Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Diagnosis of atrial fibrillation based on AI-detected anomalies of ECG segments

Sanghoon Choi [a,b,1], Kyungmin Choi [a,b,1], Hong Kyun Yun [a,b], Su Hyeon Kim [a,b], Hyeon-Hwa Choi [a,b], Yi-Seul Park [a,b], Segyeong Joo [a,b,*]

[a] *Department of Biomedical Engineering, University of Ulsan College of Medicine, Seoul, Republic of Korea*
[b] *Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, Seoul, Republic of Korea*

ARTICLE INFO

Keywords:
Anomaly detection
Electrocardiogram
Atrial fibrillation
Clinical evidence
Autoencoder

ABSTRACT

Early detection of atrial fibrillation (AF) is crucial for its effective management and prevention. Various methods for detecting AF using deep learning (DL) based on supervised learning with a large labeled dataset have a remarkable performance. However, supervised learning has several problems, as it is time-consuming for labeling and has a data dependency problem. Moreover, most of the DL methods do not provide any clinical evidence to physicians regarding the analysis of electrocardiography (ECG) for classification or detection of AF. To address these limitations, in this study, we proposed a novel AF diagnosis system using unsupervised learning for anomaly detection with three segments, PreQ, QRS, and PostS, based on the normal ECG. Two independent datasets, PTB-XL and China, were used in three experiments. We used a long short-term memory (LSTM)-based autoencoder to train the segments of the normal ECG. Based on the threshold of anomaly scores using mean squared error (MSE), it distinguished between normal and AF segments. In Experiment A, the best score was that of PreQ, which detected AF with an AUROC score of 0.96. In Experiment B and C for cross validation of each dataset, the best scores were also of PreQ, with AUROC scores of 0.9 and 0.95, respectively. To verify the significance of the anomaly score in distinguishing between AF and normal segments, we utilized an XG-Boosted model after generating anomaly scores in the three segments. The XG-Boosted model achieved an AUROC score of 0.98 and an F1 score of 0.94. AF detection using DL has been controversial among many physicians. However, our study differentiates itself from previous studies in that we can demonstrate evidence that distinguishes AF from normal segments based on the anomaly score.

1. Introduction

Atrial fibrillation (AF) is a critical disease and the most common arrhythmia worldwide. The prevalence of AF has attained 37,574 million cases, increasing by 33 % during the last 20 years [1]. AF can lead to severe health problems such as ischemic stroke and heart failure [2]. As the aging population grows, the prevalence of AF increases, emphasizing the growing significance of early detection [3]. In addition, early detection of AF is crucial for initiating appropriate therapeutic interventions, which in turn can mitigate the risk of

[*] Corresponding author. Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, 388-1 Pungnap-dong, Songpa-gu, Seoul, Republic of Korea.

E-mail address: sgjoo@amc.seoul.kr (S. Joo).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2023.e23597
Received 11 October 2023; Received in revised form 6 December 2023; Accepted 7 December 2023
Available online 12 December 2023

potential complications associated with this condition [4,5]. Electrocardiography (ECG) signals are commonly utilized for the diagnosis of heart diseases owing to their ability to identify heart disease through morphology and rhythm [6]. ECG morphology consists of the P wave, QRS complex, and T wave, which are recorded sequentially for each cardiac cycle. The P wave represents the atrial depolarization and should have a height that does not exceed 2.5 mm and a width not greater than 0.1 s in a normal ECG. The QRS complex corresponds to ventricular depolarization and should have a width not exceeding 0.095 s in a normal ECG. The T wave is generated during ventricular repolarization. When diagnosing AF, physicians typically observe the ECG signal and analyze various ECG characteristics, such as heart rate, rhythm, P-wave, QRS complex, and T-wave, based on normal signals [7,8]. Using these parameters as the basis of their analysis allows them to make an accurate diagnosis.

In recent years, deep learning (DL) methods have demonstrated remarkable performance in accurately diagnosing AF and have the potential to assist physicians in making faster and more accurate diagnoses [9–12]. Most of the studies to detect AF use supervised learning methods. These approaches based on a large amount of labeled data have been successful. However, the results can be expensive and time-consuming to obtain. Due to these limitations, unsupervised learning approaches are also being actively researched for detecting AF using ECG data [13,14].

Anomaly detection is a type of unsupervised learning with large datasets. Anomalies in ECG signals can include various arrhythmias that differ from normal ECG signals. Various DL models in anomaly detection, such as the autoencoder (AE) and generative adversarial network (GAN), have been introduced using large volume of normal ECG signals for detecting or predicting abnormal ECG. The AE model is usually utilized in anomaly detection. Thill et al. [14] achieved an F1 score of 0.92 using a temporal convolutional autoencoder model on the MIT-BIH Arrhythmia database. Jang et al. [13] also used a convolutional variational AE model for detecting four types of anomalies based on normal signals, which achieved an F1 score of 0.86 in all types and 0.76 in AF using an ECG record of 10 s. To preserve the temporal feature in ECG signals, Hou et al. [15] trained a model as a long short-term memory (LSTM)-based AE to distinguish normal and abnormal ECG signals with the MIT-BIH Arrhythmia database, which achieved an average accuracy of 0.994. Additionally, GAN-based models have been studied and their performance in anomaly detection has improved. Zhu et al. [16] used an LSTM-GAN model for anomaly detection to distinguish between normal and anomalous classes, which performed an accuracy of 0.81. Qin et al. [17] used an ECG-ADGAN model trained with normal ECG and detected anomaly ECG in an MIT-BIH database, which achieved an F1 score of 0.94. Recent studies have shown that leveraging the GAN framework in AE models yields excellent performance in anomaly detection research. Wang et al. [18] used an AE with memory module in a GAN framework to distinguish between normal and abnormal ECG and classify abnormal types, which performed an area under the receiver operating characteristic curve (AUROC) score of 0.95. In general, most of the mentioned studies have been conducted on ECG data at the 10 s, beat, or R–R interval.

Although these approaches have exhibited signi-ficant performance, the results predicted by current DL models lack clinical evidence and remain a black box to many physicians [19]. To overcome this limitation, the class activation mapping (CAM) method has been used to explain how the model was able to distinguish between different classes by focusing on distinctive features [20–23]. Additionally, the attention-based model, which selectively focuses on those parts of the input data that are most relevant for making a prediction, has also been used for detecting various arrythmias [24–27]. However, these approaches have various limitations. They rely on the input features and tend to be random, and inconsistency is critical in the medical field and diagnosis system. Therefore, these challenges should be considered for applying a DL system in a real-medical environment.

In this study, we proposed a new approach for an AF diagnosis system using anomaly detection in ECG segments that contain PreQ (before Q-wave), QRS (QRS complex), and PostS (after S-wave) segments (Fig. 1).

The contributions of our study are as follows:

- We demonstrate that using unsupervised learning for anomaly detection can overcome the challenges related to time-consuming data labeling and limited availability of labeled datasets.
- To facilitate clinical interpretation of results predicted by DL models, we divided the ECG signals into PreQ, QRS, and PostS segments, and calculated an anomaly score to distinguish between normal and AF segments.

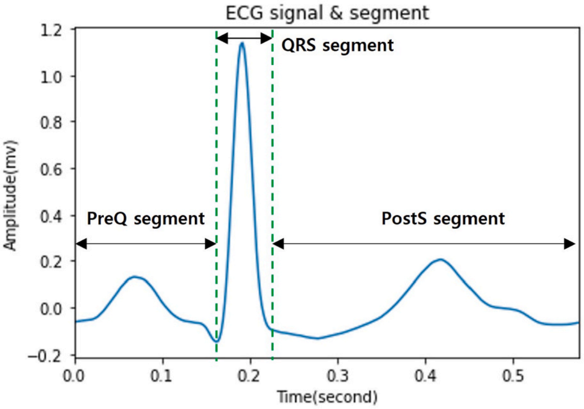


Fig. 1. Segments of ECG beat.

- We proposed an AF diagnosis system that diagnoses AF by comparing the anomaly scores in each ECG segment, which is the first of its type and offers more reliability to physicians and the medical field in general.

2. Material and methods

Fig. 2 shows an overview of the overall study, which comprises preprocessing, training dataset for the model, DL model, and anomaly detection. We provide a detailed description of the datasets used in our study in Section 2.1, the preprocessing steps for training our model are described in Section 2.2, our proposed AE model-based LSTM is explained in Section 2.3, and the method for evaluation of the model for anomaly detection is explained in Section 2.4.

2.1. Dataset

Our study utilized two public datasets, the PTB-XL dataset and the China dataset provided by PhysioNet. The PTB-XL dataset was recorded from 1989 to 1996 in Germany, and it comprises 21,837 clinical 12-lead ECG records from 18,885 patients [28]. The China dataset was collected by the GE MUSE ECG system in Shaoxing People's hospital of China and consists of 10,646 patients aged 50 or above; it corresponds to more than 60 % of the records [29]. Both datasets were recorded 10 s and sampled at 500 Hz. Our dataset has 7528 normal ECG records and 1514 records of AF from the PTB-XL dataset, whereas the China dataset contains 5419 normal ECG records and 1780 records of AF. We used only the lead II from both datasets, which was selected because it is known to provide a clear and reliable representation of cardiac electrical activity and is widely used in clinical practice. As shown Fig. 3, the training and test dataset were split at a ratio of 8:2; the training set was split into a validation set at a ratio of 8:2 during the training phase. We implemented three experiments. Experiment A used two independent datasets to train and evaluate the model. To verify the generalization of our model, we trained and evaluated the model on each of the two datasets separately to perform cross-validation in Experiment B and C.

2.2. Preprocessing

The preprocessing included three parts: filtering, dividing segments, and normalization. Both datasets were filtered using a bandpass filter to remove baseline wandering and high frequency noise. We used a 4th order Butterworth bandpass filter, with a frequency range set from 0.5 to 50 Hz.

After applying a filtering process, we employed the Pan–Tompkins algorithm to identify R-peaks in 10 s ECG recordings [30]. The identified R-peaks were used to segment the ECG recordings into individual ECG beats based on a 1:2 R–R interval ratio between the preceding and subsequent R peaks. We then calculated the Q peak and S peak based on the R peak identified earlier to separate the PreQ, QRS, and PostS segments. To extract the Q peak, we first selected the signal from 0.08 s before the R peak and then computed the gradient at each point. The Q peak was identified as the point where the slope of the Q–R segment changed from positive to negative for the first time. Similarly, the S peak was identified by extracting the signal from the R peak to 0.1 s after the R peak and identifying the point where the slope of the R–S segment transitioned from negative to positive for the first time. However, these methods for finding

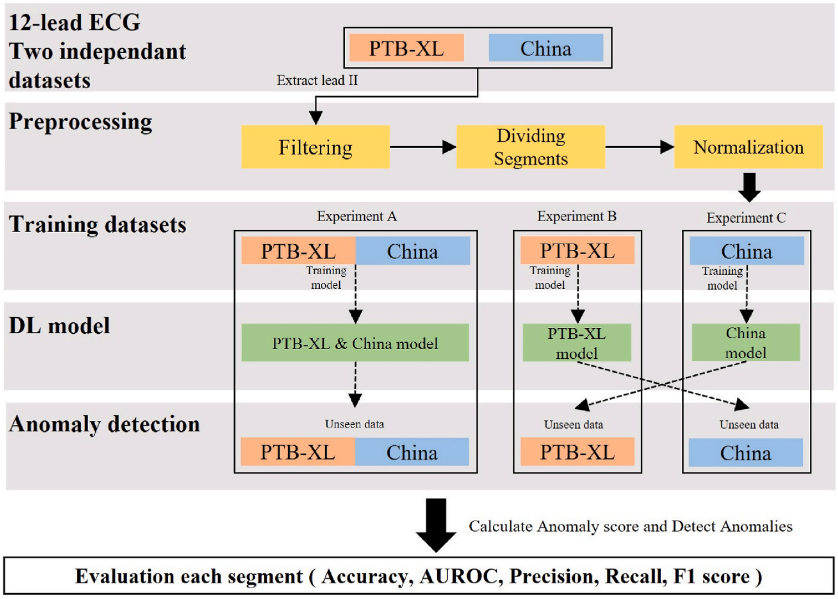


Fig. 2. Overview of the experiment.

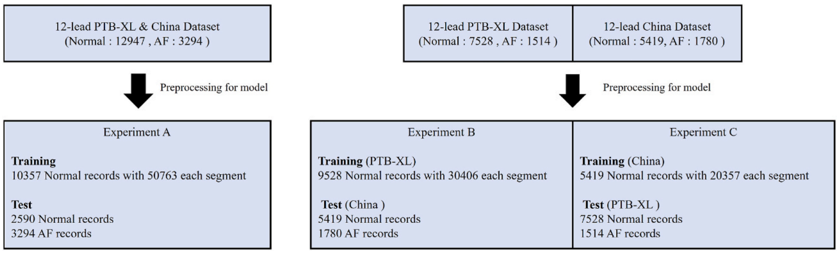


Fig. 3. Dataset split in each experiment.

the Q and S peaks have a drawback in some cases where the ECG beat has ST-segment patterns such as ascending, horizontal, or descending, or a J wave. These patterns can cause errors where the S peak is delayed beyond its expected location. To solve this problem, we detected these cases by calculating the slope between the R peak and R–S points and identifying the slope of the expected S peak as the smallest value. If this is the case, we determined the baseline of the ECG beat and calculated the intersection of the R–S signal and the baseline to obtain the location of the modified S peak.

Due to different length of PreQ, QRS, and PostS in each record, we opted to select only those beats that fell within the top 95 % and bottom 5 % of the distribution for each segment value, which excluded the segments that were either too short or long. Fig. 4 shows the distribution of each segment, and the selected ranges for analysis were from 77 to 174 samples for PreQ, from 28 to 62 samples for QRS, and from 167 to 362 samples for PostS. All the segments were normalized by applying min-max normalization to scale the data to within [−1,1]. The normalization was calculated as Equation (1).

$$Scale = \frac{Segment - \min\ (Segment)}{\max(Segment) - \min\ (segment)} \times 2\ - 1 \tag{1}$$

To standardize the length of each segment, we applied zero-padding to the length of each segment to make it the nearest power of 2, resulting in lengths of 256, 64, and 512.

2.3. Deep learning model

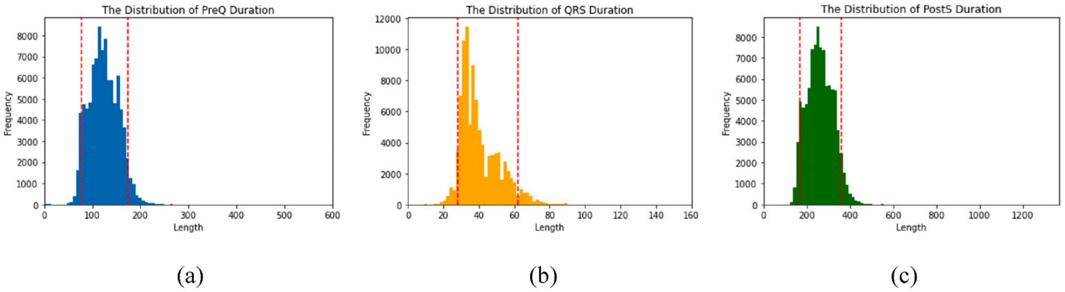
We trained three AE models based on LSTM according to each segment. LSTM is a type of RNN model that processes sequential data and selectively stores and extracts important information from the input data, allowing it to capture long-term dependencies. LSTM has shown excellent performance in time-series data, and ECG data exhibits the characteristics of time-series data. The AE model is a neural network architecture that consists of two main parts: an encoder and decoder. The encoder takes an input $x_t = [x^1,\cdots ,x^t]$ (*t:length of input*) and compresses it into a lower-dimensional representation as latent space $Z^d = [Z^1, \cdots Z^d]$(*d:dimension of latent space*) through a series of LSTM layers. The encoder captures the most important features of the input, which represent as Equation (2).

$$h_{output\ of\ encoder} = f(w_i x + b_i\ ) = f(Z) \tag{2}$$

where *f* is the activation function (a hyperbolic tangent (tanh) activation function was used). ***w*** is the weight matrix, ***b*** is the bias, and Z is the latent space. The decoder reconstructed the original input by generating through a series of LSTM layers that unsampled the compressed representation, which is represented as Equation (3)

$$\hat{x} = f^{'}\left(w_j h + b_j\right) \tag{3}$$

where $f^{'}$ is the activation function used in the same manner as the encoder. ***w*** is the weight matrix, ***b*** is the bias for decoder and ***h*** is the input of decoder.



(a) (b) (c)

Fig. 4. Distributions of length in each segment. (a), (b), and (c) represent the distribution of length in PreQ, QRS, and PostS respectively.

We utilized the reconstruction loss as the mean squared error (MSE) that is calculated to minimize the difference between the original signal and reconstruction signal generated by the decoder. The loss function is expressed as Equation (4),

$$Reconstruction\ Loss\ (MSE) = \frac{1}{n}\sum_{k=1}^{n}\left(x^{k} - \widehat{x}^{\,k}\right)^{2} \tag{4}$$

where n is the total number of input signals. In our study, we selected as anomaly score the MSE function, which is often used in anomaly detection tasks. Fig. 5 shows the LSTM-based autoencoder model, which takes as input the PreQ, QRS, and PostS segments. The number of units in the LSTM layer starts at 64 in the encoder and decreases by half at each layer. In the decoder, it starts at 16 and doubles in size at each layer during training. Table 1 presents the hyperparameters of the model for each experiment. The batch size and learning rate were set to 64 and 0.0005, respectively.

2.4. Anomaly detection

To evaluate model, we computed the anomaly score for each PreQ, QRS, and PostS segment of the 10 s test data. The anomaly scores were calculated using the MSE for the non-zero padding parts of each segment. Subsequently, the average score in each record was used for evaluation. We determined the threshold for detecting anomalies in the normal ECG utilizing the best threshold calculated using the Youden index based on the AUROC score.

3. Results

As evaluation metrics, accuracy, precision, recall, and F1 score were used, which are calculated as Equations (5)–(8).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1\ score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{8}$$

where TP represents the number of true positive predictions, FP represents the number of false positive predictions, FN represents the number of false negative predictions, and TN represents the number of true negative predictions. We used micro-average scores for all experiments. In addition, the AUROC, which provides the model with performance ability to discriminate between positive and negative classes across all possible classification thresholds, was used for evaluation. In Table 2, in Experiment A, the ratios of the differences in anomaly scores for each class were 14.4, 2.3, and 4.6, respectively, which indicated that PreQ is the most different segment for normal and.AF. In Table 3 and Fig. 6, with each threshold, the PreQ segment exhibited the highest performance with an AUROC score of 0.96. Meanwhile, the QRS segment had an AUROC score of 0.75. This indicates that the most significant differences between the normal and AF rhythm were found in the ST and TP intervals during the R–R interval, rather than in the QRS complex. This was determined through analysis of the anomaly scores, indicating that these intervals are the most distinguishing factors between the two classes. We also conducted cross-validation using two independent datasets, the PTB-XL and China datasets. In Experiment B,

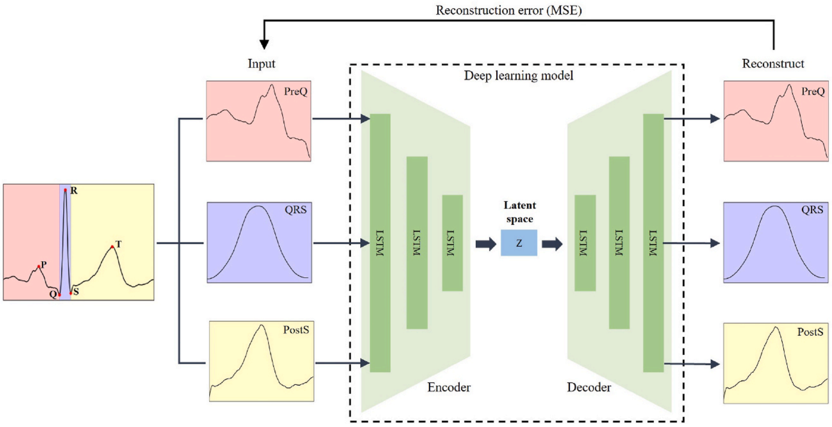


Fig. 5. The deep learning model in our proposed study.

Table 1
The Hyperparameters of our model in this study.

Experiment	Activation function	Loss function	Batch size	Learning rate	Epoch
Experiment A Experiment B Experiment C	Tangent hyperbolic	MSE	64	0.0005	200

Table 2
The Anomaly scores of three segments in each Experiment, and threshold based each segment in normal.

Experiment	Components	Normal	AFIB	Threshold
Experiment A	PreQ	0.00126	0.0182	0.00284
	QRS	0.0247	0.056	0.0784
	PostS	0.0184	0.086	0.0251
Experiment B	PreQ	0.00393	0.0311	0.00774
	QRS	0.0208	0.123	0.0011
	PostS	0.0486	0.143	0.0543
Experiment C	PreQ	0.00423	0.0346	0.00863
	QRS	0.0279	0.167	0.000693
	PostS	0.036	0.149	0.0914

Table 3
The results of classification using anomaly score.

Experiment	Components	AUROC	Precision	Recall	F1 score	Accuracy
Experiment A	PreQ	0.96	0.92	0.92	0.92	0.92
	QRS	0.75	0.69	0.7	0.7	0.7
	PostS	0.95	0.9	0.89	0.9	0.9
Experiment B	PreQ	0.9	0.84	0.84	0.84	0.84
	QRS	0.76	0.7	0.7	0.7	0.7
	PostS	0.89	0.79	0.79	0.79	0.79
Experiment C	PreQ	0.96	0.9	0.9	0.9	0.9
	QRS	0.74	0.56	0.56	0.56	0.56
	PostS	0.95	0.87	0.87	0.87	0.87

the PTB-XL dataset used the training model and evaluated the China dataset. The results in Table 3 show that the AUROC scores of PreQ, QRS, and PostS were also 0.9, 0.76, and 0.89, respectively, which means that the most different segment in ECG signals is the PreQ segment, regardless of the datasets. In Experiment C, the China dataset used the training model and evaluated the PTB-XL dataset. The AUROC scores of each segment were 0.96, 0.74, and 0.95. These performances had a similar tendency to those of previous experiments. These results demonstrate that our proposed method performs well without relying on the characteristics of the training data. The detailed results of the error distribution, ROC curve, and confusion matrix for Experiment B and C can be found in the supplementary material section.

4. Discussion

We also experimented with the anomaly scores of three segments in the dataset used in Experiment A, which were classified using the XG-Boosted model to verify the anomaly score calculated by the DL model. As presented in Table 4 and Fig. 7, the results achieved an AUROC score of 0.98 and an F1 score of 0.94. In Table 5, we compared them with the results of previous studies that evaluated on AUROC score and F1 score because they used a different dataset at the same task [31–37]. The selected methods were feature extraction with DL or machine learning (ML) and only using DL. The DL methods, especially [9,35], have remarkable performance, with F1 score of 0.97, to detect AF automatically. Further [31], used the frequency domain in the ECG signal with the PTB-XL dataset and achieved the best performance at an AUROC of 0.98. However, these studies have not provided clinical evidence. Meanwhile, using only the PreQ score, our study detected AF with an AUROC score of 0.96, a performance comparable with that of other previous studies. Additionally, it enables clinical interpretation of abnormalities in comparison to normal ECG in the segment preceding the Q-wave.

Moreover, our study of the three anomaly scores with the XG-boosted model had the best performance at an AUROC score of 0.98, which demonstrated that the anomaly scores for the three segments were verified as effective features for detection of AF. In contrast to previous studies that lacked clinical explanations, our study revealed significant differences in PreQ and PostS compared with normal ECG patterns. However, we observed no significant impact on the QRS complex, which exhibited the highest AUROC score. Fig. 8 exhibits examples of FNs. Fig. 8(a) presents the 10 s record with the lowest anomaly score among all AF cases that remained within the thresholds based in normal class. Fig. 8(b) shows the anomaly scores in the 10 s record. While most ECG beats in AF exhibit indistinct P-

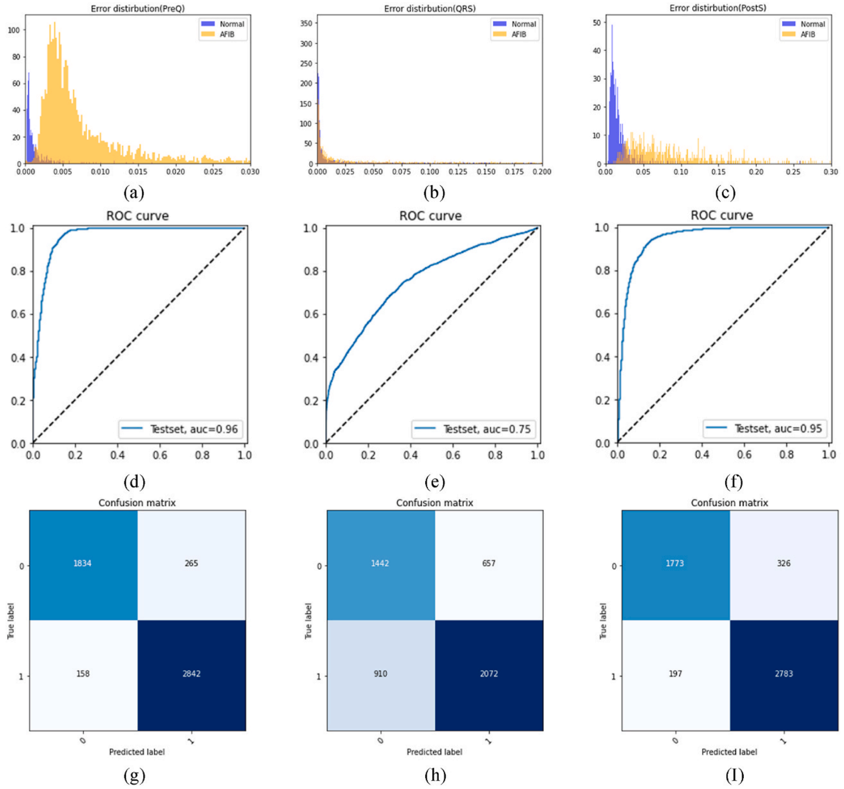


Fig. 6. Results of Experiment A. (a), (b), and (c) show results of error distribution between normal and AF case in each segment. (d), (e), and (f) show results of ROC curve in each segment. (g), (h), and (I) show results of confusion matrix. Class 0 represents Normal, while Class 1 corresponds to AF.

Table 4
The performance of XG-boosted model using three segments.

Class	Precision	Recall	F1 score	Support
Normal (0)	0.93	0.9	0.91	412
AF (1)	0.93	0.95	0.94	603
Accuracy			0.93	1015
AUROC			0.98	1015

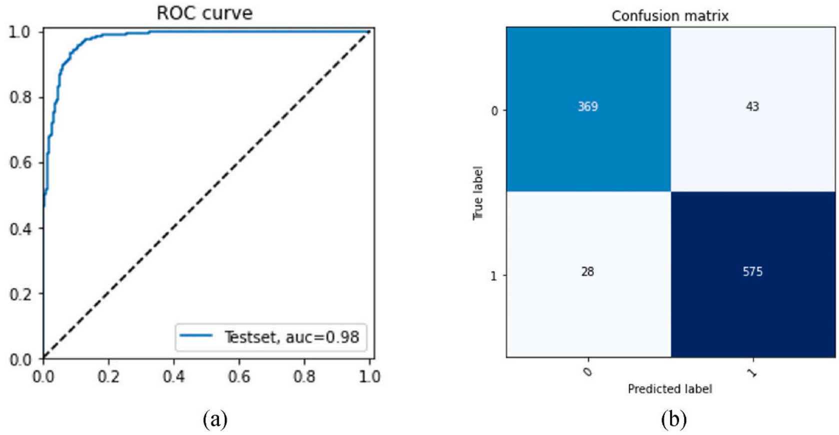


Fig. 7. Result of classification using anomaly scores based on the XG-boosted model. (a) and (b) show results of ROC curve and confusion matrix. Class 0 represents Normal, while Class 1 corresponds to AF.

Table 5
The comparison of previous study.

Study	Method	Dataset	AUROC	F1 score	Clinical explain
Kent et al. [31]	Feature extraction + DL	PTB-XL	0.98	–	x
Xu et al. [32]		MIT-BIH	0.95	–	x
Jo et al. [33]		PTB-XL	0.97	0.93	o
B Chen et al. [34]	DL	Own dataset	0.98	–	x
Anderson et al. [9]		MIT-BIH	0.94	0.97	x
Petmezas et al. [35]		MIT-BIH	–	0.97	x
Kropf et al. [36]	Feature extraction + ML	CINC (2017) dataset	–	0.81	x
Czabanski et al. [37]		MIT-BIH	–	0.97	x
Our study		PTB-XL + China dataset	0.98	0.94	o
Our study	Anomaly score (PreQ)	PTB-XL + China dataset	0.96	0.92	o

wave morphology, this case presented slight P-wave morphology. In addition, it is possible for the T-wave of the preceding beat to overlap with the post best when creating a beat from a 10 s ECG recording due to the irregular R–R interval. This result indicated that not all beats demonstrate AF within a 10 s timeframe at the beat level. We propose a new approach for an AF diagnosis system that utilizes an explainable model to assist physicians clinically and addresses the limitations of the black box nature of DL models. Despite the contribution of this study, there are several limitations that need to be considered and addressed in future research. First, we utilized a single lead ECG, specifically lead II. Although lead II has important features for diagnosing AF, other arrhythmias may require additional leads in the ECG.

Second, our study defined anomalies exclusively as AF; however, it is necessary to address various other cardiac diseases beyond AF in the future study. Third, in our research, although the defined segments, PreQ, QRS, and PostS, are crucial in determination of cardiac diseases, it is also important to consider evaluating more precise segments, such as the PR interval and QT segment. Finally, the dataset utilized in our study comprises records contained in 10 s rather than continuous records, which may not be conducive to continuous AF diagnosis. Therefore, it would be developed to apply out study to continuous long-term data, such as Holter data, to monitor three segments for anomaly scores.

5. Conclusion

The AF diagnosis system using DL has been controversial to many physicians, even though it has achieved remarkable performance to detect or predict AF. To address the problem, we proposed a novel AF diagnosis system that considers anomaly detection in segments, PreQ, QRS, and PostS, compared with the normal ECG. The highest and lowest AUROC scores were 0.96 and 0.75 in the PreQ and QRS segments, respectively. This means that the PreQ segment, which is the section from the P-wave to the Q-wave, has important features for detecting AF. Meanwhile, the QRS, which is the section from the Q-wave to the S -wave, exhibits a relatively low prevalence of anomaly between AF and normal ECG. In addition, we conducted cross-validation by training and testing the models on separate datasets. We used the independent PTB-XL and China datasets and verified that the best score of AUROC was achieved in PreQ at 0.9 and 0.96. Through cross-validation, the potential for a generalized model that can yield promising results regardless of race has been demonstrated, addressing the inherent data dependency issue in conventional DL models. For applying a DL model in the medical field, the reason for diagnosis should be explainable and medically justified to physicians and experts. In this respect, our study distinguishes itself from other previous studies in that we can clearly demonstrate the evidence that distinguishes normal from AF based on the anomaly score. In our future work, our approach will be developed for use in detection of various anomalies.

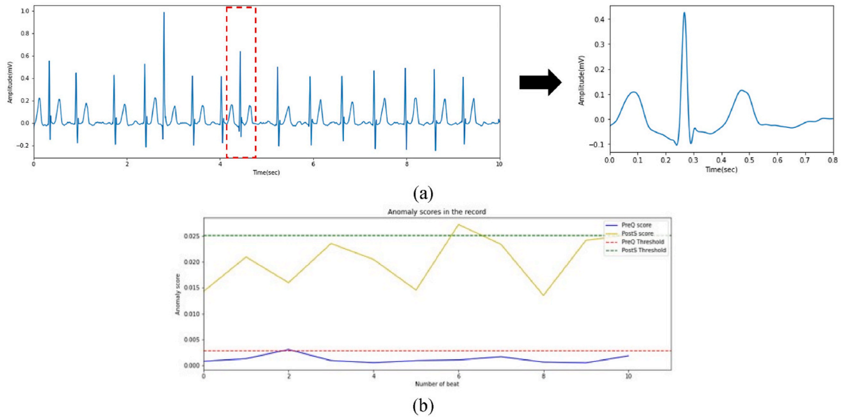


Fig. 8. Example of false negative in anomaly scores of PreQ and PostS. (a) Shows a 10-s record and the beat with the lowest anomaly score in the AF class. (b) Shows the anomaly score of PreQ and PostS in (a) recording.

Data availability statement

Datasets associated with this study have been deposited at
https://physionet.org/content/ptb-xl/1.0.3/and
https://physionet.org/content/ecg arrhythmia/1.0.0/.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Sanghoon Choi: Writing – original draft, Software, Methodology, Investigation, Data curation. **Kyungmin Choi:** Writing – review & editing, Validation, Software, Resources, Methodology, Conceptualization. **Hong Kyun Yun:** Visualization, Software. **Su Hyeon Kim:** Writing – review & editing, Software, Investigation. **Hyeon-Hwa Choi:** Validation, Software, Methodology. **Yi-Seul Park:** Data curation. **Segyeong Joo:** Supervision, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Segyeong Joo reports financial support was provided by Asan Medical Center.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF2021R1A2C1013755)

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e23597.

References

[1] G. Lippi, F. Sanchis-Gomar, G. Cervellin, Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge, Int. J. Stroke 16 (2) (2021) 217–221.
[2] S.S. Chugh, et al., Worldwide epidemiology of atrial fibrillation: a global burden of disease 2010 study, Circulation 129 (8) (2014) 837–847.
[3] M. Zoni-Berisso, et al., Epidemiology of Atrial Fibrillation: European Perspective, Clinical epidemiology, 2014, pp. 213–220.
[4] P. Kirchhof, et al., ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS, 2016, Europace 18 (11) (2016) 1609–1678.
[5] M.P. Turakhia, et al., Economic burden of undiagnosed nonvalvular atrial fibrillation in the United States, Am. J. Cardiol. 116 (5) (2015) 733–739.
[6] P.M. Rautaharju, B. Surawicz, L.S. Gettes, AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part IV: the ST segment, T and U waves, and the QT interval: a scientific statement from the American heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the American college of cardiology foundation; and the heart rhythm society: endorsed by the international society for computerized electrocardiology, Circulation 119 (10) (2009) e241–e250.
[7] W.B. Kannel, et al., Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: population-based estimates, Am. J. Cardiol. 82 (7) (1998) 2N–9N.
[8] A.T.F. Members, et al., Acc/aha/esc 2006 guidelines for the management of patients with atrial fibrillation–executive summary: a report of the american college of cardiology/american heart association task force on practice guidelines and the european society of cardiology committee for practice guidelines (writing committee to revise the 2001 guidelines for the management of patients with atrial fibrillation) developed in collaboration with the european heart rhythm association and the heart rhythm society, Eur. Heart J. 27 (16) (2006) 1979–2030.
[9] R.S. Andersen, A. Peimankar, S. Puthusserypady, A deep learning approach for real-time detection of atrial fibrillation, Expert Syst. Appl. 115 (2019) 465–473.
[10] P. Cao, et al., A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation, Biomed. Signal Process Control 56 (2020), 101675.
[11] Z.I. Attia, et al., Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction, J. Cardiovasc. Electrophysiol. 30 (5) (2019) 668–674.
[12] P. Zhang, et al., Semi-supervised learning for automatic atrial fibrillation detection in 24-hour holter monitoring, IEEE Journal of Biomedical and Health Informatics 26 (8) (2022) 3791–3801.
[13] M. Thill, et al., Temporal convolutional autoencoder for unsupervised anomaly detection in time series, Appl. Soft Comput. 112 (2021), 107751.
[14] J.-H. Jang, et al., Unsupervised feature learning for electrocardiogram data using the convolutional variational autoencoder, PLoS One 16 (12) (2021), e0260612.
[15] B. Hou, et al., LSTM-based auto-encoder model for ECG arrhythmias classification, IEEE Trans. Instrum. Meas. 69 (4) (2019) 1232–1240.
[16] G. Zhu, et al., A novel LSTM-GAN algorithm for time series anomaly detection, in: 2019 Prognostics and System Health Management Conference (PHM-Qingdao), IEEE, 2019.
[17] J. Qin, et al., A novel temporal generative adversarial network for electrocardiography anomaly detection, Artif. Intell. Med. (2023), 102489.
[18] Z. Wang, S. Stavrakis, B. Yao, Hierarchical deep learning with Generative Adversarial Network for automatic cardiac diagnosis from ECG signals, Comput. Biol. Med. 155 (2023), 106641.
[19] D. Jin, et al., Explainable deep learning in healthcare: a methodological survey from an attribution view, WIREs Mechanisms of Disease 14 (3) (2022) e1548.

[20] S. Sawano, et al., Deep learning model to detect significant aortic regurgitation using electrocardiography, J. Cardiol. 79 (3) (2022) 334–341.

[21] N. Sobahi, et al., Explainable COVID-19 detection using fractal dimension and vision transformer with Grad-CAM on cough sounds, Biocybern. Biomed. Eng. 42 (3) (2022) 1066–1080.

[22] V. Jahmunah, et al., Explainable detection of myocardial infarction using deep learning models with Grad-CAM technique on ECG signals, Comput. Biol. Med. 146 (2022), 105550.

[23] S. Vijayarangan, et al., Interpreting deep neural networks for single-lead ECG arrhythmia classification, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020.

[24] P. Singh, A. Sharma, Attention-based convolutional denoising autoencoder for two-lead ECG denoising and arrhythmia classification, IEEE Trans. Instrum. Meas. 71 (2022) 1–10.

[25] Z. Liu, X. Zhang, ECG-based heart arrhythmia diagnosis through attentional convolutional neural networks, in: 2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS), IEEE, 2021.

[26] Y. Zhao, et al., An explainable attention-based TCN heartbeats classification model for arrhythmia detection, Biomed. Signal Process Control 80 (2023), 104337.

[27] Q. Yao, et al., Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network, Inf. Fusion 53 (2020) 174–182.

[28] P. Wagner, et al., PTB-XL, a large publicly available electrocardiography dataset, Sci. Data 7 (1) (2020) 154.

[29] J. Zheng, et al., A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients, Sci. Data 7 (1) (2020) 48.

[30] L. Sathyapriya, L. Murali, T. Manigandan, Analysis and detection R-peak detection using Modified Pan-Tompkins algorithm, in: 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, IEEE, 2014.

[31] M. Kent, et al., Fourier space approach for convolutional neural network (CNN) electrocardiogram (ECG) classification: a proof-of-concept study, J. Electrocardiol. 80 (2023) 24–33.

[32] X. Xu, et al., Atrial fibrillation beat identification using the combination of modified frequency slice wavelet transform and convolutional neural networks, Journal of healthcare engineering 2018 (2018).

[33] Y.-Y. Jo, et al., Detection and classification of arrhythmia using an explainable deep learning model, J. Electrocardiol. 67 (2021) 124–132.

[34] B. Chen, et al., A deep learning model for the classification of atrial fibrillation in critically ill patients, Intensive Care Medicine Experimental 11 (1) (2023) 1–10.

[35] G. Petmezas, et al., Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets, Biomed. Signal Process Control 63 (2021), 102194.

[36] M. Kropf, D. Hayn, G. Schreier, ECG classification based on time and frequency domain features using random forests, in: 2017 Computing in Cardiology (CinC), IEEE, 2017.

[37] R. Czabanski, et al., Detection of atrial fibrillation episodes in long-term heart rhythm signals using a support vector machine, Sensors 20 (3) (2020) 765.