and Wolff-Parkinson-White syndrome (WPW). The cardiologist provided the diagnosis results by observing all 12-lead signals represented in ECG paper format. Table 16 lists whether the cardiologist was able to reach diagnosis according to the generated signals. The generated signals of LVH, RBBB, and LBBB showed high diagnosability, represented by the percentage of diagnosable data. Over 85% of the generated signals could be used to diagnose diseases. However, WPW, RVH, and MI showed low diagnosability, indicating several limitations for diagnosing abnormal diseases.

	Diagnosable	Fair	Undiagnosable	All	Performance
LVH	52	2	1	55	95%
RBBB	41	2	1	44	89%
LBBB	25	4	0	29	86%
WPW	12	3	2	17	63%
RVH	3	4	3	10	20%
MI	0	2	7	9	0%

Table 16. Performance on abnormal cases

From the results, first, although the diagnosability of LVH, LBBB, and RBBB were high, a few of the generated signals resulted in uneven amplitude beat by beat. Second, the strain patterns were not shown properly as reference signals. Third, the evaluation scores of chest leads were low. Therefore, although inferior wall defects were diagnosed, diagnosis of anterior wall defects was difficult, resulting in low performance in generating MI signals. Finally, while generating WPW signals, most signals showed delta patterns; however, some were directed to opposite positions of the heart. Lack of abnormal data mainly caused the above limitations. Therefore, further studies should be performed with more abnormal data.

Table 17 compares the results for our method with those of related works on ECG synthesis by GANs. All previous studies used lead Ⅱ as input; however, it is not suitable for wearable devices, such as smartwatches. Generated leads were limited to same as input or only chest leads. Additionally, SynSigGAN, Wulan et al, and Leea et al. used denoising methods that distorted raw signals. Although denoising, data augmentation, and R-peak alignment methods may be effective in deep learning, they are not applicable in real-world generation. Moreover, previous methods had limitations in generation length. Wulan et al. generated 2.5s ECG signals, while Lee et al.'s generation method was limited to 1s generation due to the use of R-peak alignment method in preprocessing. Conversely, our method can generate more than 10s ECG signals. Finally, the evaluation scores vary due to different evaluation methods. The FD score for the SynSigGAN model was 0.936 lower than that of the proposed method. However, FD scores between reference and generated signals were not calculated for the two methods.

Method	Input lead	Generated Lead	Preprocess Method	Generation length	Result
SynSigGAN	Lead Ⅱ	Lead Ⅱ	Wavelet denoising	median R-R interval time	FD: 0.936 mV MAE: 0.218 mV
Wulan et al	Lead Ⅱ	Lead Ⅱ	Noise reduction	2.5 s	GAN train score: 89.07%
Lee et al	Lead Ⅱ	V1~V6	Down sampling R-peak alignment Data augmentation	1 s	Corr: 0.86 Amp: 0.21 mV SSIM: 0.21 mV
Proposed method	Lead Ⅰ	12 Lead	None	10 s	FD: 6.701 mV MSE: 0.017 mV

Table 17. Comparison between the proposed method and other studies related to ECG synthesis based on GANs

Finally, in response to the demand for accurate diagnoses, the Internet of Medical Things (IoMT) technologies are being developed. Furthermore, smartwatches are closely related to IoMT, as they enable the tracking and monitoring of biosignals, including ECG signals. Using our proposed method, more detailed ECG signals can be obtained from IoMT devices, such as smartwatches.

In this study, a novel method of lead conversion based on lead I was developed. Experimental results showed that the proposed method could generate 12-lead ECGs from lead I signals with low FD and MSE scores. Prior studies, such as [32] and [33], used preprocessing methods for ECG signal segmentation by beats or by R peaks, thereby distorting the morphology of ECG signals and being unable to generate the actual value corresponding to the reference signal. In contrast, our proposed method for ECG generation showed superior performance without preprocessing raw signals. Additionally, A comparison of ECG papers demonstrated that the generated signal waveforms of P-Q-R-S-T, by our method, follows the reference signals. Furthermore, the sync between leads matches. Importantly, this method can be applied to wearable devices that measure ECG signals based on a single lead, thus contributing to the growth of IoMT.

Conclusion

In this study, several DL-based digital healthcare models were developed using cardiovascular biosignal data. In the first chapter, we developed a DL model for AF detection in patients (binary classification into AF and non-AF patients) using ambulatory (Holter) ECG data obtained from three datasets. Additionally, we verified ECG data dependency for AF detection using ResNet architectures. Specifically, owing to the characteristics of medical data, such as data imbalance, we found that if a database has AF events far less than normal rhythms, which is common in biosignal databases, the trained model can be biased, resulting in reduction in performance when tested on external datasets. Then, in the next chapter, we developed a DL model, by combining ResNet and LSTM architectures, for real-time indirect continuous arterial BP measurements using ECG and PPG waveforms. We verified that our proposed method overcomes the limitations of rule-based methods in the clinical field. Finally, in the third chapter, using a GAN model, we were able to convert single-lead ECGs to 12-lead ECGs, thus obtaining more detailed readings of heart signals using wearable devices, such as smartwatches, which are less complex than devices used in hospitals. Moreover, the biosignals generated by this method could be used to monitor the heart condition and diagnose diseases in most cases. Nevertheless, the lack of abnormal data mainly imposed limitations on the diagnosis of abnormal diseases. Therefore, to overcome this problem, future studies should use more abnormal data. Our results highlight the untapped potential of DL approaches in enhancing the growth of the IoMT industry and digital healthcare applications.

Reference

[1] R. Agrawal, and S. Prabakaran, "Big data in digital healthcare: lessons learnt and recommendations for general practice," *Heredity (Edinb),* vol. 124, no. 4, pp. 525-534, Apr, 2020.

[2] A. Arora, "Conceptualising Artificial Intelligence as a Digital Healthcare Innovation: An Introductory Review," *Med Devices (Auckl),* vol. 13, pp. 223-230, 2020.

[3] S. Colilla, A. Crow, W. Petkun *et al.*, "Estimates of current and future incidence and prevalence of atrial fibrillation in the U.S. adult population," *Am J Cardiol,* vol. 112, no. 8, pp. 1142-7, Oct 15, 2013.

[4] M. H. Kim, S. S. Johnston, B. C. Chu *et al.*, "Estimation of total incremental health care costs in patients with atrial fibrillation in the United States," *Circ Cardiovasc Qual Outcomes,* vol. 4, no. 3, pp. 313-20, May, 2011.

[5] S. S. Chugh, R. Havmoeller, K. Narayanan *et al.*, "Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study," *Circulation,* vol. 129, no. 8, pp. 837-47, Feb 25, 2014.

[6] A. Clim, R. D. Zota, and G. Tinica, "Big Data in home healthcare: A new frontier in personalized medicine. Medical emergency services and prediction of hypertension risks," *International Journal of Healthcare Management,* vol. 12, no. 3, pp. 241-249, 2019/07/03, 2019.

[7] H. C. Seo, S. Oh, H. Kim *et al.*, "ECG data dependency for atrial fibrillation detection based on residual networks," *Sci Rep,* vol. 11, no. 1, pp. 18256, Sep 14, 2021.

[8] K. He, Zhang, X., Ren, S. & Sun, J, " Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[9] S. Petrutiu, A. V. Sahakian, and S. Swiryn, "Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans," *Europace,* vol. 9, no. 7, pp. 466-70, Jul, 2007.

[10] G. Moody, "A new method for detecting atrial fibrillation using RR intervals," *Computers in Cardiology*, pp. 227-230, 1983.

[11] G. B. Moody, and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng Med Biol Mag,* vol. 20, no. 3, pp. 45-50, May-Jun, 2001.

[12] D. Gu, T. N. Kelly, X. Wu *et al.*, "Blood pressure and risk of cardiovascular disease in Chinese men and women," *Am J Hypertens,* vol. 21, no. 3, pp. 265-72, Mar, 2008.

[13] R. S. Vasan, J. M. Massaro, P. W. Wilson *et al.*, "Antecedent blood pressure and risk of cardiovascular disease: the Framingham Heart Study," *Circulation,* vol. 105, no. 1, pp. 48-53, Jan 1, 2002.

[14] L. A. Geddes, and L. E. Baker, *Principles of applied biomedical instrumentation*: John Wiley

& Sons, 1975.

[15] A. S. Alghamdi, K. Polat, A. Alghoson *et al.*, "Gaussian process regression (GPR) based non-invasive continuous blood pressure prediction method from cuff oscillometric signals," *Applied Acoustics,* vol. 164, pp. 107256, 2020.

[16] M. Kachuee, M. M. Kiani, H. Mohammadzade *et al.*, "Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring," *IEEE Trans Biomed Eng,* vol. 64, no. 4, pp. 859-869, Apr, 2017.

[17] F. Miao, N. Fu, Y. T. Zhang *et al.*, "A Novel Continuous Blood Pressure Estimation Approach Based on Data Mining Techniques," *IEEE J Biomed Health Inform,* vol. 21, no. 6, pp. 1730-1740, Nov, 2017.

[18] I. Sharifi, S. Goudarzi, and M. B. Khodabakhshi, "A novel dynamical approach in continuous cuffless blood pressure estimation based on ECG and PPG signals," *Artif Intell Med,* vol. 97, pp. 143-151, Jun, 2019.

[19] X. Xing, and M. Sun, "Optical blood pressure estimation with photoplethysmography and FFT-based neural networks," *Biomed Opt Express,* vol. 7, no. 8, pp. 3007-20, Aug 1, 2016.

[20] F. Miao, B. Wen, Z. Hu *et al.*, "Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques," *Artif Intell Med,* vol. 108, pp. 101919, Aug, 2020.

[21] H. Eom, D. Lee, S. Han *et al.*, "End-to-End Deep Learning Architecture for Continuous Blood Pressure Estimation Using Attention Mechanism," *Sensors (Basel),* vol. 20, no. 8, Apr 20, 2020.

[22] Y. H. Li, L. N. Harfiya, K. Purwandari *et al.*, "Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model," *Sensors (Basel),* vol. 20, no. 19, Sep 30, 2020.

[23] G. E. Dower, "A lead synthesizer for the Frank system to simulate the standard 12-lead electrocardiogram," *J Electrocardiol,* vol. 1, no. 1, pp. 101-16, 1968.

[24] E. Frank, "An accurate, clinically practical system for spatial vectorcardiography," *Circulation,* vol. 13, no. 5, pp. 737-49, May, 1956.

[25] D. Q. Feild, C. L. Feldman, and B. M. Horacek, "Improved EASI coefficients: their derivation, values, and performance," *J Electrocardiol,* vol. 35 Suppl, pp. 23-33, 2002.

[26] D. Dawson, H. Yang, M. Malshe *et al.*, "Linear affine transformations between 3-lead (Frank XYZ leads) vectorcardiogram and 12-lead electrocardiogram signals," *J Electrocardiol,* vol. 42, no. 6, pp. 622-30, Nov-Dec, 2009.

[27] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez *et al.*, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *Lancet,* vol. 394, no. 10201, pp. 861-867, Sep 7, 2019.

[28] O. Yildirim, P. Plawiak, R. S. Tan *et al.*, "Arrhythmia detection using deep convolutional

neural network with long duration ECG signals," *Comput Biol Med,* vol. 102, pp. 411-420, Nov 1, 2018.

[29] T. Karras, S. Laine, M. Aittala *et al.*, "Analyzing and improving the image quality of stylegan." pp. 8110-8119.

[30] C. Ledig, L. Theis, F. Huszár *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network." pp. 4681-4690.

[31] J.-Y. Zhu, T. Park, P. Isola *et al.*, "Unpaired image-to-image translation using cycle-consistent adversarial networks." pp. 2223-2232.

[32] D. Hazra, and Y. C. Byun, "SynSigGAN: Generative Adversarial Networks for Synthetic Biomedical Signal Generation," *Biology (Basel),* vol. 9, no. 12, Dec 3, 2020.

[33] H. Shin, S. Sun, J. Lee *et al.*, "Complementary photoplethysmogram synthesis from electrocardiogram using generative adversarial network," *IEEE Access,* vol. 9, pp. 70639-70649, 2021.

[34] N. Wulan, W. Wang, P. Sun *et al.*, "Generating electrocardiogram signals by deep learning," *Neurocomputing,* vol. 404, pp. 122-136, 2020.

[35] J. Lee, K. Oh, B. Kim *et al.*, "Synthesis of electrocardiogram V-lead signals from limb-lead measurement using R-peak aligned generative adversarial network," *IEEE journal of biomedical and health informatics,* vol. 24, no. 5, pp. 1265-1275, 2019.

[36] J. Francis, "ECG monitoring leads and special leads," *Indian Pacing Electrophysiol J,* vol. 16, no. 3, pp. 92-95, May - Jun, 2016.

[37] H. C. Seo, G. W. Yoon, S. Joo *et al.*, "Multiple electrocardiogram generator with single-lead electrocardiogram," *Comput Methods Programs Biomed,* vol. 221, pp. 106858, Jun, 2022.

[38] P. Wagner, N. Strodthoff, R. D. Bousseljot *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Sci Data,* vol. 7, no. 1, pp. 154, May 25, 2020.

[39] J. Zheng, J. Zhang, S. Danioko *et al.*, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Scientific data,* vol. 7, no. 1, pp. 1-8, 2020.

[40] I. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, "Generative adversarial nets," *Advances in neural information processing systems,* vol. 27, 2014.

[41] P. Isola, J.-Y. Zhu, T. Zhou *et al.*, "Image-to-image translation with conditional adversarial networks." pp. 1125-1134.

[42] M. Mirza, and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[43] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation." pp. 234-241.

Abstract

Biosignals represent an important type of digital healthcare data. Although these signals have to be monitored continually in real time for long periods to provide advanced diagnostic and therapeutic capabilities for patients, healthcare workers cannot always perform this task efficiently. Additionally, such data are usually atypical, non-standardized, irregular, and extremely vulnerable to noise. Thus, there is an urgent need for an effective and comprehensive automated system that can collect and analyze such data in real time. Using biosignal data, this study develops and applies several deep learning (DL) models for digital healthcare applications.

First, to overcome the problem of data dependency prevalent in the medical field due to the differences in analog to digital (AD) resolution and sampling rate of the measuring equipment, we used meta-analysis to integrate data from various databases. Specifically, we collected ECG data from three open databases, and used different ResNet architectures for AF detection in each database. Our results showed that the accuracy was lowered when the trained models were evaluated on datasets other than the one containing the training dataset, thus confirming dependency on the training data. it is necessary to validate the DL-based AF detection algorithm using various external databases to overcome the problem of data dependency.

Next, by combining ResNet and LSTM architectures, we developed a DL model for real-time indirect continuous arterial BP measurement using ECG and PPG waveforms. We verified that our proposed method realizes continuous blood pressure estimation in compliance with international standards. In addition, it can estimate BP for AF patients.

Finally, we used a GAN model to convert single-lead ECGs to 12-lead ECGs, thus obtaining more detailed readings of heart signals using wearable and less complex devices, such as smartwatches. Moreover, the multi-electrode ECGs generated by this method could be used to monitor the heart condition and diagnose diseases in the clinical field. Thus, the proposed method can be integrated in IoMT devices and digital healthcare applications.

Through several digital healthcare studies using biosignals, we confirmed that DL-based applications can be effectively used in clinical conditions.