Paroxysmal Atrial Fibrillation Detection by Combined Recurrent Neural Network and Feature Extraction on ECG Signals

Xinqi Bao[1][a], Fenghe Hu[1], Yujia Xu[1], Mohamed Trabelsi[2] and Ernest Kamavuako[1][b]
[1]*Department of Engineering, King's College London, London, U.K.*
[2]*Department of Electronic and Communications Engineering, Kuwait College of Science and Technology, Kuwait*

Keywords: Electrocardiogram (ECG), Paroxysmal Atrial Fibrillation (Afib), Recurrent Neural Network (RNN).

Abstract: Paroxysmal atrial fibrillation (AFib) or intermittent atrial fibrillation is one type of atrial fibrillation which occurs rapidly and stops spontaneously within days. Its episodes can last several seconds, hours, or even days before returning to normal sinus rhythm. A lack of intervention may lead the paroxysmal into persistent atrial fibrillation, causing severe risk to human health. However, due to its intermittent characteristics, it is generally neglected by patients. Therefore, real-time monitoring and accurate automatic algorithms are highly needed for early screening. This study proposes a two-stage algorithm, including a BiLSTM network to classify healthy and atrial fibrillation, followed by a feature-extraction-based neural network (NN) to identify the persistent, paroxysmal atrial fibrillation onsets. The extracted features include the entropy and standard deviation of the RR intervals. The two steps can achieve 90.14% and 92.56% accuracy in the validation sets on small segments. This overall algorithm also has the advantage of the low computing load, which shows a high potential for a portable embedded device.

1 INTRODUCTION

Atrial fibrillation (AFib) is an irregular heartbeat (arrhythmia) caused by the ectopic impulses in the atrium. It may lead to blood clots, stroke, and heart failure, which are severe hidden dangers to human lives. Furthermore, the AFib is a common issue for approximately 2% of people younger than 65 and 9% older than 65 (Kornej et al., 2020). The American Heart Association guideline (January et al., 2014) classified Afib into four types: paroxysmal AFib, persistent AFib, long-standing persistent AFib, and permanent AFib based on the duration and recoverability. While in clinics, physicians usually sort them into paroxysmal and persistent types only. Paroxysmal AFib episodes can last several seconds, hours, or even days before returning to normal sinus rhythm. Lack of intervention may lead the paroxysmal into persistent AFib, which is irreversible. Due to the intermittent characteristics of the paroxysmal AFib, it is generally neglected by patients before deteriorating into a persistent type. As a result, the all-cause mortality rate is approximately 6.3% on AFib patients (Lee et al, 2018). Therefore, it is vital to have an algorithm that can work automatically in the early screening to prevent the paroxysmal AFib from worsening to persistent AFib or more severe health issues.

Electrocardiogram (ECG) is the most commonly used approach in cardiac diagnosis. It represents the electrical activity of the heart. The whole electrical process starts with the spontaneous impulse generated at the Sinoatrial node (SA node), then propagates to the atrioventricular node (AV node), causing the squeezing of the atria as represented by the P wave. Afterwards, the electrical signal is transmitted to the His bundle and Purkinje fibres, causing the contraction of the ventricles. The ventricles will be repolarized and ready for the next heart cycle. The QRS complex indicates the depolarization, and the T wave shows the repolarization of the ventricles, respectively. However, AFib is caused by irregular fast squeezing of the atria leading the heart walls quiver, or fibrillate. This phenomenon it is reflected by disorganized electrical activity (ectopic impulses instead of SA impulse) in the atrium, so its ECG

[a] https://orcid.org/0000-0002-7117-1267
[b] https://orcid.org/0000-0001-6846-2090

Bao, X., Hu, F., Xu, Y., Trabelsi, M. and Kamavuako, E.
Paroxysmal Atrial Fibrillation Detection by Combined Recurrent Neural Network and Feature Extraction on ECG Signals.
DOI: 10.5220/0010987300003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 4: BIOSIGNALS*, pages 85-90
ISBN: 978-989-758-552-4; ISSN: 2184-4305

signal differs from normal, as shown in Figure 1. Morphologically, the AFib ECG has irregular intervals, a narrow QRS complex, and undulating P waves. Thus, using ECG signals to identify the AFib is a practical approach in designing automatic classification algorithms.

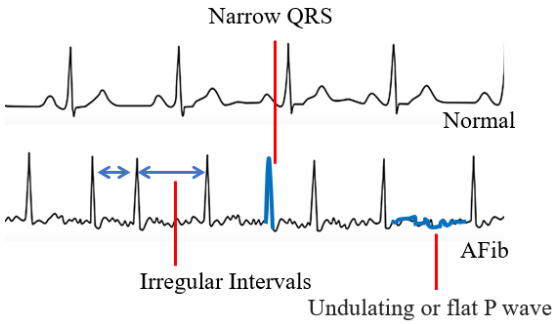

Figure 1: The cardiac cycles of normal and AFib ECG.

Computer-aided algorithms for AFib detection have been developed for decades, and the proposed algorithms covered the conventional machine learning (ML) methods such as support vector machine (SVM), k-nearest neighbours algorithm (KNN), random forest, discriminant analysis, etc (Zhou et al., 2015; De Giovanni et al., 2017; Kalidas & Tamil,2019; Pourbabaee et al., 2018; Annavarapu et al., 2016; Rizwan et al., 2020). These conventional ML approaches relied on manually extracted features such as average, standard deviation, and entropy of RR intervals in the time domain (Liu et al., 2018), power spectral density in the frequency domain, and statistical features such as kurtosis and skewness (Rizwan et al., 2020). With the development of deep learning (DL) in recent years, approaches such as convolutional neural network (CNN) and recurrent neural network (RNN) have also been tested on AFib detection (Xiong et al., 2017; Petmezas et al., 2021; Ping et al., 2020). They hold the advantage of neglecting feature extraction and using raw ECG signals as input and have also achieved promising performance. Though there are tons of researches focusing on AFib classification, only few pieces of research work have focused on paroxysmal AFib detection due to the lack of suitable databases. As a result, paroxysmal AFib is often unrecognized (Michaud & Stevenson, 2021). Therefore, it is pretty meaningful to explore the capability of the neural network (NN) in the identification of paroxysmal AFib.

In this study, the primary aim is to propose an algorithm that can classify the non-AFib, persistent AFib, paroxysmal AFib, and their onsets. The secondary task is to constrain the computing load while achieving comparable performance, making it available for a standard laptop or embedded system. All the findings will provide knowledge on using NNs to classify paroxysmal AFib and contribute to designing small-scale portable ECG devices which can do real-time monitoring of the heart conditions.

2 METHODOLOGY

2.1 Database

The database used in this research was CPSC2021 (Wang et al., 2021). It includes 1436 ECG recordings (475 Persistent AFib, 229 Paroxysmal AFib, 732 Non-AFib) from 100 subjects (24 Persistent AFib, 23 Paroxysmal AFib, 53 Non-AFib).

2.2 Proposed Algorithm

In this study, a two-stage algorithm was designed to conduct the detection of paroxysmal AFib and its onsets. The flowchart of the proposed algorithm is shown in Figure 2. In Stage I, a Bidirectional Long short-term memory (BiLSTM) network was used to classify the ECG segments into Non-AFib and AFib segments. Then the ECG signals consisting of AFib segments were transferred to Stage II and classified into Persistent AFib or Paroxysmal AFib. A moving window was employed to classify the whole signal and detect AFib onsets. The processing was conducted in Matlab® R2021a environment, using a laptop (CPU: i7-8650U, RAM: 16G, no GPU).

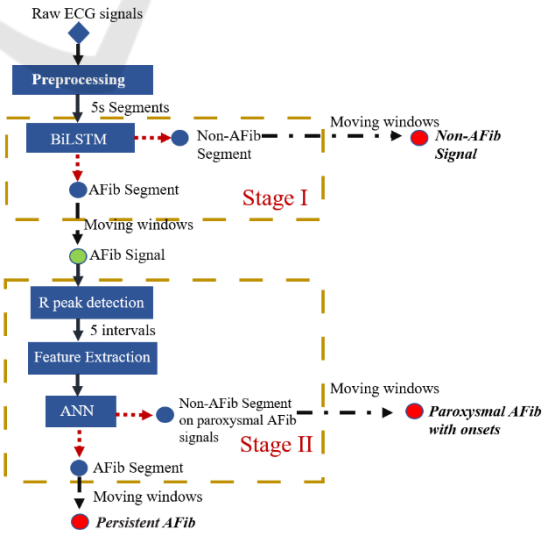

Figure 2: The flowchart of the designed algorithm.

2.2.1 Pre-processing

Before the two classification stages, the ECG signals were pre-processed. The raw ECG signals were normalized (z-score), filtered with 0.5 – 30 Hz bandpass filter (3rd order Butterworth), then segmented into 5s segments for training (without overlap). After segmentation, 699040 ECG segments were generated (421022 Non-AFib, 212098 Persistent, and 65920 Paroxysmal) for training.

2.2.2 Stage I: BiLSTM

BiLSTM is one type of RNN algorithms that showed outstanding performance in the sequence data, such as speech and text recognition (Graves et al., 2005, Liu & Guo, 2019). In the proposed algorithm, a simplified structure with two layers of BiLSTM (hidden units: 50) was applied. The inputs for the BiLSTM layers were 5s segments. After BiLSTM, it is connected with a fully connected layer to project the results into Non-AFib (0) and AFib (1) two classes. The overall structure of Stage I is shown in Figure 3(a).

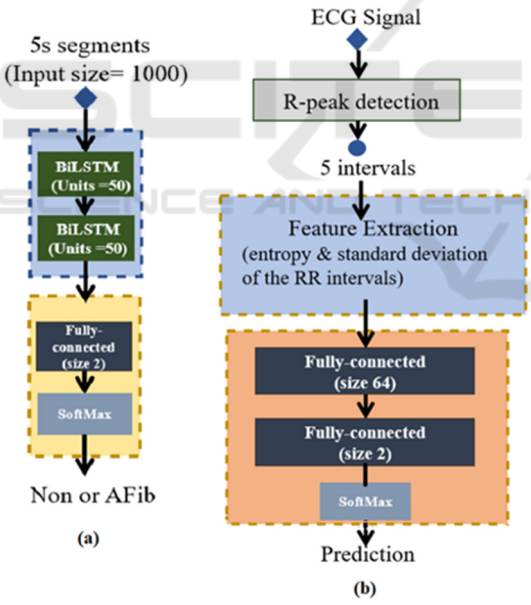

Figure 3: (a) The Stage I structure, (b) The Stage II structure.

During training, the used training sets included non-AFib segments labeled (0), persistent AFib segments labeled (1), while paroxysmal AFib segments were also labeled (1) to increase the sensitivity. Training and Validation Proportion was 7:3. 20% recordings (285) were randomly left for the whole signal testing, including 145 non-AFib, 95 persistent, and 45 paroxysmal recordings. The optimizer selected in this study was stochastic gradient descent with momentum (SGDM). The initial learning rate was 0.001 with a drop factor of 0.2, the max epoch of 10, and the batch size of 256.

The network can identify the non-AFib segments of the ECG signal. For the complete signal classification, a moving window (size: 5s, slide: 1s) was conducted on the signal to classify each segment. A majority voting was applied to avoid sudden incorrect classification. Each time frame was covered by 5 sliding windows, so the time frame is only labeled AFib when more than 3 windows (segments) were classified as AFib.

2.2.3 Stage II: Feature Extraction & ANN

In the testing phase, the use of a relatively simplified DL network (with two layers BiLSTM and three Conventional layers structure, such as Stage I) didn't perform well in the identification of paroxysmal or persistent AFib. The loss didn't go down and the training accuracy remained at 69.15%, which means the network was uncapable to learn. Deeper and complex network structure were excluded to avoid increasing the computation burden. Therefore, manual features extraction was applied in the classification stage, where entropy and standard deviation of RR intervals (which are commonly used as input features for classification) were selected. The process of Stage II is shown in Figure 3(b).

R-peaks were extracted by Pan–Tompkins algorithm (Pan and Tompkins, 1985). Five RR intervals were clipped as a segment, and the entropy and standard deviation were extracted from the segments. Afterward, they were sent to the fully connected layers to classify into non-AFib or AFib segments. Similar to Stage I, a moving window (size: 5 intervals, slide: 1 interval) was also applied to identify the whole signal as persistent or paroxysmal. The entropy calculation is given by the equation:

$$\mathrm{E}(R) = -\sum_{i=1}^{n} \mathrm{P}(R_i)\log\ \mathrm{P}(R_i)$$

where E is the entropy of the segment, R_i indicates each RR interval length and P is the occurrence probability.

The training sets were only persistent AFib labelled (1), and paroxysmal AFib segments were labelled according to the reference label. Because the paroxysmal segments were approximately 30% of persistent segments, and the non-AFib segments of the paroxysmal are less. Therefore, a moving window

(size: 5 intervals, slide: 1 interval) was applied to section more paroxysmal segments to balance the data structure. The rest training settings were the same as Stage I.

2.3 Evaluation Metrics

The validation accuracy of the two stages indicates their capability to identify the small segments (within windows). The overall performance of the algorithm can be reflected by the score of the testing recordings. In this paper, the CSPC2021 Challenge scoring scheme is considered (Wang et al., 2021).

The score includes two parts: the first part (Ur) classifies the AFib correctly, and the score matrix is shown in Figure 4. The second part (Ue) is meant to detect the AFib onsets. If the onsets and end of the AFib episodes were detected within ±1 R-peak, Ue + 1, within ±2, Ue + 0.5.

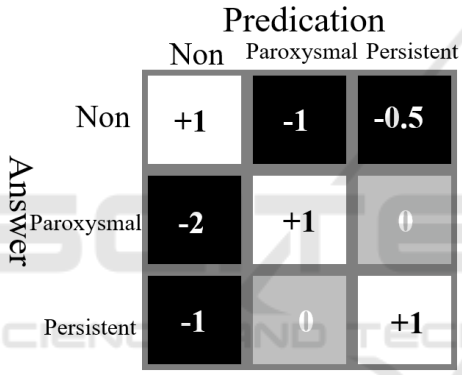

Figure 4: The score matrix for part one.

The overall score (U) is calculated by:

$$U = \frac{1}{N}\sum_{i=1}^{N}\left(Ur_i + \frac{Ma_i}{max\{Mr_i, Ma_i\}} \times Ue_i\right)$$

3 RESULTS

For Stage I, the validation sets achieved 90.14% accuracy to classify the non-AFib and AFib segments with a specificity of 93.65% and sensitivity of 84.82%, respectively. The result indicated that Stage I could identify the non-AFib segments well but may miss some AFib segments. However, it wasn't an issue for the whole signal because the majority voting and the appropriate threshold can improve the overall performance and remedy the sensitivity. In the testing phase, a 2.5% threshold was set which means if less than 2.5% of the signal is classified as AFib, the overall signal will be regarded as non-AFib. By this approach, the accuracy of non-AFib signals classification could be increased approximately from 92.62% to 96%. Theoretically, raising the threshold can improve the non-AFib accuracy on validation to almost 100%, but it will lose its sensitivity and generalization.

For Stage II, on the validation sets, it did the accuracy of 92.56% with a specificity of 86.24% and sensitivity of 95.77% to classify non-AFib and AFib segments on the AFib signals. The result showed that Stage II might tend to classify the healthy segments into AFib segments. However, because of the considered two stages design, non-AFib signals have been excluded before Stage II; thus, it won't affect the overall classification performance. It will only affect the detection of the onset of the AFib.

During the testing recordings, the two-stages method achieved 2.0953 overall mark, including 0.8714 Ur and 1.4039 Ue. It showed a satisfying performance on the classification, while the onset detection can be improved. Furthermore, the total neural network is only about 1.6 MB in Matlab (coding in Python can be smaller, approximately 500 k.), which is possible to use on a personal laptop or embedded device.

4 DISCUSSION

This study aimed to design an algorithm using NNs to detect paroxysmal AFib and make the computing load small enough for a portable embedded ECG device. This is done because patients typically neglect paroxysmal AFib due to its intermittent characteristics and lack of appropriate databases. In this study, a two-stage algorithm was designed using the CSCP2021 database, and it proofed its capability to classify the AFib segments and onsets on the validation sets.

Firstly, the use of a two-stage method rather than one NN will be justified. Before the training, our preconceived thought on paroxysmal AFib was like intermittent non-AFib and AFib waveforms in the ECG signals. However, it is not, or at least the BiLSTM or Conventional Neural Network (CNN) cannot easily learn it. For non-AFib or AFib segments from the non/persistent AFib signals, the network in Stage I can learn in a very short time within one epoch, while the segments from paroxysmal could not regress, and the loss didn't go down (training accuracy also stuck at 69.15%, which is approximately equal to the data proportion). This may indicate that the paroxysmal AFib may hold

pathological characteristics even in the healthy episodes and using a simplified network cannot classify the non-AFib or AFib episodes. There is no doubt that using the deeper neural network with complex structure, such as adding lots of CNN layers and attention layers, will learn the difference. Still, it will make the computing load quite extensive, which is contrary to the original intention. Therefore, a second phase was included for the detection of the paroxysmal onset.

Secondly, the use of Stage II to finish the whole classification task is tested. However, the performance was not satisfying due to the oversensitivity of the Stage II network and its trend to identify the segment as AFib. Besides, feature extraction relies greatly on reliable and accurate R peak detection. When the signal has massive motion artefacts, the failed R peak detection will cause an error in the algorithm. This is another advantage of the two-stage structure.

Thirdly, there is still room for the improvement of the overall performance. In the blind test of the challenge, the overall mark is decreased from 2 to approximately 1.7. This result showed that the generalization needs to be improved, especially in Stage II. Currently, only two features were used while adding more features might be a solution to improve the algorithm. Besides, appropriate window length may also affect the result. Currently, a 5s window on Stage I and five intervals on Stage II are used. Longer window length may provide more information, especially on the feature extraction of Stage II. Short duration cannot maximize the feature difference.

5 CONCLUSIONS

This study proposed a two-stage neural network algorithm that can detect paroxysmal AFib and its onsets. For performance, it can achieve 90.14% and 92.56% accuracy on non-AFib and AFib segments classification respectively in the two stages, got 2.0953 overall mark on our testing sets. As few researches have focused on paroxysmal AFib detection using NNs, the finding of this study will provide knowledge for the further researches in this area. In the meantime, the proposed method also holds the advantage of a small computing load, making it possible for embedded ECG devices.

ACKNOWLEDGEMENT

This work has been funded in part from KFAS, Kuwait Foundation for Advancement of Sciences, project no. CN20-13EE-01.

REFERENCES

January, C. T., Wann, L. S., Alpert, J. S., Calkins, H., Cigarroa, J. E., Cleveland, J. C., ... & Yancy, C. W. (2014). 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *Journal of the American College of Cardiology*, 64(21), e1-e76.

Kornej, J., Börschel, C. S., Benjamin, E. J., & Schnabel, R. B. (2020). Epidemiology of atrial fibrillation in the 21st century: novel methods and new insights. *Circulation research*, 127(1), 4-20.

Zhou, X., Ding, H., Wu, W., & Zhang, Y. (2015). A real-time atrial fibrillation detection algorithm based on the instantaneous state of heart rate. *PloS one*, 10(9), e0136544.

De Giovanni, E., Aminifar, A., Luca, A., Yazdani, S., Vesin, J. M., & Atienza, D. (2017, September). A patient-specific methodology for prediction of paroxysmal atrial fibrillation onset. In 2017 *Computing in Cardiology (CinC)* (pp. 1-4). IEEE.

Pourbabaee, B., Roshtkhari, M. J., & Khorasani, K. (2018). Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(12), 2095-2104.

Annavarapu, A., & Kora, P. (2016). ECG-based atrial fibrillation detection using different orderings of Conjugate Symmetric–Complex Hadamard Transform. *International Journal of the Cardiovascular Academy*, 2(3), 151-154.

Rizwan, A., Zoha, A., Mabrouk, I. B., Sabbour, H. M., Al-Sumaiti, A. S., Alomainy, A., ... & Abbasi, Q. H. (2020). A review on the state of the art in atrial fibrillation detection enabled by machine learning. *IEEE reviews in biomedical engineering*, 14, 219-239.

Liu, C., Oster, J., Reinertsen, E., Li, Q., Zhao, L., Nemati, S., & Clifford, G. D. (2018). A comparison of entropy approaches for AF discrimination. *Physiological measurement*, 39(7), 074002.

Kalidas, V., & Tamil, L. S. (2019). Detection of atrial fibrillation using discrete-state Markov models and Random Forests. *Computers in biology and medicine*, 113, 103386.

Xiong, Z., Stiles, M. K., & Zhao, J. (2017, September). Robust ECG signal classification for detection of atrial fibrillation using a novel neural network. In 2017 *Computing in Cardiology (CinC)* (pp. 1-4). IEEE.

Petmezas, G., Haris, K., Stefanopoulos, L., Kilintzis, V., Tzavelis, A., Rogers, J. A., ... & Maglaveras, N. (2021). Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets. *Biomedical Signal Processing and Control*, 63, 102194.

Ping, Y., Chen, C., Wu, L., Wang, Y., & Shu, M. (2020, June). Automatic detection of atrial fibrillation based on CNN-LSTM and shortcut connection. In *Healthcare* (Vol. 8, No. 2, p. 139). Multidisciplinary Digital Publishing Institute.

Michaud GF, Stevenson WG. (2021) Atrial Fibrillation. *N Engl J Med*. , 384(4):353-361.

Wang, X., Ma, C., Zhang, X., Gao, H., Clifford, G., & Liu, C. Paroxysmal Atrial Fibrillation Events Detection from Dynamic ECG Recordings: The 4th China Physiological Signal Challenge 2021.

Graves, A., Fernández, S., & Schmidhuber, J. (2005, September). Bidirectional LSTM networks for improved phoneme classification and recognition. In *International conference on artificial neural networks* (pp. 799-804). Springer, Berlin, Heidelberg.

Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing*, 337, 325-338.

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering*, (3), 230-236.

Lee, E., Choi, E. K., Han, K. D., Lee, H., Choe, W. S., Lee, S. R., ... & Oh, S. (2018). Mortality and causes of death in patients with atrial fibrillation: a nationwide population-based study. *PLoS One*, 13(12), e0209687.