Deep Learning Approaches for Arrhythmia Screening in Adhesive Patch-Type Wearable Electrocardiographs

Sangkyu Kim

The Graduate School

Yonsei University

Graduate Program of Biomedical Engineering

Deep Learning Approaches for Arrhythmia Screening in Adhesive Patch-Type Wearable Electrocardiographs

A Dissertation Submitted

to the Graduate Program of Biomedical Engineering

and the Graduate School of Yonsei University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy in Graduate Program of Biomedical Engineering

Sangkyu Kim

June 2024

This certifies that the dissertation of
Sangkyu Kim is approved.

Thesis Supervisor: Hee Cheol Kang

Thesis Committee Member: Sun Kook Yoo

Thesis Committee Member: Jong Chul Park

Thesis Committee Member: Byung Chae Lee

Thesis Committee Member: Ja Young Kim

The Graduate School
Yonsei University
June 2024

Acknowledgements

먼저 석사 및 박사 학위과정에 걸쳐 연세대학교 의과대학 의학공학교실에서 배움과 연구의 기회를 주신 유선국 교수님께 진심으로 감사의 말씀을 드립니다. 또한 귀중한 시간을 할애하여 저의 박사학위 논문을 지도해주신 강희철 교수님께 감사의 말씀을 드립니다. 바쁘신 와중에도 학위논문의 심사위원을 맡아 주시고 따뜻한 격려와 조언을 해 주신 박종철 교수님과 이병채 교수님, 김자영 교수님께 감사드립니다. 졸업 후에도 교수님들의 지도를 이어받아 연구와 학문에 정진하며, 사회에 기여할 수 있는 제자가 되도록 노력하겠습니다.

다음으로 좋은 연구 주제와 최적의 연구 환경을 제공해주신 ㈜휴이노와 인공지능팀 동료 여러분께 감사의 인사를 전합니다. 동료 여러분의 협력과 지원 덕분에 인공지능이라는 분야를 배울 수 있었고, 학위논문을 잘 마칠 수 있었습니다. 여러분의 도움 없이는 이룰 수 없는 성과였습니다. 진심으로 감사드립니다.

다음으로 석사 및 박사 학위과정에서 연구실 생활 및 진로에 많은 조언을 주셨던 UHCL 연구실의 수많은 선후배님들께 감사인사를 드립니다. 졸업 후에도 많은 선후배님들이 목표하시는 바를 이루시기를 기원 드리며, 학위과정에서의 배움을 바탕으로 선후배 님들에게 도움이 될 수 있는 동문이 되도록 노력하겠습니다.

끝으로, 제가 학업에 전념할 수 있도록 지원해주시고 응원해주신 소중한 가족 여러분께 깊은 감사의 인사를 드립니다. 힘든 상황에서도 가족들이 있어 버팀목이 되었습니다. 앞으로 기대에 어긋나지 않도록 자랑스러운 아들이 되도록 노력하겠습니다.

2024 년 6 월
김상규 올림

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Deep Learning Approaches for Arrhythmia Screening in Adhesive Patch-Type Wearable Electrocardiographs

Due to their convenience and extended measurement duration, adhesive patch-type wearable electrocardiographs are increasingly utilized for arrhythmia screening. The growing adoption of these devices in clinical settings promises enhanced capabilities for early and accurate detection and treatment of heart diseases. However, the effectiveness of current arrhythmia screening techniques when applied to patch-type wearable electrocardiographs remains uncertain, primarily due to their single-lead structure and susceptibility to noise. This study aims to develop robust methods to improve the classification performance of arrhythmias using these devices.

Firstly, SeqAFNet, a deep learning model, utilized RR interval frames specifically devised for beat-wise atrial fibrillation classification. This model was designed to classify each ECG beat sequentially, based on a recurrent neural network structure. To evaluate its performance not only on the training database but also more broadly, we compared it across three different databases. SeqAFNet demonstrates robust performance in AF classification, aligning with the 2020 European Society of Cardiology guidelines and the IEC 60601-2-47 standard in clinical practice.

To address the problems caused by noise artifacts in wearable electrocardiographs, the SE-ResNet-ViT hybrid model was developed. The SE-ResNet encoder in this model can effectively extract features from ECG data, while the transformer component focuses attention on the noise sections within the 10 second data window. Thanks to this hybrid structure, the proposed model is capable of classifying signals with not only noise alone but also those with various arrhythmias measured alongside the noise.

The methods proposed in this study hold significant promise for advancing the field of arrhythmia detection and management using wearable technologies. These robust and effective deep learning-based techniques could simplify the workload for medical professionals. Furthermore, they contribute to more accurate arrhythmia diagnoses and the early treatment of heart diseases in clinical settings.

Key words: Deep learning, Arrythmia, Atrial fibrillation, Wearable device, Patch-type electrocardiograph, RNN, ViT

REMARK

Chapter 3 of my thesis was published in the IEEE Journal of Biomedical and Health Informatics in June 2024, titled 'SeqAFNet: A Beat-Wise Sequential Neural Network for Atrial Fibrillation Classification in Adhesive Patch-Type Electrocardiographs'.

Chapter 4 of my thesis, titled 'SE-ResNet-ViT Hybrid Model for Noise Classification in Adhesive Patch-type Wearable Electrocardiographs,' was presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in July 2023.

1. Introduction

1.1 Motivation

Arrhythmia refers to an abnormal heart rhythm where the heart beats too slowly, too rapidly, or irregularly [1, 2]. This feature occurs due to the abnormal generation or transmission of electrical signals in the heart. Arrhythmia can range from mild symptoms to life-threatening conditions. Sometimes, they can lead to major complications such as heart failure, stroke, or cardiac arrest [3]. Therefore, the accurate diagnosis and monitoring of arrhythmias are essential for maintaining patients' cardiac health and administering appropriate treatment.

The 12-lead electrocardiogram (ECG) is the golden standard for arrhythmia diagnosis [4]. Each lead records the electrical activity from a specific area of the heart, allowing for a comprehensive assessment of various regions of the heart. The patient typically undergoes this test while lying down, and it is conducted over a short period of time. If arrhythmia is not detected on the 12-lead ECG, however there is a high clinical suspicion, additional work-up may be required [5]. A Holter monitor can be utilized for the early diagnosis of arrhythmia [6]. This device first launched on the market in 1962. To date, it continues to be used for the diagnosis of arrhythmia. However, traditional Holter monitors have a limitation in recording duration. They capture data only for 24–48 hours. Therefore, the short recording duration presents a diagnostic challenge [6]. This limitation is important when detecting such as paroxysmal AF, where abnormal episodes may not occur within the short monitoring period.

To address the limited recording time and inconvenient design of traditional Holter monitors, various adhesive patch-type wearable electrocardiographs have been recently introduced to the market. These adhesive patch-type electrocardiographs can measure ECG up to 14 days with the advancement of hardware technology [7, 8]. Furthermore, these devices are compact and lightweight. Therefore, these devices are more convenient for daily use by patients. Currently, these wearable devices are becoming increasingly standard in the diagnosis of arrhythmias, complementing traditional methods [9]. However, despite these advantages, analyzing data collected from these wearable devices remains challenging due to the extended measurement time, resulting in a substantial increase in the amount of ECG data that must be analyzed. In addition, because of their single-lead structure, these wearable devices can be more vulnerable in terms of signal quality and noise than traditional Holter monitors [10]. Therefore, there is a need for a method capable of effectively analyzing the large volume of data and complex signals measured by these devices.

Deep learning methodologies have revolutionized the accurate classification and diagnosis of arrhythmias [11, 12]. These methods exhibit exceptional performance in learning and predicting complex data patterns, offering higher accuracy and efficiency compared to traditional approaches. Recent advances have led to the development of numerous deep learning-based models for arrhythmia classification [13]. However, many of these studies have focused on ECG data derived from 12-lead ECGs or Holter monitors [14-16]. As a result, the compatibility and performance of these models for analyzing arrhythmias in ECG signals recorded by adhesive patch-type wearable electrocardiographs remain insufficiently investigated. Furthermore, the suitability of these deep learning models for arrhythmia screening in clinical practice, or their potential for real-time arrhythmia monitoring, has not been adequately considered.

The objective of this study is to develop and comprehensively evaluate deep learning models for their effective application in adhesive patch-type wearable electrocardiographs. It concentrates on the development and evaluation of these models for early arrhythmia screening in clinical settings.

1.2 Contributions of this Dissertation

In this dissertation, a deep learning-based atrial fibrillation (AF) classification method, which is robust for inter-device and inter-patient variability, and a noise classification method, which is conducted in real-world clinical settings, are proposed.

Obtaining a large dataset and labeling the data requires significant time and cost. If a new device is used instead of the previously used one, the performance of the deep learning model on data measured by the new device cannot be guaranteed. This is because the data from the new device may differ from the data used previously. Recently, various types of Long-term continuous cardiac monitoring (LTCM) devices have been introduced to the market. It is important to note the potential differences between the data measured by newly released LTCM devices and those obtained from Holter or 12-lead electrocardiograms. Such differences among devices may necessitate additional effort and cost in preparing data for training a new deep learning model.

Noise signals can hinder the accurate classification of arrhythmias in electrocardiogram signals, particularly in single-lead wearable electrocardiographs, which may be more vulnerable to noise. However, there is a lack of prior research utilizing real-world data due to the difficulty in obtaining data where noise signals are labeled along with electrocardiogram signals measured from patients with arrhythmias.

The contribution of this dissertation is outlined as follows. First, a deep learning model utilizing RR intervals, which can effectively reflect the irregular characteristics of AF, is proposed. This model aligns with the current standards of multi-level deep learning architectures for time series data, where local features are extracted and long-term dependencies across the entire series are captured. Seconds, the model evaluation was conducted using cross-validation across different databases and devices, specifically focusing on inter-database and devices, and intra-database variations to assess the robustness of the electrocardiogram model. Finally, a deep learning model was developed utilizing electrocardiogram signals, including arrhythmias signals, using patch-type electrocardiographs in real-world clinical settings.

1.3 Overview

The overview provides a brief introduction to the various chapters of this study.

Chapter 2 introduces the medical significance of adhesive patch-type electrocardiographs and deep learning techniques utilized in this study. It presents architectures and concepts of Residual Network (ResNet) and Recurrent Neural Network (RNN).

Chapter 3 analyzes deep learning-based methods for classifying atrial fibrillation (AF) using data from adhesive patch-type wearable electrocardiographs. The chapter highlights the innovative SeqAFNet, a beat-wise sequential neural network architecture designed for precise AF classification. This model is particularly adept at processing the continuous and irregular rhythms characteristic of AF, leveraging its sequential input and output processing capabilities. The chapter also addresses the challenges of interpreting ECG data collected from single-lead, adhesive patch-type wearables, which often involve extended measurement periods. It demonstrates how SeqAFNet effectively classifies atrial fibrillation, thus enhancing the practical utility of these wearable devices in clinical settings.

Chapter 4 focuses on classifying noise signals caused by patient movements while using adhesive patch-type wearable electrocardiographs. Such noise signals can lead to the misclassification of arrhythmia signals. To address this issue, the chapter introduces the SE-ResNet-ViT hybrid model, which is specifically tailored to differentiate between noise and arrhythmia signals.

2. Background Ambulatory cardiac monitors

Advancements in wearable technology have significantly enhanced the feasibility and effectiveness of ambulatory cardiac monitors (ACM). These technologies enable long-term, non-invasive, real-time tracking of heart rhythms and other relevant cardiac parameters, improving patient compliance and comfort. Moreover, the integration of algorithms and artificial intelligence in these systems has improved the precision of data analysis, allowing for the detection of subtle and transient cardiac events that might otherwise go unnoticed in traditional episodic monitoring settings.

One of the ACMs, adhesive patch-type electrocardiographs are among the most actively used devices in the market, their unique blend of convenience, efficacy, and healthcare cost benefits. These lightweight, wearable devices adhere directly to the skin, allowing for continuous monitoring without the need for bulky equipment or frequent medical visits.

Since their introduction to the market, adhesive patch-type electrocardiographs have not been evaluated for variations in monitoring strategy, clinical outcomes, and healthcare utilization in patients undergoing ambulatory monitoring.

Recently, the "Cardiac Ambulatory Monitor Evaluation of Outcomes and Time to Events (CAMELOT) study," a retrospective cohort study utilizing the full (100%) Medicare Fee-For-Service sample, including inpatient and outpatient medical claims, was conducted to evaluate the clinical effectiveness of adhesive patch-type electrocardiographs [17]. In this study, when comparing the new diagnosis of specified arrhythmia between adhesive patch-type electrocardiographs and Holter monitors, the Odds Ratio was 0.5, indicating a higher arrhythmia detection rate with adhesive patch-type electrocardiographs. Furthermore, with an Odds Ratio of 1.35 for retesting any ACM within 180 days, it suggests that costs associated with retesting can be reduced. However, during the cohort period, differences in the Odds Ratio for new diagnosis of specified arrhythmia and ACM retest were observed among the five groups of adhesive patch-type electrocardiographs billed to Medicare. This suggests that even when using adhesive patch-type electrocardiographs, considerations should be given to the algorithms or deep learning technologies used for diagnosing arrhythmias, as they can influence the clinical utility of adhesive patch-type electrocardiographs in a real-world setting.

In this dissertation, the MEMO Patch, a single-lead adhesive patch-type ambulatory electrocardiograph shown in Figure 2.1, used to record ECGs from patients participating in the clinical trial. This device is approved by the Ministry of Food and Drug Safety (MFDS) in the Republic of Korea. This device is capable of operating for up to 14 days, recording ECG at a sampling rate of 250 Hz and with a 12-bit resolution.

Figure 2.1. Patch-type ambulatory electrocardiograph, MEMO Patch, HUINNO Co., Ltd.

2.2 Residual Network (ResNet)

A residual network (ResNet) is one of the neural network architectures that has significantly advanced the field of computer vision and deep learning. It was introduced by Kaiming He et al. in their 2015 paper. Deep learning models encounter the issue of vanishing gradients as the number of layers increases, a challenge that becomes more noticeable in deeper networks. The fundamental idea of this architecture is to facilitate learning while making the network deeper.

The most significant innovation of ResNet is the residual block. This block directly adds the input to the output of a layer, also known as a skip connection. Through this, the network learns only the residual, or the modifications needed between the input and output. ResNet utilizes convolutional layers to extract features from input. These layers learn important information from data through a spatial hierarchical structure. In addition, batch normalization stabilizes and accelerates the learning process by normalizing the input across layers. Typically, the ReLU (Rectified Linear Unit) activation function is employed. ReLU introduces non-linearity, enabling the network to learn more complex patterns.

2.3 Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN) is a neural network specifically designed for processing sequential and time-series data. It is widely used in fields such as natural language processing, speech recognition, and time-series forecasting. The key feature of an RNN is its internal memory, which retains previous information and combines it with current input to generate output. This allows the model to learn patterns and relationships in data over time. RNNs have a recurrent structure, allowing them to reflect information from previous time steps. This assists the network in considering past data when determining current output. Basic RNNs face the problem of long-term dependencies, meaning the network encounters difficulties in retaining and utilizing older information. To solve this issue, advanced RNN models such as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Units) have been developed. These models can more effectively handle long-term dependencies through their complex gating mechanisms.

RNN has various types of input and output sequences leading to different architectures like many-to-many, many-to-one, and one-to-many. Each of these architectures leverages the ability of RNNs to process sequential data, making them suitable for a wide range of applications in natural language processing, computer vision, and beyond.

3. SeqAFNet: A Beat-Wise Sequential Neural Network for Atrial Fibrillation Classification in Adhesive Patch-Type Wearable Electrocardiographs

3.1 Motivation

Atrial fibrillation (AF) characterized by rapid and irregular contractions in the atria is known as the most common type of arrhythmia [18]. When the atria contract irregularly, it can disrupt the blood flow to the ventricles, subsequently increasing the risk of heart failure [19]. These irregular contractions may also lead to the formation of a thrombus, which can cause a stroke, a major complication associated with arrhythmia [3]. AF may initially occur with mild symptoms or remain asymptomatic but over time, these symptoms can gradually become worse [20]. For diagnosing AF, the 12-lead ECG is gold standard [4]. The 12-lead ECG is capable of identifying an irregular rhythm, detecting the absence of observable P waves, and the presence of F waves [5]. If AF is not identified through this procedure, yet there is a strong clinical suspicion of AF, further investigation may be required [5]. In such cases, a Holter monitor is employed for the detection of arrhythmias, including AF [6].

The European Society of Cardiology (ESC) categorizes AF into various types based on its duration and presentation, first diagnosed, paroxysmal, persistent, long-standing persistent, and permanent. These classification schemes are very useful in guiding appropriate treatment [5]. The diagnosis and classification of AF are essential for patient management. These classifications can guide the choice of treatment methods. While Holter monitors are helpful in diagnosing AF, their limited recording duration, typically capturing data for only 24–48 hours, encounters a diagnostic challenge [6]. This shortcoming is particularly significant in detecting paroxysmal AF where episodes may not occur during the monitoring period [6]. To overcome these recording duration constraints and improve patient convenience in daily life, various adhesive patch-type wearable electrocardiographs have been developed and introduced to the market. These wearable devices are now widely used for diagnosing AF [7-9]. Despite their various benefits, wearable electrocardiographs encounter several challenges. The extended measurement duration significantly amplifies the amount of ECG data needing analysis. Furthermore, their single-lead structure makes these devices potentially more susceptible to signal quality problems and noise, in contrast to the more robust Holter monitors [10]. Consequently, developing a more robust and effective method tailored for wearable devices to process and analyze the substantial volume and complexity of these signals is essential. However, it is remarkable that numerous previous studies focusing on classifying AF using ECG signals have primarily used data from 12-lead or Holter monitors [14, 15, 21]. Therefore, the effectiveness of these traditional methods in analyzing ECG signals captured by wearable electrocardiographs remains unverified.

Analysis on a fine-grained, sample-wise, or beat-wise basis may enhance the classification of ECG signals, compared to a window-based approach [14, 16, 22]. Particularly for rhythm-type arrhythmias like atrial flutter, supraventricular tachyarrhythmia, and AF, analyzing ECG signals on a beat- or sample-wise basis offers a more comprehensive examination at the change in rhythm [23, 24]. After the 4th China Physiological Signal Challenge in 2021 (CPSC2021), several studies have conducted detailed sample- or beat-wise analyses of AF [16, 22]. However, they encounter limitations in objectively comparing their performance [25]. The underlying reason for this is that the testing dataset from the challenge has not been released to the public, and the scoring method was tailored specifically for ranking participants in the challenge. Additionally, another complication arises from the fact that many studies have not employed diverse datasets for both their training and testing phases [14, 26]. In other words, many studies have relied on a single dataset for both training and evaluating their models. Previous research has predominantly utilized the MIT-BIH Atrial Fibrillation Database (AFDB), and some studies have also made use of the Long-Term AF Database (LTAFDB) or CPSC2021 training datasets [16, 22]. In a recent study conducted by Yating Hu et al., it was observed that their model, which was trained on one dataset and evaluated on another, experienced a reduction in classification performance [22]. While the model achieved a high F1-score of 0.985 on the CPSC2021 training dataset, this score significantly decreased when the model was tested on different datasets. Specifically, the F1-score dropped to 0.9 on the AFDB and further declined to 0.74 on the LTAFDB after the process of label merging. Another issue arises from the fact that some studies did not specify whether the patients (or recordings) used in the training dataset were excluded from the testing dataset. The absence of clear separation between patients (or recordings) in the training and testing datasets can lead to overfitting [27, 28]. Because of these limitations, there is an increased risk of models inaccurately classifying unseen ECG signals, especially when these signals are compared to recordings obtained from wearable electrocardiographs[27-29].

To overcome the limitations, this study focused on developing a many-to-many recurrent neural network (RNN)-based model for the sequential classification of AF, utilizing the R-peak intervals of ECG signals. This sequential approach allows the model to analyze each ECG beat at individual time steps. Thanks to its many-to-many configuration, the model is particularly adept at accurately classifying irregular AF patterns. Moreover, to take advantage of the many-to-many structure, the input to the model is successively strided by one beat, and the ensuing inferences are ensembled to produce a merged output. The ECG data, including AF were collected from patients who had either been previously diagnosed with arrhythmia or exhibited symptoms of suspected arrhythmia, using the MEMO Patch. The proposed model, trained using the AFDB and LTAFDB, was subjected to a performance evaluation [30, 31]. This evaluation involved comparing data from adhesive patch-type wearable electrocardiographs against those from public databases. Furthermore, to prevent overfitting, a strict separation of recordings was maintained between the training and testing datasets.

3.2 MethodsDatabases

The model was trained with data from the AFDB and LTAFDB public databases [30, 31]. In order to evaluate the model's performance on data obtained from wearable electrocardiographs, we used ECG signals from a clinical trial that utilized the MEMO Patch, depicted in Figure 3.1, an adhesive patch-type wearable electrocardiograph manufactured by HUINNO Co., Ltd [32]. Table 3.1 presents detailed statistical information for both the public databases and the patch database.

Figure 3.1. MEMO Patch, an adhesive patch-type wearable electrocardiograph, manufactured by HUINNO Co., Ltd.

Table 3.1. Statistics for each database.

	AFDB	LTAFDB	Patch
Number of participants	-	-	17
Age (years)	-	-	58.9 (42–69)
Male sex, n (%)	-	-	13 (76.5)
Number of records (chunks)	23	82[a]	57
Duration of records (chunks)	9.25–10.23 h	6.13–26.35 h	0.5–24 h
Mean (σ) duration of records (chunks)	10.19 (0.2) h	23.32 (2.75) h	12.08 (7.59) h
Beats in records (hunks)	34,837–61,915	31,190–184,809	2322–137321
Episodes with paroxysmal AF	291	7,355	593
AF episodes that tare <30 s	65	4,703	150
Paroxysmal AF duration	1.7–36001.8 s	0.3–88378.8 s	2.7–51877.5 s
Mean (σ) paroxysmal AF duration	1132.8 (4292.3) s	489.5 (4550) s	811.3 (3799.4) s

[a] Record 64, 113 was excluded.

AFDB includes 23 long-term ECG recordings from human subjects with AF (mostly paroxysmal) [30]. The duration of each recording is 10 h, and the individual recordings contained two ECG signals sampled at 250 Hz with 12-bit resolution over a range of ±10 mV. To evaluate the intra-database performance of the model, 23 records were divided into 5 folds. To prevent patient overlap within a fold, training and testing data were split based on the record name. The patients comprised in the testing data for each fold are presented in Table 3.2.

Table 3.2. Composition of 5-fold for AFDB.

Fold number	**Record names**	**Number of records**
Fold 1	'04015','04043','04048', '04126'	4
Fold 2	'04746','04908','04936', '05091'	4
Fold 3	'05121','05261','06426','06453', '06995'	5
Fold 4	'07162','07859','07879','07910', '08215'	5
Fold 5	'08219','08378','08405','08434', '08455'	5
	Total	23

LTAFDB includes 84 long-term ECG recordings from human subjects with sustained paroxysmal AF [31]. Each record included two simultaneously recorded ECG signals digitized at 128 Hz with 12-bit resolution over a 20-mV range. Although the record durations varied, they typically lasted between 24–25h. The record named 64 was excluded due to missing annotation and, the record named 113 was excluded because the ECG signal occasionally vanished. To evaluate the intra-database performance of the model, 82 records were divided into 5 folds. The patients comprised in the testing data for each fold are presented in Table 3.3.

Table 3.3. Composition of 5-fold for LTAFDB.

Fold number	**Record names**	**Number of records**
Fold 1	'00', '01', '03', '05', '06', '07', '08', '10', '11', '12', '13', '15', '16', '17', '18', '19'	16
Fold 2	'20', '21', '22', '23', '24', '25', '26', '28', ‘30', '32', '33', '34', '35', '37', '38', '39'	16
Fold 3	'42', '43', '44', '45', '47', '48', '49', '51', '53', '54', '55', '56', '58', '60', '62', '65'	16
Fold 4	'68', '69', '70', '71', '72', '74', '75', '100', '101', '102', '103', '104', '105', '110', '111', '112', '114'	17
Fold 5	'115', '116', '117', '118', '119', '120', '121', '122', '200', '201', '202', '203', '204', '205', '206', '207', '208'	17
	Total	82

The patch dataset was collected in a multi-center clinical trial conducted at Seoul National University Bundang Hospital and Korea University Hospital (IRB numbers: B-2105/686-002 and 2021AN0247, respectively). In total, 149 participants were included in the clinical trial. Among them, 17 presented with AF. The patients visited the hospital to get the device attached to their chests and returned after 14 days to hand it back. Within 14 days, lead-off in the ECG signal could occur if the patients detach the device from their body during showering or due to the inadequate contact of electrodes. To exclude the lead-off, we only extracted the well-recorded ECG signals between the lead-off sections, and these sections were referred to as chunk. Based on the data of 17 participants with AF, we created 57 chunks that may contain either AF, non-AF, or both types of signals. Procedure for making the chunk from the patch dataset depicted in Figure 3.2. All recordings were sampled at 250 Hz with 12-bit resolution using MEMO Patch.

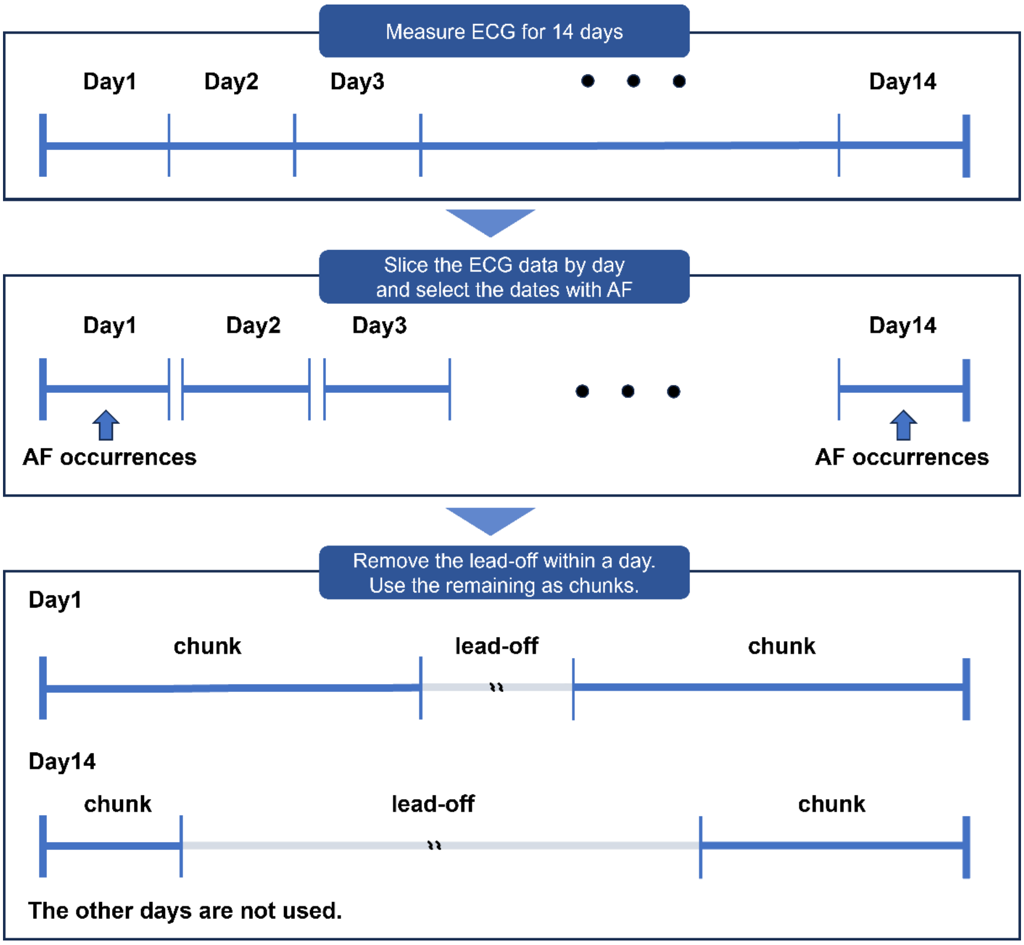

Figure 3.2. Procedure for making the chunk from the patch dataset.

3.2.2 Preprocessing

The RR interval (RRI) of the ECG signal served as the input for our model. Previous research has employed the RRI as a characteristic to effectively depict the irregular or rapid rhythm of AF [14, 16, 21, 28, 30]. RRIs were calculated using beat annotations from AFDB and LTAFDB. For the patch dataset, beats were identified using the MEMO Care software provided by HUINNO Co., Ltd. Subsequently, the RRIs were computed based on the number of samples between a current R peak and its preceding R peak. The LTAFDB, having a different sampling rate of 128 Hz compared to other datasets, had its RRIs up-sampled to 250 Hz. Each R peak was labeled as either non-AF or AF.

According to the 2020 ESC Guidelines for AF diagnosis, a standard 12-lead ECG recording or a single-lead ECG tracing lasting longer than 30 s that shows cardiac rhythm without discernible repeating P waves and exhibits irregular RR intervals (provided there is no atrioventricular conduction impairment) is considered diagnostic of clinical AF [5]. Consequently, AF durations of less than 30 s were relabeled as non-AF. Table 3.4 displays the statistics after applying preprocessing for AF in each database.

Table 3.4. AF in each database applying preprocessing.

	AFDB	LTAFDB	Patch
With Paroxysmal AF	226	26,51	443
Paroxysmal AF duration	30.8–36001.8 s	30–88378.8 s	30–51877.5 s
Mean (σ) paroxysmal AF duration	1454.7 (4822.7) s	1342.6 (7503.4) s	1079.8 (4363.3) s

3.2.3 Sequential input

A frame consisting of 20 RRIs was generated for embedding into the input layer of our model, starting with the initial 20 RRIs in the first frame. To make a sequential input structure, the subsequent frame was shifted by one ECG beat. Thus, the second frame contains the 2nd to the 21st RRIs. Each frame received a label depending on the quantity of AF beats present. If this number was equal to or exceeds a certain threshold, the frame was labeled AF. If not, it was labeled as non-AF. This threshold was determined to be 11, more than half of the 20 RRIs per frame, and the label was placed at the central ECG beat of the frame. Figure 3.3 illustrates the process for creating sequential input from the ECG signal.

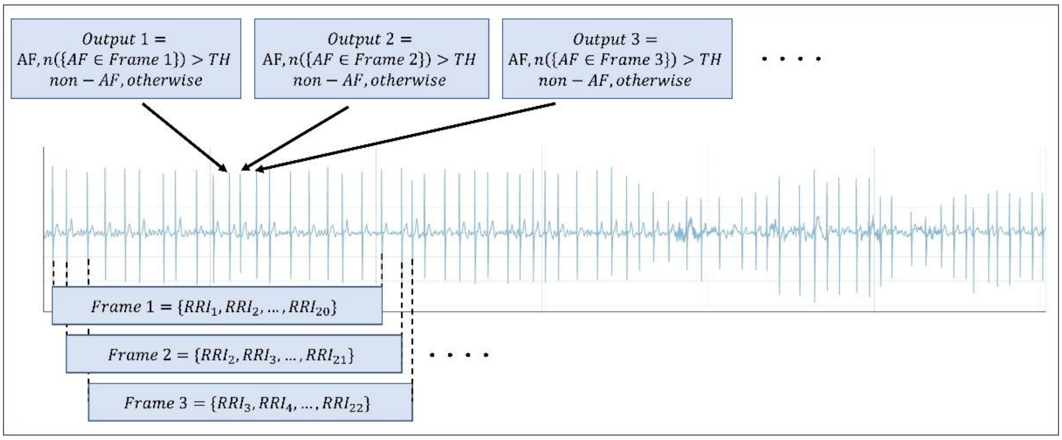

Figure 3.3. Method of generating sequential input from the ECG signal.

3.2.4 SeqAFNet

The proposed SeqAFNet employs a two-stage bidirectional RNN featuring a many-to-many structure, specifically designed for handling sequential data like time series and sequence data [33]. A distinct attribute of RNNs is their capacity to integrate information from past data with the current input, offering significant advantages for predicting time-series challenges [33, 34]. However, as the sequence length increases, RNNs may encounter vanishing or exploding gradient problems [35]. To address these challenges, both the long short-term memory (LSTM) and its simplified version, the gated recurrent unit (GRU), have been developed [36]. This model incorporates bidirectional GRU cells in its initial and subsequent stages, named a local-wise RNN layer and a sequence-wise RNN layer, respectively. Figure 3.4 illustrates overall architecture of SeqAFNet.

Local-wise RNN layer: The local-wise RNN layer comprises 30 GRU cells, each cell configured to handle a singular frame consisting of 20 time steps. Each individual time step in the GRU cell is designed to process a one RRI. This layer is aims to identify short regional arrhythmias utilizing only the local RRI frames. The output from each GRU cell has 20 time steps, including both the forward and backward hidden states.

Flatten layer: Positioned after the local-wise RNN layer, the flatten layer converts its multi-dimensional output into a one-dimensional vector. This reshaping facilitates efficient data transmission to the following fully connected layer, thus enhancing the model’s ability for high level representations from the sequential data.

Fully connected layer: In this model, two fully connected layers are integrated, each reducing the input and output sizes by a factor of four. This process of feature abstraction reduces the size of the model and ensures that it retains only the most important features, which can potentially enhance generalization capability.

Multi-head attention: Prior to its integration into the sequence-wise RNN layer, the multi-head attention layer was utilized, offering several advantages. The multi-head attention mechanism enables the model to simultaneously concentrate on different segments of the input sequence, effectively identifying different relationships and dependencies. This attention enhances the model's ability to recognize patterns and handle long-range dependencies, which can be especially beneficial for processing complex sequences.

Sequence-wise RNN layer: Features extracted by the preceding local-wise layer were input to the sequence-wise RNN layer. By utilizing local RRI patterns from each period, this layer enhances the effective prediction of arrhythmia, particularly in AF where irregular rhythms occur sequentially and continuously.

Fully connected layer: To maintain the intrinsic sequential output features of the many-to-many structure, independent processing is required. The fully connected layer is comprised of two distinct stages, and each sequence-wise layer operates independently, without interconnection. The preceding stage reduces the output shape to half of the input size, and the subsequent stage adjusts the output to a size of two. This configuration enables the binary categorization of each time step into either AF or non-AF, employing a sigmoid function.

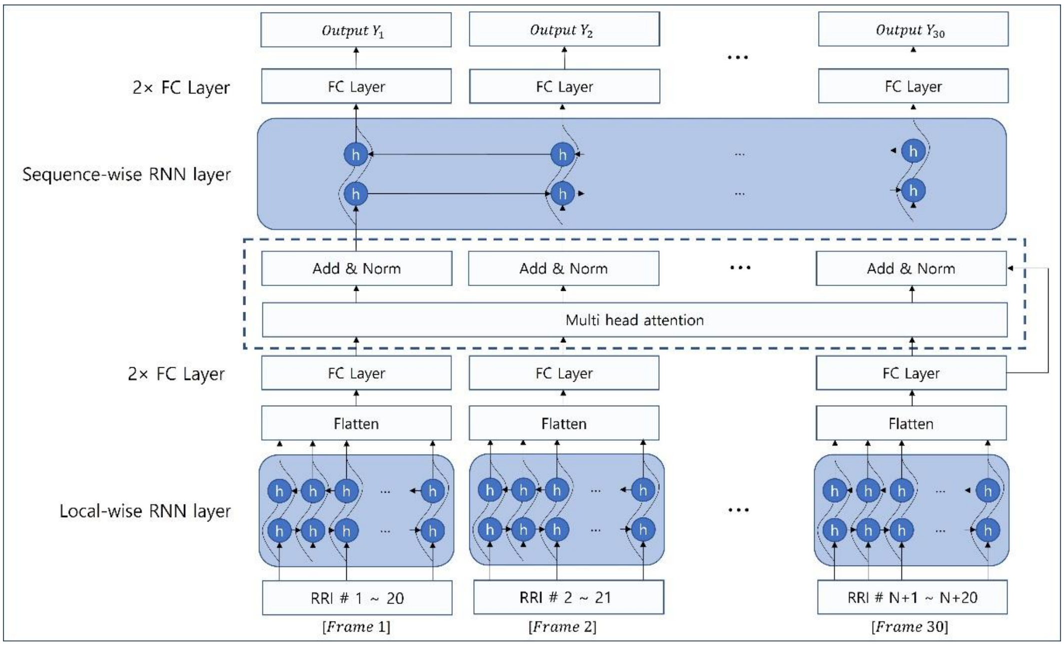

Figure 3.4. Illustration of SeqAFNet.

3.2.5 Ensemble decision and post-processing

To take advantage of SeqAFNet's many-to-many structure, an ensemble decision strategy, depicted in Figure 3.5, was employed. Initial inference of SeqAFNet returns 30 outputs, denoted as $\{Y_i\}$, each corresponding to the R peak of the ECG at time $t(n)$. In next inferences, the input frames are shifted by one beat to progress to the subsequent temporal sequence. This progression prompts the model to predict the next series of 30 outputs, designated as $\{Y_{i+1}\}$. At the specific moment of $t(n)$, corresponding to the single R peak in the ECG, there exist 30 outputs denoted as $Y_i[t(n)]$ from different temporal sequences. Every distinct output $Y[t(n)]$ within the assembled set $\{Y_i\}$ represents a prediction of AF. At the position $t(n)$, the classification is determined as AF if the count of AF predictions is equal to or exceeds the predefined threshold. The threshold is designated at 16. When the predictions do not reach this threshold, the classification is settled on non-AF. Following the ensemble decision approach, AF predictions with durations less than 30 s are relabeled as non-AF.

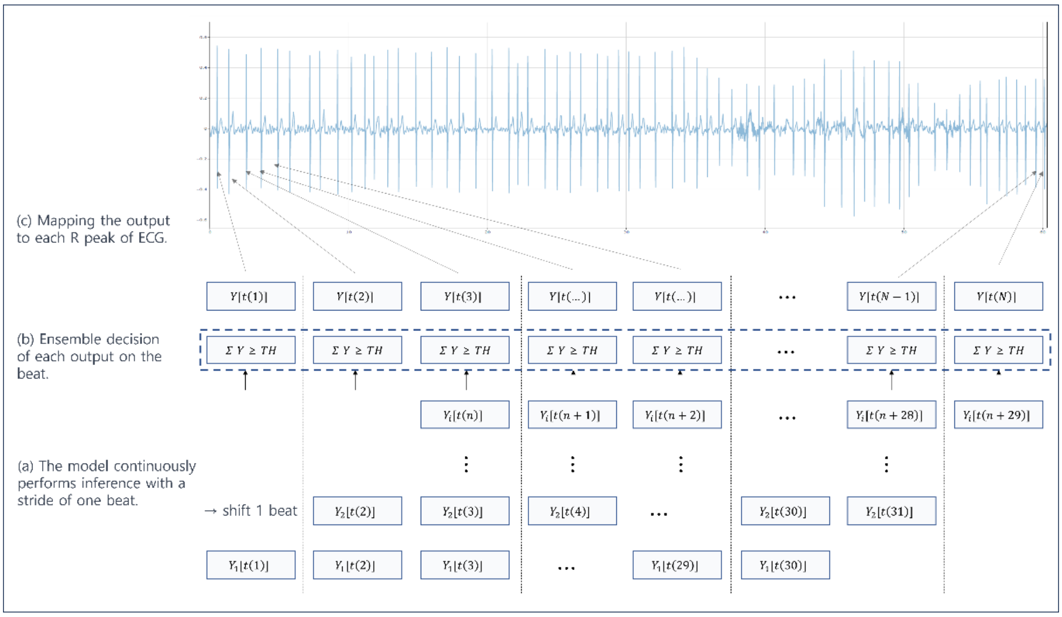

Figure 3.5. Illustration of ensemble decision process.

3.2.6 Metrics

The model's performance was assessed using two distinct metrics. The first metric focused on the beat-by-beat classification performance of AF, employing measures such as accuracy, precision, sensitivity, specificity, and F-1 score. The second metric, aligned with Subclause 201.12.1.101.1.5.3 of the IEC 60601-2-47 standard, aimed at evaluating the sensitivity and positive predictive value specific to AF duration [37].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.4}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3.5}$$

$$AF\ duration\ Se = \frac{Duration\ of\ overlap}{Duration\ of\ reference - Annoted\ AF} \tag{3.6}$$

$$AF\ duration + P = \frac{Duration\ of\ overlap}{Duration\ of\ algorithm - Annoted\ AF} \tag{3.7}$$

3.3 Results

3.3.1 Experiment setups

All experiments were conducted using the RTX3090 GPU within the PyTorch framework. We employed the Adam optimizer, beginning with an initial learning rate of $1e^{-3}$ and a weight decay of $1e^{-5}$. In the optimization process, a binary cross-entropy loss function was utilized. The model training continued for up to 15 epochs with a batch size of 256, and the optimal epoch was selected based on the performance in the validation dataset. Table 3.5 shows the detailed parameters of the model. The notation '30 ×' in front of the layers was used to indicate that there are 30 layers in parallel.

Table 3.5. Detailed parameters of the proposed model.

Layer	Input shape	Output shape	Parameter
Input	(30, 20, 1)	30 × (20, 1)	
30 × GRU	30 × (20)	30 × (20, 128)	Hidden size: 64
30 × Flatten	30 × (20, 128)	30 × (1, 2560)	
30 × FC layer	30 × (1, 2560)	30 × (1, 640)	Activation: ReLU
Dropout	-	-	Rate: 0.1
30 × FC layer	30 × (1, 640)	30 × (1, 160)	Activation: ReLU
Multi-head attention	(30, 160)	(30, 160)	Num head: 4
Add & Layer Norm	-	-	
GRU	(30, 160)	(30, 128)	Hidden size: 64
30 × FC layer	(30, 128)	30 × (1, 64)	Activation: ReLU
Dropout	-	-	Rate: 0.1
30 × FC layer	30 × (1, 64)	30 × (1, 2)	Activation: sigmoid

3.3.2 Comparison of performance based on the number of RRIs per frame

To determine the optimal number of RRIs for each frame, we evaluated the performance using frames containing 10, 15, 20, 25, and 30 RRIs, respectively. We set the threshold for determining the frame's label at a level exceeding half the number of RRIs. Table 3.6 shows performance of model at the number of RRIs per frame. We set the threshold for determining the ensemble decision at 16. The highest performance on the Patch database was when the frame contained 20 RRIs. However, there were no significant differences with other values.

Table 3.6. Performance of model at the number of RRIs per frame.

# RRIs per frame	Label TH	Train	Test	Accuracy	Precision	Sensitivity	Specificity	F1 score
30	≥ 16	LTAFDB	Patch	0.983	0.975	0.976	0.988	0.975
25	≥ 13	LTAFDB	Patch	0.981	0.977	0.969	0.992	0.973
20	≥ 11	LTAFDB	Patch	0.986	0.981	0.979	0.992	0.98
15	≥ 8	LTAFDB	Patch	0.969	0.951	0.961	0.976	0.959
10	≥ 6	LTAFDB	Patch	0.983	0.969	0.983	0.983	0.976

3.3.3 Comparison of performance based on the threshold for the label of frame

To determine the optimal number of thresholds for the label of frame, we evaluated the performance using threshold 1, 6, 11, 16, 20. If the number of AF beats within a frame exceeded the threshold, the frame was labeled as AF; otherwise, it was labeled as non-AF. The number of RRIs per frame was fixed at 20, which demonstrated the best performance in previous experiments. Table 3.7 shows performance of model at the threshold for label of frame. We set the threshold for the ensemble decision at 16. The highest performance on the Patch database was when the threshold at 11.

Table 3.7. Performance of model at the threshold for label of frame.

Label TH	Train	Test	Accuracy	Precision	Sensitivity	Specificity	F1 score
≥ 20	LTAFDB	Patch	0.981	0.975	0.971	0.989	0.973
≥16	LTAFDB	Patch	0.981	0.968	0.979	0.983	0.973
≥ 11	LTAFDB	Patch	0.986	0.981	0.979	0.992	0.98
≥ 6	LTAFDB	Patch	0.984	0.97	0.984	0.983	0.977
≥ 1	LTAFDB	Patch	0.965	0.934	0.975	0.957	0.952

3.3.4 Comparison of performance based on the threshold for determining the ensemble decision

To determine the optimal number of thresholds for the ensemble decision, we evaluated the performance using threshold 6, 11, 16, 21, 26. We set the number of RRIs per frame at 20 and, threshold for label at 11, which demonstrated the best performance in previous experiments. Table 3.8 shows performance of model at the threshold for ensemble decision. We observed that the performance at thresholds of 11 and 16 was nearly identical for the Patch database. Due to the negligible difference between the two values, we opted to use 16.

Table 3.8. Performance of model at the threshold for ensemble decision.

Ensemble TH	Train	Test	Accuracy	Precision	Sensitivity	Specificity	F1 score
≥ 26	LTAFDB	Patch	0.979	0.983	0.957	0.997	0.969
≥21	LTAFDB	Patch	0.985	0.984	0.973	0.995	0.978
≥ 16	LTAFDB	Patch	0.986	0.981	0.979	0.992	0.98
≥ 11	LTAFDB	Patch	0.986	0.977	0.983	0.988	0.98
≥ 6	LTAFDB	Patch	0.982	0.967	0.983	0.982	0.975

3.3.5 Intra-database performance of the AF classification

To assess the model's performance within each AFDB and LTAFDB, we implemented an inter-patient 5-fold cross validation. Generally, k-fold cross validation was sufficient for assessing the model's generalization ability on unseen data. However, ECG data from a single patient exhibit almost similar morphology and patterns that repeat. Therefore, if ECG data from a single patient are mixed across the training and testing sets within each fold, the model may learn to recognize the specific patterns of that patient rather than generalized patterns of ECG. To prevent this overfitting problem, we utilized record names to separate the training and testing sets within each fold. The patients comprised in the testing sets for each fold are presented in Table 3.2 and Table 3.3. Table 3.9 shows the results of cross validation of AFDB and Table 3.10 shows the results of cross validation of LTAFDB.

Table 3.9. Inter-patient 5-fold cross validation score of AFDB.

Fold	Accuracy	Precision	Sensitivity	Specificity	F1 score	duration Se	duration +P	FN	FP	TN	TP
Fold 1	0.962	0.885	0.909	0.975	0.896	0.846	0.741	164,892	4,161	3,530	15,735
Fold 2	0.976	0.977	0.973	0.987	0.975	0.965	0.983	121,874	1,498	3,332	73,172
Fold 3	0.971	0.966	0.967	0.961	0.967	0.97	0.954	119,596	4,769	3,368	112,618
Fold 4	0.979	0.951	0.972	0.973	0.961	0.969	0.989	55,066	1,504	5,367	173,931
Fold 5	0.896	0.979	0.98	0.978	0.979	0.983	0.977	142,374	3,191	2,229	114,827
Average	0.971	0.952	0.96	0.975	0.956	0.947	0.929	100,760	3,025	3,565	98,057
Std	0.006	0.035	0.026	0.008	0.03	0.051	0.095	36,696	1,342	1,013	52,242

Abbreviations: Std, Standard deviation; FN, False Negative; FP, False Positive, TN, True Negative; TP, True Positive.

Table 3.10. Inter-patient 5-fold cross validation score of LTAFDB.

Fold	Accuracy	Precision	Sensitivity	Specificity	F1 score	duration Se	duration +P	FN	FP	TN	TP
Fold 1	0.988	0.988	0.989	0.981	0.988	0.996	0.967	886,208	5,451	5,586	796,196
Fold 2	0.957	0.957	0.956	0.953	0.957	0.955	0.951	666,525	30,407	37,483	754,181
Fold 3	0.983	0.985	0.978	0.959	0.981	0.995	0.972	580,539	16,798	8,198	1,087,356
Fold 4	0.958	0.96	0.955	0.933	0.957	0.974	0.932	761,873	35,519	43,565	1,119,435
Fold 5	0.986	0.988	0.98	0.963	0.984	0.996	0.981	525,232	17,899	8,980	1,253,081
Average	0.974	0.976	0.972	0.958	0.973	0.983	0.961	684,075	21,215	20,762	1,002,050
Std	0.014	0.014	0.014	0.016	0.014	0.016	0.017	128,972	10,659	16,288	193,849

Abbreviations: Std, Standard deviation; FN, False Negative; FP, False Positive, TN, True Negative; TP, True Positive.

3.3.6 Inter-database performance of the AF classification

To assess the model's performance with the unseen data that was measured from different devices, we implemented inter-database cross validation. The model was trained separately on the AFDB and LTAFDB databases, and then tested on the LTAFDB and AFDB, respectively. Additionally, Patch dataset is used to evaluate its performance. Table 3.11 shows the results of inter-databases performance of AF classification. To enhance performance on the Patch dataset, all parameters in this paper have been optimized specifically for the Patch dataset. Finally, we trained the model using the entirety of both the AFDB and LTAFDB databases and then evaluated its performance on the Patch dataset. However, we obtained lower performance when using both AFDB and LTAFDB for training, compared to using only LTAFDB.

Table 3.11. Inter-database performance of LTAFDB.

Train	Test	Accuracy	Precision	Sensitivity	Specificity	F1 score	duration Se	duration +P	FN	FP	TN	TP
AFDB	LTAFDB	0.961	0.96	0.96	0.954	0.96	0.958	0.958	3,404,205	122,246	207,163	4,906,898
AFDB	Patch	0.974	0.973	0.953	0.992	0.963	0.932	0.967	2,332,601	18,360	59,097	623,957
LTAFDB	AFDB	0.975	0.974	0.975	0.97	0.975	0.98	0.967	600,480	18,445	10,782	497,327
LTAFDB	Patch	0.986	0.981	0.979	0.992	0.98	0.97	0.97	2,332,222	18,739	23,187	659,867
AFDB + LTAFDB	Patch	0.978	0.981	0.957	0.996	0.968	0.943	0.982	2,341,282	9,679	55,607	627,447

Abbreviations: FN, False Negative; FP, False Positive, TN, True Negative; TP, True Positive.

3.3.7 Comparison of performance before and after applying ensemble decision

For the ensemble decision, ECG was shifted by 1 beat at a time for inference. In the performance experiment without applying the ensemble decision, ECG was shifted by 30 beats at a time without overlap. Table 3.12 shows performance of model before and after applying ensemble decision. When the batch size is 256, it takes approximately 5.76 ms for the model to perform one inference (CPU: Intel i9-12900K, GPU: RTX3090). This means that 30 ECG beats take 22.5 us and a single ECG beat takes 0.75us. Since the ensemble decision uses the results of 30 inferences for a single ECG beat, it takes 22.5 us.

Table 3.12. Performance of model before and after applying ensemble decision.

Ensemble Decision	Train	Test	Accuracy	Precision	Sensitivity	Specificity	F1 score
X	AFDB	LTAFDB	0.956	0.955	0.956	0.955	0.955
O	AFDB	LTAFDB	0.961	0.96	0.96	0.954	0.96
X	AFDB	Patch	0.97	0.969	0.945	0.991	0.956
O	AFDB	Patch	0.974	0.973	0.953	0.992	0.963
X	LTAFDB	AFDB	0.967	0.966	0.967	0.964	0.966
O	LTAFDB	AFDB	0.975	0.974	0.975	0.97	0.975
X	LTAFDB	Patch	0.98	0.974	0.97	0.989	0.972
O	LTAFDB	Patch	0.986	0.981	0.979	0.992	0.98

3.4 Discussion

In this work, we have developed SeqAFNet, an RNN-based architecture with a many-to-many approach, specifically designed for sequentially classifying AF utilizing the RRIs of ECG signals. The novel configuration of input frames in a sequential shift, illustrated in Figure 3.3, allows the model to consider surrounding rhythm variations for more accurate AF classification. Moreover, SeqAFNet's beat-wise output structure, as shown in Figure 3.5, not only contributes to the effective classification of each ECG beat, enhancing overall results, but also supports continuous output inference. This approach leads to an ensemble of these outputs, enabling more accurate decision-making by accumulating predictions from various time points. In the context of diagnosing AF using 12-lead or single-lead ECG recordings, the ESC guidelines propose that clinical AF should be identified when AF symptoms lasting for longer than 30 s [5]. However, applying these guidelines is challenging for many existing automated AF screening due to their specific output structures [21, 38, 39]. Many previous AF screening methods rely on their own criteria for AF screening, which are not always acceptable in clinical practice [21]. In contrast, SeqAFNet, with its beat-wise output structure, can effectively determine the duration of predicted AF exceeds 30 s, thereby aligning more closely with established clinical guidelines. Furthermore, the proficiency of our method in AF screening was evaluated based on its performance in determining AF duration, adhering to the IEC 60601-2-47 standards titled 'Particular Requirements for the Basic Safety and Essential Performance of

Ambulatory Electrocardiographic Systems.' The results, detailed in Table 3.6, highlight the effectiveness of our approach, especially when applied to adhesive patch-type wearable electrocardiographs.

Previous research has demonstrated the ability of their methods to accurately classify AF in a specific dataset. However, as shown in Table 3.13, these methods often exhibit decreased performance when applied to unseen data. In contrast, our proposed method, SeqAFNet, maintains consistent and reliable performance across different datasets, including AFDB, LTAFDB, and the patch dataset, as detailed in Table 3.13. Therefore, we believe that our approach, which has been effective in screening for AF using data measured by adhesive patch-type wearable electrocardiographs in this study, will also be effective in screening for AF in a variety of unseen data.

Table 3.13. Comparisons of the AF classification with recent deep learning models.

Training	Testing	Input	Output	Accuracy	Precision	Sensitivity	Specificity	F1	Method	Ref
AFDB	AFDB[a]	100 RRI	Single output	0.998	0.997	0.999	0.996	0.998	LSTM	[14]
CPSC2021	LTAFDB	100 RRI		0.909	0.905	0.912	0.912			[16]
AFDB	AFDB[b]	10sec RRI	Single output	0.996	0.994	0.996	0.995	0.995	Random Forest	[40]
AFDB	AFDB[c]			0.967	0.953	0.969	0.966	0.961		
MITDB	MITDB[b]			0.996	0.995	0.979	0.999	0.987		
MITDB	MITDB[c]			0.958	0.852	0.924	0.965	0.887		
LTAFDB	AFDB	10sec ECG	Single output	-	0.924	0.953	-	0.938	ResNet	[29]
AFDB	LTAFDB			-	0.833	0.806	-	0.819		
AFDB	AFDB	5sec ECG	Single output	0.993	0.993	0.993	-	0.993	Res-BiLSTM	[41]
AFDB	LTAFDB			0.913	0.919	0.912	-	0.915		
CPSC2021	CPSC2021	3beat ECG	Beat-wise	0.982	-	0.981	0.983	0.985	CNN with Transformer	[22]
CPSC2021	MITDB			0.963	-	0.806	0.98	0.854		
CPSC2021	AFDB			0.987	-	0.877	0.986	0.901		
CPSC2021	LTAFDB			0.862	-	0.657	0.831	0.74		
LTAFDB	AFDB	26.88sec ECG	Sample-wise	-	0.92	0.982	-	0.95	Swin Transformer	[42]
CPSC2021	LTAFDB			-	0.965	0.948	-	0.956		
AFDB	LTAFDB			-	0.947	0.949	-	0.948		
AFDB	AFDB[d]	RRI frames	Beat-wise	0.971	0.952	0.96	0.975	0.956	This study	
AFDB	LTAFDB			0.961	0.96	0.96	0.954	0.96		
AFDB	Patch			0.974	0.973	0.953	0.992	0.963		
LTAFDB	LTAFDB[d]			0.974	0.976	0.972	0.958	0.973		
LTAFDB	AFDB			0.975	0.974	0.975	0.97	0.975		
LTAFDB	Patch			0.986	0.981	0.979	0.992	0.98		

[a] This experiment was conducted on an inter-patient, with the testing dataset comprising records named 08405, 08434, and 08455.

[b] These experiments were conducted on intra-patient 10-fold cross-validation, meaning that data from the same patients could be used in both the training and testing sets.

[c] These experiments were conducted using 10-fold cross-validation, with each fold being divided on an inter-patient. There is no overlap of patients between the training data and the testing data.

[d] These experiments were conducted using 5-fold cross-validation, with each fold being divided on an inter-patient. There is no overlap of patients between the training data and the testing data.

Deep learning-based research for atrial fibrillation (AF) classification often overlooks practical challenges encountered in real-world clinical settings. Specifically, clinical technicians are required to review and correct misclassified results generated by automatic screening systems, making this a labor-intensive step in ensuring accurate diagnosis [43, 44]. In a recent study by Hao Wen et al., a beat-wise AF classification model based on LSTM was developed [16]. They utilized the recording '00' from the LTAFDB to assess its performance over time. The model achieved macro-averaged scores of 0.938 for sensitivity, 0.938 for specificity, 0.982 for precision, and 0.974 for accuracy, respectively. Initially, these results might indicate the model can classify AF effectively. However, an output displayed over time on a graph, as illustrated in Figure 3.6(a), reveals a significant number of misclassifications. Consequently, clinical technicians expend a significant amount of effort and time correcting such segmentalized misclassifications during ECG signal analysis. On the same recording '00', our proposed model, which was trained on the LTAFDB fold 1 trainset, achieved score of 0.988 for accuracy, 0.988 for precision, 0.989 for sensitivity, 0.981 for specificity, and 0.988 for F1 score. The output of our model was also plotted over time, as illustrated in Figure 3.6. Employing the proposed ensemble decision method enhances the robustness of the analysis, particularly in diminishing segmentalized misclassifications during beat-wise analysis, demonstrated in Figure 3.6(b). In Figure 3.6(b), the labels indicating non-AF correspond to instances of ventricular tachycardia (VT), characterized by the occurrence of three or four premature ventricular complexes in succession during episodes of AF. Figure 3.6(c), presenting a magnified view of the section from 40,000 to 50,000 beats, demonstrates our model's capability to accurately classify transitions into and termination of AF. Consequently, our proposed method substantially lightens the workload of clinical technicians by minimizing the necessity for reviewing and correcting misclassifications. As a result of its efficiency, it holds significant potential as a valuable tool in clinical practice.

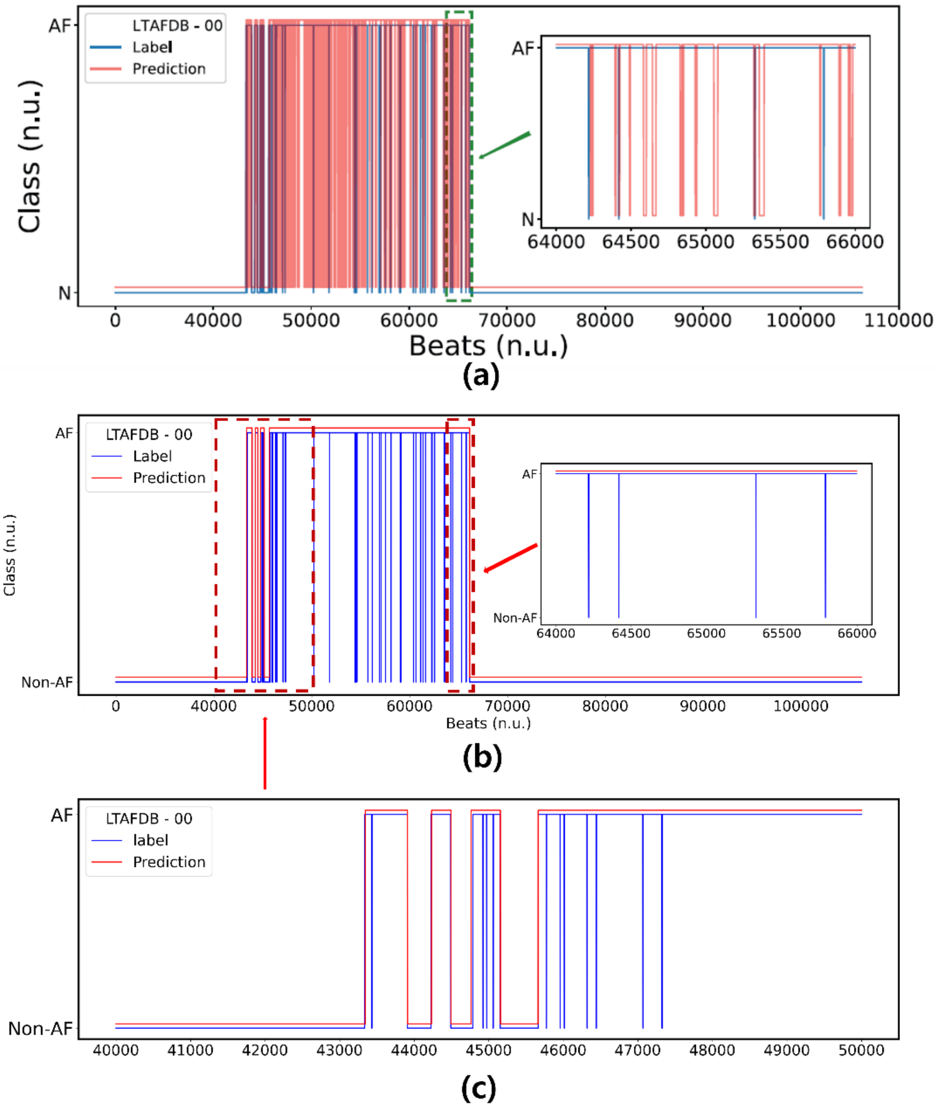

Figure 3.6. Comparison of beat-wise evaluation results on recording 00 from LTAFDB: Study by Hao Wen et al. and our Study. (a) Study by Hao Wen et al. [16] (b) Results of our proposed method. (c) Magnified view of section from 40,000 to 50,000 beats.

Despite its significant results, the current study is not without limitations. The primary limitation lies in the model's reliance on RRIs of ECG waveforms for input. While this approach enables our model to effectively capture irregular cardiac rhythms, which is a characteristic of AF, it overlooks the detailed morphological features associated with AF. These include the lack of distinct, repeating P waves or the presence of F waves, which are crucial for a comprehensive AF analysis. Incorporating morphological features along with the irregular intervals currently utilized in AF classification could potentially elevate the model's accuracy in diagnosing AF. As a secondary limitation, our proposed methods are vulnerable to VT events during AF episodes. In the LTAFDB, the occurrence of three or four consecutive premature ventricular complexes, indicative of VT, can coincide with AF episodes. These ventricular arrhythmias may obscure the exact start and end points of AF episodes, thereby complicating the classification process. Considering the morphological features of the ECG during such episodes could provide crucial context, potentially enhancing the accuracy of classification. For future work, it is essential to investigate methodologies that combine RR intervals and detailed ECG waveform features as sequential inputs within a frame-based model. This advancement could notably enhance the performance and accuracy of AF classification systems.

3.5 Summary

This study presented the use of SeqAFNet, a many-to-many RNN-based model designed for the beat-wise sequential classification of AF. This model is suitable for wearable electrocardiographs. By employing an ensemble decision strategy, our approach refines the model's many-to-many output, thereby effectively correcting mispredicted beats. The performance of SeqAFNet was comprehensively tested by training it on two separate databases (AFDB and LTAFDB). Further, its capabilities were evaluated using data from MEMO Patch, a type of adhesive patch-type wearable electrocardiograph. SeqAFNet had a consistently high efficacy, indicating a robust performance with not only the MEMO Patch dataset but also the tested datasets from public databases. Furthermore, our model was evaluated based on the Subclause 201.12.1.101.1.5.3 of the IEC 60601-2-47 standard, which involves the basic safety and essential performance of ambulatory electrocardiographic systems, with a specific focus on AF duration assessment. In conclusion, our proposed method improves the screening of clinical AF using adhesive patch-type wearable electrocardiographs that can be used for up to 14 days, which is in accordance with the 2020 ESC Guidelines for AF diagnosis in clinical practice.

4. SE-ResNet-ViT Hybrid Model for Noise Classification in Adhesive Patch-Type Wearable Electrocardiographs

4.1 Motivation

Arrhythmia is a heart rhythm abnormality where the heart beats too slowly, too rapidly, or irregularly [1, 2]. Certain arrhythmias pose risks of serious complications, including stroke, heart failure, and cardiac arrest. Detecting these arrhythmias, which can occur intermittently, is challenging when patients undergo electrocardiogram (ECG) measurements during hospital visits [45]. Hence, to diagnose arrhythmia, ECG signals are commonly recorded using a Holter monitor over a period of 24 or 48 hours. However, previous studies indicate that 24- or 48-hour Holter monitoring is often ineffective in diagnosing certain clinically significant asymptomatic arrhythmias, including episodes of atrial fibrillation and transient bradyarrhythmia [46, 47]. Furthermore, these Holter monitors, also known as memory recorders, can be cumbersome for patients due to their bulky size or complex design.

Recently, to minimize patient inconvenience, patch-type single-lead electrocardiographs have been launched in market, such as the Zio Patch (iRhythm in the United States), Ezypro (SIGKNOW in Taiwan), and MEMO Patch (HUINNO in Korea). These patch-type electrocardiographs improve patient convenience in daily life, thanks to their lightweight and compact design. The patch-type electrocardiograph, designed for low-power consumption, can record ECG signals for up to 14 days. Previous studies have shown that recording ECG for approximately 14 days is effective in detecting most symptomatic arrhythmias [7]. Therefore, using this long-term patch-type electrocardiograph can enable more accurate diagnosis of arrhythmia and prevention of serious complications like stroke, heart failure, and cardiac arrest, compared to the 24- or 48-hour monitoring using a Holter monitor. Analyzing signals from patch-type electrocardiographs is more labor-intensive compared to those from Holter monitors, as the longer recording time leads to an increase in the absolute amount of noise within the signal. Furthermore, the fact that the morphologies of some noise signals are similar to some of arrhythmia signals complicates the task for machine learning models or algorithms in distinguishing between noise and arrhythmia signals. Therefore, patch-type long-term electrocardiographs require more robust software support for automated ECG analysis and arrhythmia classification compared to Holter monitors. Previous studies have proposed deep learning models to classify noise and ECG signals. However, these methods used only ECG data from ICU or Holter monitors for training and evaluating their models [48, 49]. Therefore, the efficacy of these models in classifying ECG signals from wearable electrocardiographs has yet to be evaluated. Additionally, some studies face the issue of not including arrhythmia signals in their data, consequently making these models less effective for arrhythmia diagnosis [50].

In this study, we introduce a SE-ResNet-ViT hybrid model designed to classify noise signals from arrhythmic ECG signals in patch-type wearable electrocardiographs. ECG signals were collected over 14 days using HUINNO's MEMO Patch, which included both arrhythmia and noise signals, from patients with a history of arrhythmia or symptoms indicative of arrhythmia. The proposed model was trained, and its performance evaluated in classifying signals as either noise or non-noise.

4.2 Methods

4.2.1 Data collection

Data were gathered from a multi-center clinical trial conducted at Korea University Hospital and Seoul National University Bundang Hospital, receiving approval from the Institutional Review Board of each institution. The IRB numbers for the clinical trial are 2021AN0247(Korea University Hospital) and B-2105/686-002(Seoul National University Bundang Hospital) respectively. Patients in need of ambulatory ECG monitoring were considered eligible if they had been diagnosed with stroke or transient ischemic attack with no identifiable causes, or if they exhibited symptoms such as palpitation, dizziness, or syncope. Patients were invited to participate in the study if they were aged between 19 and 80 years old, capable of providing voluntary informed consents, and able to adhere to the study protocol for 14 days of attaching a MEMO Patch for monitoring.

Figure 4.1 depicts the overall process of data collection. In the clinical trial, a total of 149 people participated. For this study, data from 70 individuals were randomly selected and analyzed. The labeling process for ECG signals involved several steps. Initially, non-clinical experts manually reviewed and selected 117,000 noisy ECG signals. Following this, clinical technicians reviewed and labeled 2,084 noise signals, 7,552 normal sinus rhythm (NSR) signals, and 8,086 arrhythmia signals, which were then further inspected by a cardiologist. Noise signals not only consist of pure noise but also include signals that are a mixture of noise and ECG. The arrhythmia signals in this study included various types, such as atrial premature contractions (APC), ventricular premature contractions (VPC), atrial fibrillation (AF), supraventricular tachycardia (SVT), atrioventricular block (AVB), and other arrhythmias. Table 4.1 shows the detailed classes and quantities of arrhythmia. To increase the quantity of noise signals, we collected data from 21 healthy individuals using the MEMO Patch during their daily life. The training and testing datasets were split into a 7:3 ratio across the noise, NSR, and arrhythmia ECG signal categories, with no patient overlap between the sets to prevent overfitting. The training dataset was divided into an 8:2 ratio for the purposes of training and validating the model.

Table 4.1. Quantities of the collected and labeled data.

Arrhythmia	Training	Testing
Normal sinus rhythm	5242	2310
Noise signal	5113	2354
Atrial premature contractions	3476	1189
Ventricular premature contractions	1018	902
Atrial fibrillation	934	87
Supraventricular tachycardia	265	84
Atrioventricular block	3	102
Other arrhythmias	16	10
Total	16067	7038

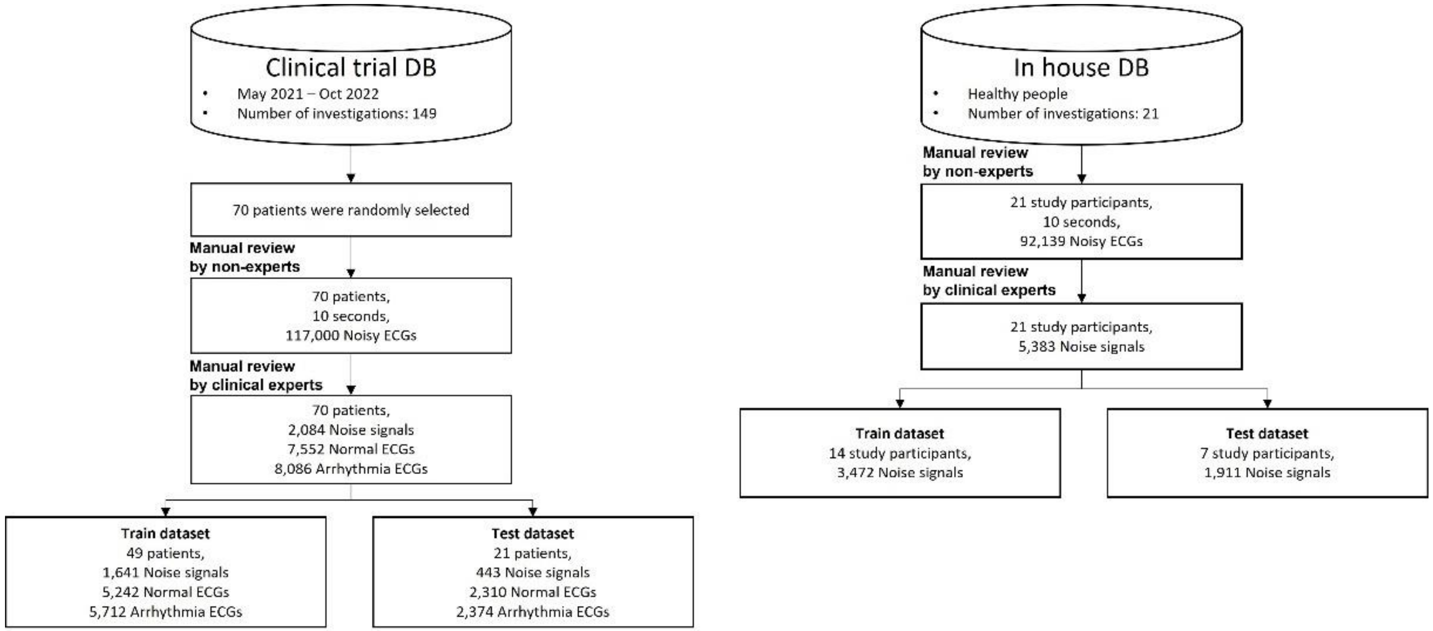

Figure 4.1. Flowchart of database collection process.

4.2.2 Device and software

Figure 2.1 displays the MEMO Patch, a single-lead adhesive patch-type ambulatory electrocardiograph, used to record ECGs from patients participating in the clinical trial. This device is approved by the Ministry of Food and Drug Safety (MFDS) in the Republic of Korea. This device is capable of operating for up to 14 days, recording ECG at a sampling rate of 250 Hz and with a 12-bit resolution. Patients visited the hospital to attach the device to their bodies and then went about their daily lives, measuring their ECG signals. After 14 days, they returned to the hospital to hand back the device. Upon its return, a technician downloaded the ECG data stored in the device's memory. The ECG data are initially pre-annotated using a machine learning model for arrhythmia classification, known as MEMO Care, provided by HUINNO, the manufacturer of the device. Subsequently, all the data referenced in this paper is reviewed by clinical technicians.

4.2.3 Preprocessing

Some noise signals within the ECG can be eliminated using a simple digital filter. For instance, baseline wander, a low-frequency noise, arises due to factors like breathing, movement, or electrically charged electrodes [51]. This type of noise can be removed by applying a high-pass filter with a cut-off frequency below 1 Hz. Additionally, the ECG signal can be contaminated by high-frequency EMG signals during patient movement, which can be filtered out using a low-pass filter [52]. Increasing the order of the filters and narrowing the cut-off frequency can effectively eliminate these noises from the signals. However, this approach may distort the ECG signal, potentially leading to reduced performance in arrhythmia classification. We applied a second-order band-pass Butterworth filter with a 0.5-50Hz range in this study to remove baseline drift and high-frequency noise from each 10-second ECG signal. Subsequently, the signals were normalized from 0 to 1 using min-max scaling.

4.2.4 Architecture of SE-ResNet-Vit model

To classify noise signals within ECG signals, we propose a SE-ResNet-ViT hybrid model-based architecture. The overall structure of the model is depicted in Figure 4.2. Since the introduction of the hybrid model that combines Convolutional Neural Network (CNN) with Vision Transformer (ViT) in Dosovitskiy's 2020 paper, numerous studies have adopted this hybrid architecture [53]. The hybrid model employs a technique that applies feature maps extracted from a CNN to the patch embedding projection for enhanced image analysis. The hybrid model demonstrates superior performance in image classification compared to the ResNet model, and it also outperforms ViT in smaller-sized models.

One of the ResNet models, SE-ResNet, is recognized for its high classification performance among CNN-structured models, and our previous study confirmed it surpasses the classification performance of the standard ResNet model [54, 55]. Given the high performance of SE-ResNet, we aimed to create a hybrid by combining SE-ResNet and ViT. This hybrid model projects the output of the CNN feature map, using a 1x1 patch size, into the Transformer dimension. The difference between our previous study and the current one is that the sampling rate has been changed from 200 Hz to 250 Hz.

The input shape for the model is set to batch size × 2500 × 1. The stem layer of the model comprises a convolution layer with a kernel size of 7 and a stride of 2, followed by a max pooling layer with a window size of 3, a stride of 2, and padding of 1. The layer composition, as represented by the stride block, aligns with that of SE-ResNet. However, within the ResNet block, the stride is altered to 2 for the convolution process.

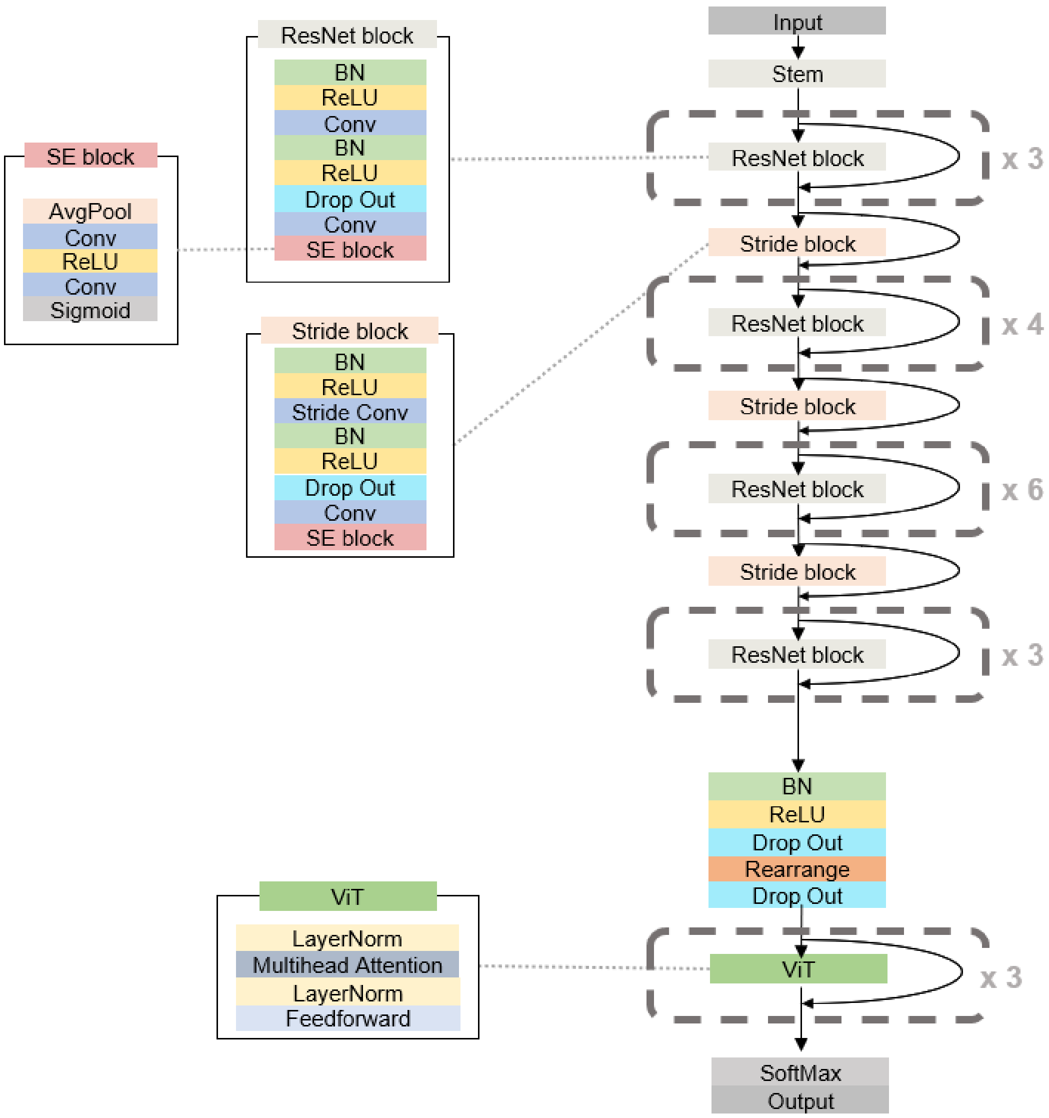

Figure 4.2. Architecture of the SE-ResNet-ViT hybrid model.

4.3 Results

4.3.1 Experiment setups

For model optimization, we employed the Adam optimizer with an initial learning rate of 0.0005. Cross entropy was used as the loss function, and the training was conducted for up to 40 epochs with a batch size of 512. All experiments were conducted using PyTorch on an RTX2080TI GPU.

4.3.2 Performance of noise classification

Model performance was assessed using a test dataset, which was categorized into noise and non-noise classes. The evaluation metrics, including precision, recall, and F1 score, are detailed in Table 4.2. The average score for both classes was computed as a weighted average to account for the differing quantities in each class.

Table 4.2. Score of noise classification.

	Precision	recall	F1 score
Noise	0.932	0.962	0.947
Non-noise	0.980	0.965	0.973
Weighted Avg	0.964	0.964	0.964

The confusion matrix is presented in Figure 4.3. The weighted averages of the F1 score, precision, and recall were all calculated to be 0.964. However, the precision for noise signals was slightly lower at 0.932. This discrepancy is likely due to the smaller number of noise signals in the test dataset compared to non-noise signals, or a larger number of non-noise predictions. The supposed reason for the lower precision in noise signals is the fewer number of noise signals in the test dataset compared to non-noise signals, or a higher number of predictions classified as non-noise. The specific classification details of ECG signals identified as non-noise are presented in Table 4.3. Upon reviewing the misclassified classes, we noted that VPC signals were most frequently misclassified as noise. A comparison of the misclassified noise and VPC signals revealed similar shapes, especially after min-max scaling, as illustrated in Figure 4.4. Both waveforms displayed wide QRS complexes and abnormal shapes.

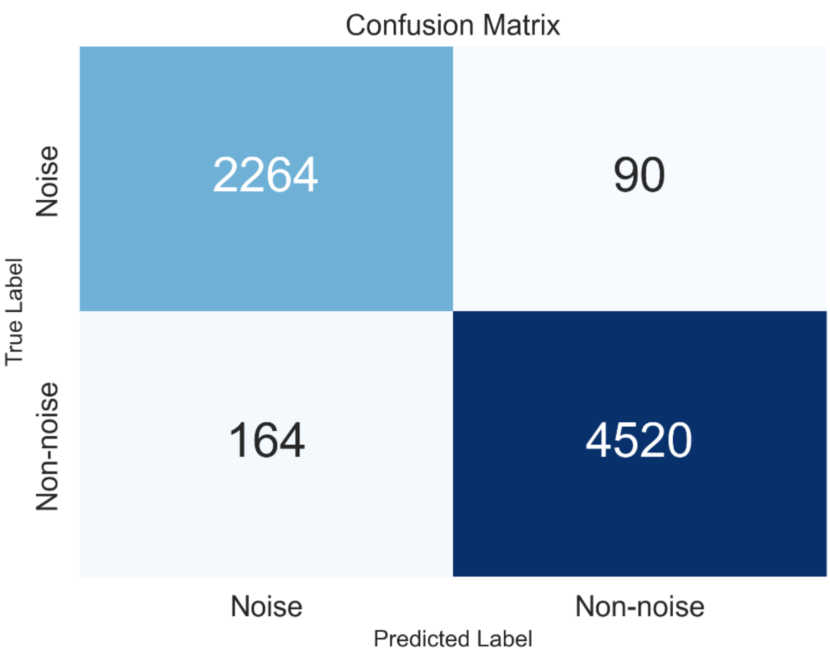

Figure 4.3. Confusion matrix of noise classification.

Table 4.3. Number of count non-noise was incorrectly classified as noise.

Arrhythmia	Counts
Normal sinus rhythm	44
Atrial premature contractions	15
Ventricular premature contractions	101
Atrial fibrillation	1
Supraventricular tachycardia	3
Atrioventricular block	0
Other arrhythmias	0
Total	164

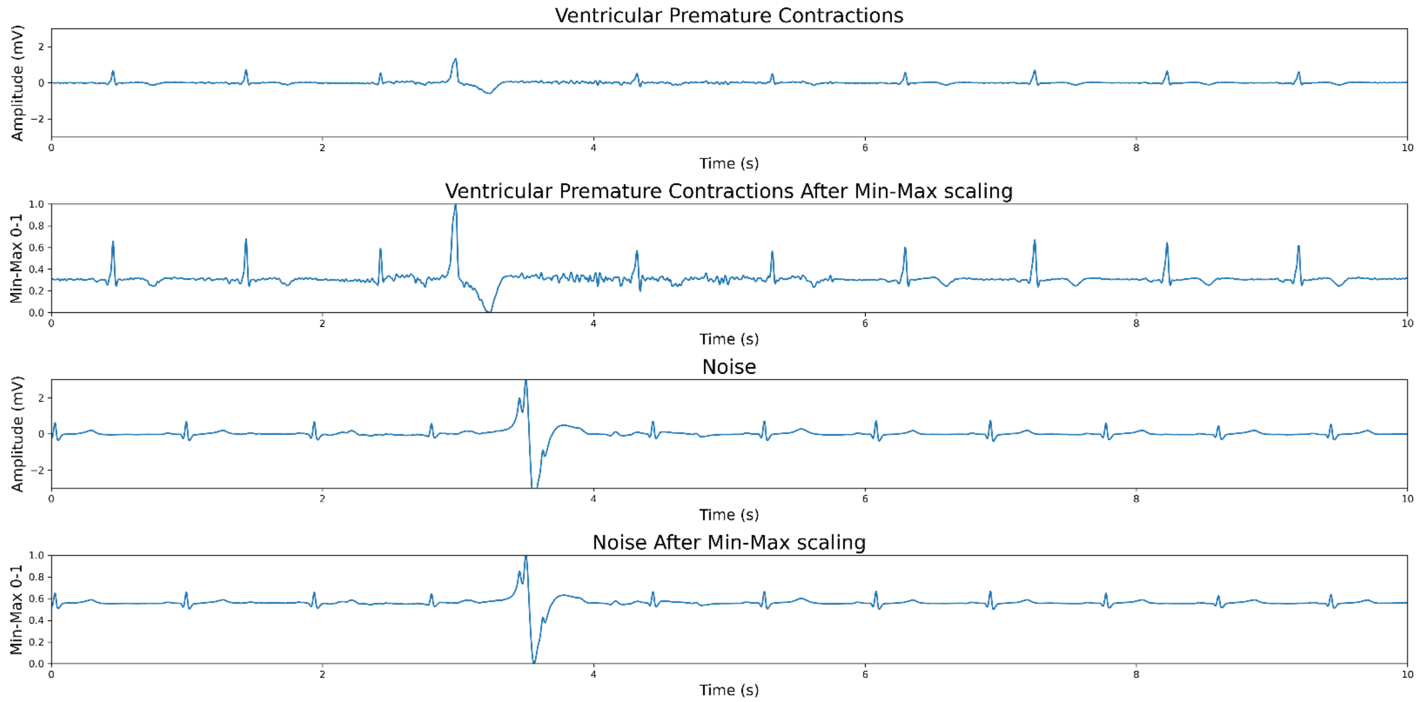

Figure 4.4. Comparison of the waveform VPC and noise after minmax scaling, they become similar to each other.

4.3.3 Comparison with other noise classification studies

We compare the performance of noise classification in ECG with previous studies [56-59]. Table 4.4 shows comparison performance with other noise classification studies. Due to the absence of ECG signals containing arrhythmias in public data, many previous studies trained models on noise classification using NSR and then evaluated these models for their tendency to misclassify arrhythmias as noise. The 2017 PhysioNet database used in the study by Smisek et al, includes both arrhythmic signals and noise signals [58]. However, the noise signals in this database consist of only 46 signals with an average length of 27.2 seconds, which may contribute to the poor classification performance of noise. The data used in previous studies consist of ECGs measured in ICUs or 12-lead resting ECGs. The 2017 PhysioNet data was collected using AliveCor's single-lead electrocardiograph, but this ECG is also a resting ECG. Therefore, noise classification models trained on resting ECGs could face greater challenges in classifying noise arising from various activities during daily life.

Table 4.4. Comparison performance with other noise classification studies.

Method	Data	Accuracy	Precision	Recall	F1 score	Ref
CNN	ICU (without Arrythmia)	0.89	0.957	0.893	0.924	[56]
	ICU (Arrythmia only)	0.666	-	-	-	
Spectrum analysis	ICU (without Arrythmia)	0.909	-	-	-	[57]
	ICU (AF only)	0.862	-	-	-	
SVM	2017 Physionet (Overall)	-	-	-	0.81	[58]
	2017 Physionet (AF)				0.81	
	2017 Physionet (Others)				0.72	
	2017 Physionet (Noise)	-	-	-	0.55	
ResNet50	2020 Physionet (Overall)	-	-	-	0.77	[59]
	2020 Physionet (NSR)	-	-	-	0.74	
	2020 Physionet (AF)	-	-	-	0.92	
ResNet-Vit hybrid (This study)	Patch (Overall)	0.974	0.964	0.964	0.964	
	Patch (NSR)	0.983	-	-	-	
	Patch (APC)	0.987	-	-	-	
	Patch (VPC)	0.888	-	-	-	
	Patch (AF)	0.989	-	-	-	
	Patch (SVT)	0.964	-	-	-	
	Patch (AV block)	1				
	Patch (Others)	1	-	-	-	

4.4 Discussion

Analyzing signals from long-term wearable electrocardiographs can be more labor-intensive, as the extended recording time results in a higher amount of noise being included in the signal. To minimize the time-consuming aspect of ECG signal analysis, this study introduces a SE-ResNet-ViT hybrid model considered for classifying noise signals in long-term ECG data from wearable patch-type devices. ECG data for training and evaluating the model were collected from participants in clinical trials, who were either diagnosed with or suspected of having arrhythmia. The collected data were reviewed and labeled by clinical experts. Upon evaluating the trained model, it achieved a weighted average F1 score of 0.964, demonstrating its effectiveness in accurately classifying noise signals measured by patch-type wearable ECG devices. Nevertheless, it is observed that some VPC signals, resembling noise signals in shape, are occasionally misclassified as noise. In the future, we aim to minimize misclassification caused by the shape similarities between VPC signals and certain noise signals. We expect that the proposed noise classification method will help in screening arrhythmias more accurately.

5. Conclusion

This study has presented reliable methods for classifying arrhythmias using deep learning techniques to apply on adhesive patch-type wearable electrocardiographs.

In Chapter 3, we have designed SeqAFNet, a many-to-many RNN-based model, specifically for the beat-wise sequential classification of AF. This model sequentially utilized RRIs from ECG signal. Thanks to its robustness, SeqAFNet demonstrated effective classification of AF in ECG data, encompassing both records from public databases obtained via ambulatory ECG recorders and those derived from wearable electrocardiographs. Additionally, SeqAFNet aligns with the 2020 ESC Guidelines for AF diagnosis, attributable to its beat-wise input and output structure.

In Chapter 4, the SE-ResNet-ViT Hybrid model is introduced, designed to classify noise signals occurring during the attachment of wearable electrocardiographs in daily life of patients. Numerous deep learning and algorithm-based arrhythmia classification methods are highly vulnerable to noise signals. To accurately classify arrhythmia in ECG signals, an effective noise classification method is crucial. The SE-ResNet structure was able to effectively extract morphological features from the ECG signals, and the ViT structure enabled attention to the areas where noise occurred in the ECG signals. The SE-ResNet-ViT Hybrid model, combining these advantages, was able to effectively classify noise signals in ECG signals measured from wearable electrocardiographs.

Despite the significant results, the current study had several limitations. First, SeqAFNet depends on a method that identifies R peak locations in the ECG waveform. Should the R peak detection methods malfunction, SeqAFNet's performance could deteriorate. To prevent such issues, high-performance R peak detection methods are required. Additionally, it is anticipated that the noise detection model introduced in Chapter 4 could be utilized to prevent malfunctions in noisy ECG waveforms.

We believe that the proposed methods will be more effective with adhesive patch-type wearable electrocardiographs, which offer a longer measurement duration compared to traditional Holter monitors, particularly for patients suspected of arrhythmias. The extended measurement duration of wearable electrocardiographs and the structural limitations that make ECG signals more challenging and voluminous to analyze. Our proposed methods would provide valuable assistance to medical professionals in clinical practice.

References

[1] J. Norris, G. Froggatt, and V. Hachinski, "Cardiac arrhythmias in acute stroke," *Stroke*, vol. 9, no. 4, pp. 392-396, Jul. 1978, doi: 10.1161/01.STR.9.4.392.

[2] W.H. Maisel and L.W. Stevenson, "Atrial fibrillation in heart failure: epidemiology, pathophysiology, and rationale for therapy," *The American journal of cardiology*, vol. 91, no. 6, pp. 2-8, Mar. 2003, doi: 10.1016/S0002-9149(02)03373-8.

[3] D. Vella, A. Monteleone, G. Musotto, G.M. Bosi, and G. Burriesci, "Effect of the alterations in contractility and morphology produced by atrial fibrillation on the thrombosis potential of the left atrial appendage," *Frontiers in Bioengineering and Biotechnology*, vol. 9, p. 586041, Feb. 2021, doi: 10.3389/fbioe.2021.586041.

[4] B. Freedman *et al.*, "Screening for atrial fibrillation: a report of the AF-SCREEN International Collaboration," *Circulation*, vol. 135, no. 19, pp. 1851-1867, May. 2017, doi: 10.1161/CIRCULATIONAHA.116.026693.

[5] G. Hindricks *et al.*, "2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS) The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC," *European heart journal*, vol. 42, no. 5, pp. 373-498, Feb. 2021, doi: 10.1093/eurheartj/ehaa612.

[6] R. Stahrenberg *et al.*, "Enhanced detection of paroxysmal atrial fibrillation by early and prolonged continuous holter monitoring in patients with cerebral ischemia presenting in sinus rhythm," *Stroke*, vol. 41, no. 12, pp. 2884-2888, Oct. 2010, doi: 10.1161/STROKEAHA.110.591958.

[7] P.M. Barrett *et al.*, "Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring," *The American journal of medicine*, vol. 127, no. 1, pp. 95. e11-95. e17, Jan. 2014, doi: 10.1016/j.amjmed.2013.10.003.

[8] F. Sana, E.M. Isselbacher, J.P. Singh, E.K. Heist, B. Pathik, and A.A. Armoundas, "Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review," *Journal of the American College of Cardiology*, vol. 75, no. 13, pp. 1582-1592, Apr. 2020, doi: 10.1016/j.jacc.2020.01.046.

[9] Y.-C. Wang *et al.*, "Current advancement in diagnosing atrial fibrillation by utilizing wearable devices and artificial intelligence: A review study," *Diagnostics*, vol. 12, no. 3, p. 689, Mar. 2022, doi: 10.3390/diagnostics12030689.

[10] A. Khunte *et al.*, "Detection of left ventricular systolic dysfunction from single-lead electrocardiography adapted for portable and wearable devices," *npj Digital Medicine*, vol. 6, no. 1, p. 124, Jul. 2023, doi: 10.1038/s41746-023-00869-w.

[11] A.Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, pp. 65-69, Jan. 2019, doi: 10.1038/s41591-018-0268-3.
[12] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, and A.Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, Jul. 2017, doi: 10.48550/arXiv.1707.01836.
[13] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," *Expert Systems with Applications: X*, vol. 7, p. 100033, Sep. 2020, doi: 10.1016/j.eswax.2020.100033.
[14] O. Faust, A. Shenfield, M. Kareem, T.R. San, H. Fujita, and U.R. Acharya, "Automated detection of atrial fibrillation using long short-term memory network with RR interval signals," *Computers in biology and medicine,* vol. 102, pp. 327-335, Nov. 2018, doi: 10.1016/j.compbiomed.2018.07.001.
[15] P. Zhang, Y. Chen, F. Lin, S. Wu, X. Yang, and Q. Li, "Semi-supervised learning for automatic atrial fibrillation detection in 24-hour Holter monitoring," *IEEE Journal of Biomedical and Health Informatics,* vol. 26, no. 8, pp. 3791-3801, May. 2022, doi: 10.1109/JBHI.2022.3173655.
[16] H. Wen and J. Kang, "A comparative study on neural networks for paroxysmal atrial fibrillation events detection from electrocardiography," *Journal of Electrocardiology,* vol. 75, pp. 19-27, Nov. 2022, doi: 10.1016/j.jelectrocard.2022.10.002.
[17] M.R. Reynolds *et al.*, "Comparative effectiveness and healthcare utilization for ambulatory cardiac monitoring strategies in Medicare beneficiaries," (in eng), *Am Heart J,* vol. 269, pp. 25-34, Mar 2024, doi: 10.1016/j.ahj.2023.12.002.
[18] S. Nattel, "New ideas about atrial fibrillation 50 years on," *Nature,* vol. 415, no. 6868, pp. 219-226, Jan. 2002, doi: 10.1038/415219a.
[19] E.J. Benjamin, P.A. Wolf, R.B. D'Agostino, H. Silbershatz, W.B. Kannel, and D. Levy, "Impact of atrial fibrillation on the risk of death: the Framingham Heart Study," *Circulation,* vol. 98, no. 10, pp. 946-952, Sep. 1998, doi: 10.1161/01.CIR.98.10.946.
[20] G.B. Forleo *et al.*, "Exploring the potential role of catheter ablation in patients with asymptomatic atrial fibrillation: should we move away from symptom relief?," *Journal of Atrial Fibrillation,* vol. 6, no. 3, Oct. 2013, doi: 10.4022/jafib.961.
[21] P. Zhang *et al.*, "Automatic screening of patients with atrial fibrillation from 24-h Holter recording using deep learning," *European Heart Journal-Digital Health,* vol. 4, no. 3, pp. 216-224, May. 2023, doi: 10.1093/ehjdh/ztad018.
[22] Y. Hu, T. Feng, M. Wang, C. Liu, and H. Tang, "Detection of Paroxysmal Atrial Fibrillation from Dynamic ECG Recordings Based on a Deep Learning Model," *Journal of Personalized Medicine,* vol. 13, no. 5, p. 820, May. 2023, doi: 10.3390/jpm13050820.

[23] Y. Liu *et al.*, "Generalizable beat-by-beat arrhythmia detection by using weakly supervised deep learning," *Frontiers in Physiology*, vol. 13, p. 850951, Mar. 2022, doi: 10.3389/fphys.2022.850951.

[24] T. Stracina, M. Ronzhina, R. Redina, and M. Novakova, "Golden standard or obsolete method? Review of ECG applications in clinical and experimental context," *Frontiers in Physiology*, vol. 13, p. 867033, Apr. 2022, doi: 10.3389/fphys.2022.867033.

[25] X. Wang, C. Ma, X. Zhang, H. Gao, G.D. Clifford, and C. Liu, "Paroxysmal atrial fibrillation events detection from dynamic ECG recordings: The 4th China physiological signal challenge 2021," *Physionet*, pp. 1-83, Jun. 2021, doi: 10.13026/h37m-r024.

[26] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Computers in biology and medicine*, vol. 93, pp. 84-92, Feb. 2018, doi: 10.1016/j.compbiomed.2017.12.007.

[27] X. Zhang, J. Li, Z. Cai, L. Zhang, Z. Chen, and C. Liu, "Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection," *Medical & Biological Engineering & Computing*, vol. 59, pp. 165-173, Jan. 2021, doi: 10.1007/s11517-020-02292-9.

[28] R.S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Systems with Applications*, vol. 115, pp. 465-473, Jan. 2019, doi: 10.1016/j.eswa.2018.08.011.

[29] H.-C. Seo, S. Oh, H. Kim, and S. Joo, "ECG data dependency for atrial fibrillation detection based on residual networks," *Scientific Reports*, vol. 11, no. 1, p. 18256, Sep. 2021, doi: 10.1038/s41598-021-97308-1.

[30] G. Moody, "A new method for detecting atrial fibrillation using RR intervals," *Proc. Comput. Cardiol.*, vol. 10, pp. 227-230, 1983.

[31] 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. 466-470, Jul. 2007, doi: 10.1093/europace/eum096.

[32] S. Kim, J. Lim, M. Shin, and S. Jung, "SE-ResNet-ViT Hybrid Model for Noise Classification in Adhesive Patch-type Wearable Electrocardiographs," *Int. Conf. of the IEEE EMBC*, pp. 1-4, Jul. 2023, doi: 10.1109/EMBC40787.2023.10340882.

[33] M. Schuster and K.K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673-2681, Nov. 1997, doi: 10.1109/78.650093.

[34] J. Zhang and K.-F. Man, "Time series prediction using RNN in multi-dimension embedding phase space," in *IEEE International Conference on Systems*, Oct. 1998, vol. 2: IEEE, pp. 1868-1873, doi: 10.1109/ICSMC.1998.728168.

[35] P. Le and W. Zuidema, "Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive LSTMs," *arXiv*, 2016, doi: arXiv.1603.00423.

[36] A. Shewalkar, D. Nyavanandi, and S.A. Ludwig, "Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU," *Journal of Artificial Intelligence and Soft Computing Research,* vol. 9, no. 4, pp. 235-245, Aug. 2019, doi: 10.2478/jaiscr-2019-0006.
[37] I.E. Commission, "Medical electrical equipment-part 2-47: Particular requirements for the basic safety and essential performance of ambulatory electrocardiographic systems, IEC 60601-2-47," ed, 2012.
[38] J. Rahul and L.D. Sharma, "Artificial intelligence-based approach for atrial fibrillation detection using normalised and short-duration time-frequency ECG," *Biomedical Signal Processing and Control,* vol. 71, p. 103270, Jan. 2022, doi: 10.1016/j.bspc.2021.103270.
[39] S. Nurmaini *et al.*, "Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks," *Future Generation Computer Systems,* vol. 113, pp. 304-317, Dec. 2020, doi: 10.1016/j.future.2020.07.021.
[40] J.L. Peng Xiong, Ming Liu, Jieshuo Zhang, Jianli Yang, Hong Zhang, Haiman Du, Xiuling Liu, "Short-term paroxysmal atrial fibrillation detection with intra- and inter-patient paradigm based on R-R intervals," *Biomedical Signal Processing and Control,* vol. 89, no. 105750, 2024, doi: 10.1016/j.bspc.2023.105750.
[41] Z. Huang, S. Yang, Q. Zou, X. Gao, and B. Chen, "A portable household detection system based on the combination of bidirectional LSTM and residual block for automatical arrhythmia detection," *Biomedical Engineering/Biomedizinische Technik,* vol. 69, no. 2, pp. 167-179, 2024, doi: 10.1515/bmt-2021-0146.
[42] D. Yun *et al.*, "Automatic segmentation of atrial fibrillation and flutter in single-lead electrocardiograms by self-supervised learning and Transformer architecture," *Journal of the American Medical Informatics Association,* vol. 31, no. 1, pp. 79-88, 2024, doi: 10.1093/jamia/ocad219.
[43] Z.I. Attia *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," *The Lancet,* vol. 394, no. 10201, pp. 861-867, Sep. 2019, doi: 10.1016/S0140-6736(19)31721-0.
[44] C. Li *et al.*, "DeepECG: Image-based electrocardiogram interpretation with deep convolutional neural networks," *Biomedical Signal Processing and Control,* vol. 69, p. 102824, Aug. 2021, doi: 10.1016/j.bspc.2021.102824.
[45] P. Zimetbaum and A. Goldman, "Ambulatory arrhythmia monitoring: choosing the right device," *Circulation,* vol. 122, no. 16, pp. 1629-1636, Oct. 2010, doi: 10.1161/CIRCULATIONAHA.109.925610.
[46] J.A. Reiffel, R. Schwarzberg, and M. Murry, "Comparison of autotriggered memory loop recorders versus standard loop recorders versus 24-hour Holter monitors for arrhythmia detection," *The American journal of cardiology,* vol. 95, no. 9, pp. 1055-1059, May. 2005, doi: 10.1016/j.amjcard.2005.01.025.

[47] J. Liao, Z. Khalid, C. Scallan, C. Morillo, and M. O'Donnell, "Noninvasive cardiac monitoring for detecting paroxysmal atrial fibrillation or flutter after acute ischemic stroke: a systematic review," *Stroke*, vol. 38, no. 11, pp. 2935-2940, Sep. 2007, doi: 10.1161/STROKEAHA.106.478685.

[48] J. Pestana, D. Belo, and H. Gamboa, "Detection of Abnormalities in Electrocardiogram (ECG) using Deep Learning," in *BIOSIGNALS*, 2020, pp. 236-243, doi: 10.5220/0008967302360243.

[49] G.B. Moody and R.G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45-50, May. 2001, doi: 10.1109/51.932724.

[50] J.N. John, C. Galloway, and A. Valys, "Deep convolutional neural networks for noise detection in ecgs," *arXiv preprint arXiv:1810.04122*, Oct. 2018, doi: 10.48550/arXiv.1810.04122.

[51] G. Lenis, N. Pilia, A. Loewe, W.H. Schulze, and O. Dössel, "Comparison of baseline wander removal techniques considering the preservation of ST changes in the ischemic ECG: a simulation study," *Computational and mathematical methods in medicine*, vol. 2017, Mar. 2017, doi: 10.1155/2017/9295029.

[52] I. Christov, T. Neycheva, R. Schmid, T. Stoyanov, and R. Abächerli, "Pseudo-real-time low-pass filter in ECG, self-adjustable to the frequency spectra of the waves," *Medical & biological engineering & computing*, vol. 55, pp. 1579-1588, Feb. 2017, doi: 10.1007/s11517-017-1625-y.

[53] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, Jun. 2020, doi: 10.48550/arXiv.2010.11929.

[54] K.-S. Lee *et al.*, "Compressed deep learning to classify arrhythmia in an embedded wearable device," *Sensors*, vol. 22, no. 5, p. 1776, Feb. 2022, doi: 10.3390/s22051776.

[55] J. Park, J.-k. Kim, S. Jung, Y. Gil, J.-I. Choi, and H.S. Son, "ECG-signal multi-classification model based on squeeze-and-excitation residual neural networks," *Applied Sciences*, vol. 10, no. 18, p. 6495, Sep. 2020, doi: 10.3390/app10186495.

[56] D. Yoon, H.S. Lim, K. Jung, T.Y. Kim, and S. Lee, "Deep learning-based electrocardiogram signal noise detection and screening model," *Healthcare informatics research*, vol. 25, no. 3, p. 201, 2019, doi: 10.4258/hir.2019.25.3.201.

[57] S.K. Bashar, E. Ding, A.J. Walkey, D.D. McManus, and K.H. Chon, "Noise detection in electrocardiogram signals for intensive care unit patients," *IEEE Access*, vol. 7, pp. 88357-88368, 2019, doi: 10.1109/ACCESS.2019.2926199.

[58] R. Smíšek *et al.*, "SVM based ECG classification using rhythm and morphology features, cluster analysis and multilevel noise estimation," in *2017 Computing in Cardiology (CinC)*, 2017: IEEE, pp. 1-4, doi: 10.22489/CinC.2017.172-200.

[59] J. Venton, P.M. Harris, A. Sundar, N.A. Smith, and P.J. Aston, "Robustness of convolutional neural networks to physiological electrocardiogram noise,"

Philosophical Transactions of the Royal Society A, vol. 379, no. 2212, p. 20200262, 2021, doi: 10.1098/rsta.2020.0262.

[60] S. Kim, J. Lim, and J. Jang, "SeqAFNet: A Beat-Wise Sequential Neural Network for Atrial Fibrillation Classification in Adhesive Patch-Type Electrocardiographs," *IEEE Journal of Biomedical and Health Informatics*, 2024, doi: 10.1109/JBHI.2024.3411056.

Abstract in Korean

패치형 웨어러블 심전도계를 이용한 부정맥 검사를 하기 위한 딥러닝 기법

패치형 부착형 심전도계는 편리함과 더 늘어난 데이터 측정기간으로 부정맥 검사에 홀터 심전계를 대체하며 점점 더 보편화되고 있다. 이러한 웨어러블 장비들이 진료환경에서 더욱 보편적으로 사용됨에 따라 심장 질환의 보다 정확한 진단과 질병의 조기 발견 및 치료가능성이 증가하고 있다. 그럼에도 불구하고 패치형 심전도계의 단일 리드 구조와 노이즈에 대한 취약성으로 인하여, 기존에 12 리드 심전도 또는 홀터 심전계에서 측정한 데이터를 분석하는데 사용되던 부정맥 진단 기법들을 적용할 경우 그 유효성이 입증되지 않았다. 이 연구는 이러한 패치형 심전도계를 사용하였을 때 부정맥 분류 성능을 향상시키기 위한 방법을 개발하고자 한다.

첫째로, 심전도 신호의 RR 간격 프레임을 사용하여 심방세동을 비트 단위로 분류하는 SeqAFNet 을 설계하였다. 이 모델은 순환신경망 구조를 기반으로 심전도 신호 각각 비트를 순차적으로 분류하기 위하여 설계되었다. 이 모델의 성능이 학습에 사용된 데이터베이스에서만 국한되지 않았다는 것을 평가하기 위하여 3 개의 서로 다른 데이터베이스에 대하여 비교하였다. SeqAFNet 은 심방세동을 분류에 확고한 성능을 보여주며, 임상 기준에 대하여

2020 유럽 심장학회 심방세동 분류와 관리에 관한 가이드라인과 IEC 60601-2-47 표준에 부합하였다.

패치형 심전도계의 노이즈로 인한 문제를 해결하기 위해, SE-ResNet-ViT 하이브리드 모델을 개발하였다. 이 모델의 SE-ResNet 인코더는 심전도 신호의 특징을 효과적으로 추출할 수 있으며, 트랜스포머 구조는 10 초 단위의 심전도 신호 안에서 노이즈 구간에 집중하여 분류할 수 있다. 이러한 하이브리드 구조 덕분에 제안된 모델은 노이즈만 있는 신호뿐만 아니라 노이즈와 함께 측정된 다양한 부정맥 신호 또한 분류할 수 있다.

이 연구에서 제안된 방법들은 웨어러블 기기를 이용하여 부정맥 검사와 관리 분야를 발전하는데 주요한 역할을 할 것으로 기대한다. 정확하고 높은 성능을 갖고 있는 딥러닝 기반의 기법들이 임상환경에서 의료진들의 업무량을 간소화할 수 있을 것으로 생각한다. 또한, 임상 환경에서 더 정확한 부정맥 진단과 심장 질환의 조기 치료에 기여할 수 있을 것으로 기대한다.

Key words: 딥러닝, 부정맥, 심방세동, 웨어러블 디바이스, 패치형 심전도계, 순환신경망, 비전 트랜스포머