

# ECGNET: Learning Where to Attend for Detection of Atrial Fibrillation with Deep Visual Attention

Sajad Mousavi, Fatemeh Afghah, Abolfazl Razi

School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, USA

U. Rajendra Acharya

Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore

**Abstract**—The complexity of the patterns associated with atrial fibrillation (AF) and the high level of noise affecting these patterns have significantly limited the application of current signal processing and shallow machine learning approaches to accurately detect this condition. Deep neural networks have shown to be very powerful to learn the non-linear patterns in various problems such as computer vision tasks. While deep learning approaches have been utilized to learn complex patterns related to the presence of AF in electrocardiogram (ECG) signals, they can considerably benefit from knowing which parts of the signal is more important to focus on during learning. In this paper, we introduce a two-channel deep neural network to more accurately detect the presence of AF in the ECG signals. The first channel takes in an ECG signal and automatically learns where to attend for detection of AF. The second channel simultaneously takes in the same ECG signal to consider all features of the entire signal. Besides improving detection accuracy, this model can guide the physicians via visualization that what parts of the given ECG signal are important to attend while trying to detect atrial fibrillation. The experimental results confirm that the proposed model significantly improves the performance of AF detection on well-known MIT-BIH AF database with 5-s ECG segments (achieved a sensitivity of 99.53%, specificity of 99.26% and accuracy of 99.40%).

**Index Terms**—Atrial fibrillation, ECG analysis, deep learning, attention mechanism.

## I. INTRODUCTION

Atrial fibrillation (AF) is the most prevalent type of arrhythmia leading to hospital admissions, and is currently affecting the lives of more than 3 million people in the U.S. and over 33 million worldwide, while the number of AF patients in the US is expected to double by 2050 [1]. AF's incidence is associated with an increase in the risk of stroke, congestive heart failure and overall mortality. This condition is commonly diagnosed by analyzing the patients' ECG signals; however, interpretation of these signals by the cardiologists and medical practitioners is usually a time-consuming task and prone to errors. Moreover, the complexity of the patterns associated with AF and the high level of noise affecting these collected signals have significantly limited the accuracy and reliability of the monitoring systems designed for AF detection [2], [3]. Therefore, it is desirable to develop algorithms for automatic detection of AF with high diagnostic accuracy and reliability.

Several algorithms have been introduced to automatically detect the presence of AF based on ECG signal characteristics. Most of these methods rely on accurate detection of P-waves and R-peaks. Thus, their performance significantly degrades when their underlying signal processing algorithm fails to detect the relevant peaks or waves of the ECG signal due

to the presence of noise in the signal. Although, there are some research works that eliminate the need for detection of P-wave and R-peak in their methodologies [4], [5], they still need to extract hand-crafted features that might not be totally representative features if the dataset changes in terms of size and the presence of other arrhythmias.

Deep learning (DL) can model high-level abstractions in data using deep neural networks in order to learn from multiple levels of abstractions. Over past years, DL-based methods have been used in ECG analysis and classification. However, the performance of these methods have not been quite significant compared to achieved performance with DL in other domains such as image processing. The main reason is the developed DL architectures were not suited enough for the addressed problems. Thus, developing new DL architectures that match specific medical problems and can capture the specific characteristics of ECG signals is still a challenge.

Motivated by the aforementioned limitations, we propose an end-to-end deep visual network for automatic detection of AF called ECGNET. The model is a two-channel deep neural network to more accurately detect AF presented in the ECG signal. The first channel takes in a preprocessed ECG signal and automatically learns where to attend for detection of AF. The second channel simultaneously takes in the preprocessed ECG signal to consider all features of the entire signal. This method gives more weights to the related parts of ECG signal with higher potential relevance to AF, and at the same time considers the whole cycle (i.e., the beat) to extract other consecutive dependencies between each wave (i.e., P-, QRS-, T-waves, etc.). Moreover, the proposed approach visualizes the parts of a given ECG signal that are more important to attend while trying to detect atrial fibrillation. It is also worth mentioning that, despite the majority of current AF detection techniques, our proposed method is capable of detecting AF in very short ECG recordings (e.g. duration around 5 s).

To the best of our knowledge, this paper is the first study that uses the whole information provided by the input signal and the visual attention at the same time for the purpose of AF detection. Recently, reference [6] reported an attention mechanism to detect AF, where they considered a deep recurrent neural network on 30-second ECG windowed inputs and took advantage of some time series covariates. The key contribution of our method is to develop an end-to-end two-channel deep network that automatically 1) extracts features from the focused parts of the signal with the capability of focusing on each part of a cycle (i.e., P-, QRS-, T-waves, etc.) instead of each windowed segment, and 2) at the same time, considers the abstracted features of the whole segment, just 5-s ECG segments. It is worth mentioning that our method does not rely on any hand-crafted features to the network as considered in [6]. We also visualize which regions of the signal

This material is based upon work supported by the National Science Foundation under Grant Number 1657260. Research reported in this publication was supported by the National Institute On Minority Health And Health Disparities of the National Institutes of Health under Award Number U54MD012388.

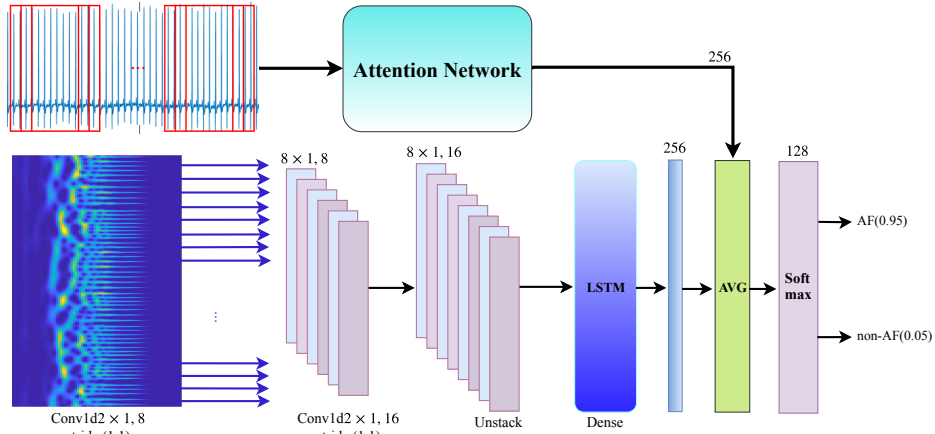


Figure 1: The network architecture for AF detection method. The top channel gets a sequence of split ECG signal (i.e., window = 128 and stride = 30) and the bottom channel gets the wavelet power spectrum of the sequence. Then, the average of two sections is computed and fed into a softmax layer. AVG: average.

are important while there is an underlying AF arrhythmia in the signal. Therefore, the proposed method can potentially assist the physicians in AF detection and can be also utilized to recognize complex pattern in the signals related to other arrhythmias that cannot be easily seen in the signals.

## II. DATASET AND DATA PREPARATION

The proposed method has been evaluated using the PhysioNet MIT-BIH Atrial Fibrillation Database (AFDB) [7]. The AF database is comprised of 25 long-term ECG recordings of human subjects with mostly atrial fibrillation. It includes two 10-hours long ECG recordings for each individual. Here, each ECG signal of AFDB is divided into 5-s segments and each segment is labeled based on a threshold parameter,  $p$ . When the percentage of annotated AF beats of the 5-s segment is greater than or equal to  $p$ , we considered it as AF, otherwise non-AF arrhythmia. Similar to previous reported studies in [4], [8], we selected  $p = 50\%$ . It is worth noticing that that we do not apply any noise removal approaches to the ECG signals.

## III. PROPOSED APPROACH

An overview of the proposed model is depicted in Fig. 1. The model architecture is a two-channel deep neural network. The top channel takes the row windowed signal as input and includes an attention strategy to emphasis on important visual task-relevant features of the given signal. This section of the architecture is called Attention Network. We divided the given ECG signal into several windows with fixed lengths of 128 and an overlap of 25%. The bottom channel considers a deep recurrent convolutional network that takes wavelet power spectrum of the windowed ECG signal. The output of the network is a vector of decimal probabilities regarding the classes. A more detailed explanation of each section of the network is provided below.

### A. Attention Network

In general, there are two types of attention models: i) soft attention and ii) hard attention models. The soft attention models are end-to-end approaches and differentiable deterministic mechanisms that can be learned by gradient based methods. However, the hard attention models are stochastic processes and not differentiable. Thus, they can be trained by using the REINFORCE algorithm in the reinforcement learning framework. In this paper, a soft attention mechanism is used

because the back propagation seems to be more effective [9]. The attention network includes three main parts as follow:

**1) Convolutional neural network (CNN):** The CNN consists of two consecutive one-dimensional convolutional layers followed by Rectified Linear Unit (ReLU) non-linear layers. They have 32 and 64 filters of  $2 \times 1$  with strides 1 for each one. Figure 2 depicts the detailed architecture. Sequences of windowed ECG signals are fed into the CNN for feature extraction. At each time-step  $t$ , a windowed frame is fed into the network and the last convolutional layer of the 1-Dimensional CNN part outputs  $D$  feature maps of size  $K \times 1$  (e.g, we concluded 64 feature maps  $8 \times 1$ ). Then, the feature maps are converted to  $K$  vectors in which each vector has  $D$  dimensions as follows:  $F_t = [F_{t,1}, F_{t,2}, \dots, F_{t,K}]$ ,  $F_{t,i} \in \mathbb{R}^D$ .

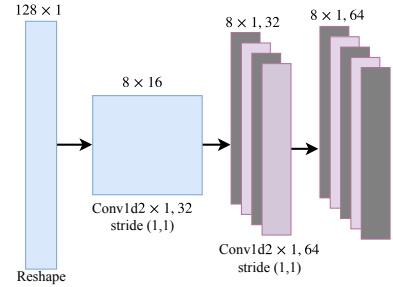


Figure 2: A diagram of convolutional layers used in the proposed model. The CNN part of the model takes the windowed ECG signal as input (i.e., a sequence of frames) and computes vertical feature slices,  $F_t$  with dimension  $D$ .

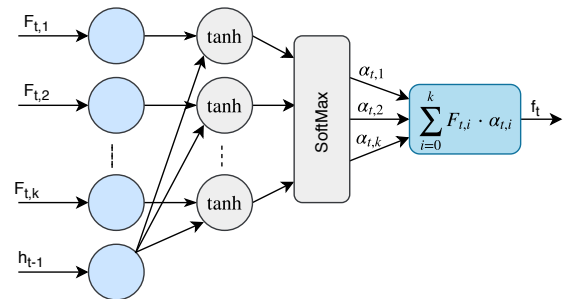


Figure 3: The structure of the attention mechanism used in the proposed model. At each time step  $t$ , the attention module utilizes  $F_t$  and the previous hidden state of the RNN part,  $h_{t-1}$  to calculate an expected value,  $f_t$  with respect to vertical feature slices,  $F_{t,i}$  and the importance of each region of input window frame,  $\alpha_{t,i}$ .

**2) Attention layer (i.e., a soft attention mechanism):** The extracted features of the CNN part are sequentially passed to the attention layer to compute the probabilities corresponding to the importance of each part of the windowed frame (e.g., P-, QRS- and T-waves, etc.). In other words, the input window is divided into  $K$  regions and the attention mechanism attempts to attend to the most relevant regions which are related to AF. Figure 3 shows the structure of the attention mechanism. The attention layer gets two separate inputs: 1)  $K$  vectors,  $F_{t,1}, F_{t,2}, \dots, F_{t,K}$ , where each  $F_{t,i}$  is a representation of different regions of the input window frame, and 2) A hidden state  $h_{t-1}$ , which is the internal state of the LSTM at the previous time step. Then, it computes a vector,  $f_t$  which is a linear weighted combination of the values of  $F_{t,i}$ . Therefore, the attention mechanism can be formulated as follows:

$$c_{att}(F_{t,i}, h_{t-1}) = \tanh(W_h h_{t-1} + W_f F_{t,i}), \quad (1)$$

$$\alpha_{t,i} = \frac{\exp(c_{att}(F_{t,i}, h_{t-1}))}{\sum_{j=1}^k \exp(c_{att}(F_{t,j}, h_{t-1}))} \quad i \in 1, 2, \dots, k \quad (2)$$

$$f_t = \sum_{i=1}^k \alpha_{t,i} F_{t,i}, \quad (3)$$

where  $\alpha_{t,i}$  is the importance of the region  $i$  of input window frame. At each time step  $t$ , the attention module calculates  $c_{att}$ , a composition of the values of  $F_{t,i}$  and  $h_{t-1}$  followed by a tanh layer. Then, it is passed to a softmax layer to compute  $\alpha_{t,i}$  over  $k$  regions. Indeed, each  $\alpha_{t,i}$  is considered as the amount of importance of the corresponding vector  $F_{t,i}$  amongst  $K$  vectors in the input window. Finally, the attention layer computes  $f_t$ , a weighted sum of all the vectors  $F_{t,i}$  based on calculated  $\alpha_{t,i}$ 's. Thus, the network can learn to put more emphasize on the important parts (e.g., P-, QRS- and T-waves, etc.) of the input window frame with higher probabilities of presence of AF in the input ECG.

**3) Recurrent neural network (i.e., Long Short-Term Memory (LSTM) units):** The attention layer is followed by LSTM units (which are a stack of two LSTM layers with the LSTM sizes of 64) for long-term learning to capture temporal dependencies between windows of each input signal. The RNN part of the network utilizes the previous hidden state  $h_{t-1}$  and the output of the attention module  $f_t$ , to calculate the next hidden state  $h_t$ . The parameter  $h_t$  is used as the input of the attention module in order to calculate the value of  $f_{t+1}$  at the next time-step. In addition, it is utilized as the input of a fully-connected linear layer with 256 neurons.

#### B. Deep Recurrent Convolutional Neural Network (RCNN):

The first layer consists of 8 1-D convolution filters of size  $2 \times 1$  with a stride 1 followed by a Rectified Linear Unit (ReLU) non-linearity. The second layer is comprised of 16 1-D convolution filters of size  $2 \times 1$  with stride 1, again followed by a rectifier non-linearity. The third layer is a RNN layer with the LSTM units of size 256 followed by a fully connected layer with 256 hidden units. Here, the spectrogram size is  $90 \times 300 \times 3$ . It can be considered as a sequence of column vectors (300 vectors) that each consists of 270 values. For the purpose of feature extraction, we feed these sequences to the first 1-D convolutional layers of the deep RCNN.

Similar to other deep learning-based AF detectors [8], the deep neural network part of our model takes a 2-D representation with wavelet power spectrum of the ECG segment. In [8], a 2-D convolution that operators on the entire input is utilized, while our method applies 1-D convolution operators

to the each frequency vector (i.e., at each time step) of the given the spectrograms obtained from each segment, and feeds the output of the 1-D convolutional layers to long short-term memory units to capture dependencies between each frequency vector. Therefore, our proposed architecture can capture the temporal potential patterns that may present in an AF arrhythmia. In other words, a CNN with two-dimensional filters shares weights of the  $x$  and  $y$  dimensions and considers the extracted features have the same meaning apart from their locations. However, in spectrograms, the two dimensions shows the strength of frequency and the time, and are completely diffident. In a 2-D convolution operator, frequency shifts of a signal (in a spectrogram representation) can change its spatial extent. Hence, the ability of 2-D CNNs to learn the spatial invariant features might not be well for the spectrograms [10]. This is the main reason, we included 1-D CNNs followed by LSTM units instead of 2-D CNNs. Moreover, using 1-D CNNs in the network would bring lower number of parameters and as a result further complexity reduction.

Finally, the outputs of the attention and RCNN sections are averaged and fed into a softmax layer. Then, the softmax assigns decimal probabilities to each class of interest (i.e., AF and non-AF).

## IV. EXPERIMENTAL RESULTS

We evaluated the performance of our proposed method using the MIT-BIH AFIB database. There are a total of 162,536 5-s segments, where the number of AF segments is 61,924, and the number of non-AF segments is 100,612. We randomly selected the same number of segments for each class (i.e., AF and non-AF), as 20,000; totally 40,000 samples, to remove the effect of imbalanced data samples on training the model. Then, 70% of data were used to train the network, 10% were used to validate the model, and the remaining 20% were used to test the model.

Table I presents a performance comparison between the proposed method and several state-of-the-art works using the MIT-BIH AFIB database, where the 5-s data segments were considered. As it is clear in Table I, overall, our proposed AF detector shows better results in terms of the sensitivity, specificity and accuracy evaluation metrics compared to all methods presented in the table.

A visualization example of attended parts of an ECG signal with an AF is illustrated in Figure 4. The white regions, showed with circles, indicate where the model learned to look while the patient had the AF. We should note that the two main indicators of AF in ECG signals as considered in the majority of the previous works are: 1) the absence of P-waves that can be replaced by a series of low-amplitude oscillations called fibrillatory waves, and 2) the irregularly irregular rhythm (i.e., irregularity of R-R intervals) [12]–[14]. It is worth noting that

Table I: Comparison of performance of the proposed model against other algorithms on the MIT-BIH AFIB database with the ECG segment of size 5-s ( $\leq 7$  Beats).

| Method                          | Best Performance (%) |              |              |
|---------------------------------|----------------------|--------------|--------------|
|                                 | Sensitivity          | Specificity  | Accuracy     |
| ECGNET                          | <b>99.53</b>         | <b>99.26</b> | <b>99.40</b> |
| Xia, et al. (2018) [8]          | 98.79                | 97.87        | 98.63        |
| Asgari, et al. (2015) [4]       | 97.00                | 97.10        | —            |
| Lee, et al. (2013) [11]         | 98.20                | 97.70        | —            |
| Jiang, et al. (2012) [12]       | 98.20                | 97.50        | —            |
| Huang, et al. (2011) [13]       | 96.10                | 98.10        | —            |
| Babaeizadeh, et al. (2009) [14] | 92.00                | 95.50        | —            |
| Dash, et al. (2009) [15]        | 94.40                | 95.10        | —            |
| Tateno, et al. (2001) [16]      | 94.40                | 97.20        | —            |

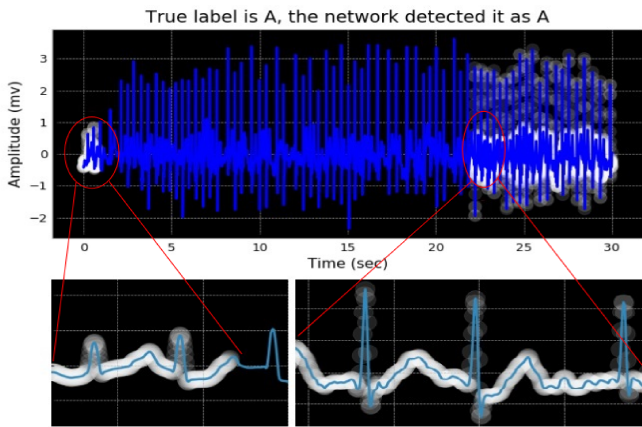


Figure 4: Visualization of the attention network's result on an ECG sample with AF arrhythmia. The white circles depicts the most important regions of the ECG signal to attend. More brightness means more attention.

the attention network focused on the regions where there is no P-waves or fibrillatory waves in the ECG signal. Furthermore, there are some attentions on the R-peaks that may show the irregularity of R-R intervals. As the R-R intervals of the signal were not computed here, we do not have a reference point to show if the focused R-peaks are because of the irregularity. We encourage interested readers to view the clip video through the link below <sup>1</sup> that visualizes the attended locations by proposed model as well. As shown in Tables I, our attention-based DL model with the proposed combination architecture of the attention network and the deep recurrent neural network (i.e., deep net part of the entire model) has superior performance in detection of AF compared to the existing techniques.

## V. DISCUSSION AND CONCLUSION

Several algorithms have been reported in the literature to detect AF ranging from RR interval variability, P-wave detection based to machine learning paradigms including deep learning-based methods. In this paper, we propose an attention-based AF detection method that automatically identifies the potential AF regions in the ECG signal. Despite the majority of the previously reported works in this area [11]–[14], the performance of the proposed method do not rely on the accuracy of the hand-crafted algorithms for detecting P-waves and R-R intervals. In contrast, the attention network in our proposed model automatically focuses on the most relevant regions of each heartbeat which are prone to be a part of an AF, and puts more weights on those regions in the network to distinguish between the AF and non-AF classes. The performance of our method is compared against the majority of existing algorithms for detection of AF using the same databases and same evaluation metrics. The proposed method achieved an accuracy of 99.40, a sensitivity of 99.53 and a specificity of 99.26 validated on the MIT-BIH AFIB database with 5-s data segments that significantly outperforms the results of other studies. One of the other key issues in AF detection methods is their poor performance in detecting AF episodes in short signal recordings (i.e., less than 30-s). While the majority of the state-of-the-art algorithms require a 30-s episode or at least 127/128 beats to achieve an acceptable detection performance [13], [17], our proposed method offers a great performance on

very short ECG segments of size 5-s which here are less than a 7-beats window. Furthermore, unlike the works such as [17], we do not need tuning any parameters which might affect on the detection performance.

In this study, we proposed a novel deep network architecture to classify the given signal as AF or non-AF. The proposed AF detector shows the better performance compared to the other detectors in the literature. One key aspect of our AF detector is that it simultaneously gives more weights to the related parts of the ECG signal with higher potential prevalence of AF, and also considers the whole cycle (i.e., the beat) to extract other consecutive dependencies between each wave (i.e., P-, QRS-, T-waves, etc.). Moreover, The proposed method obtains significant detection results by using a short ECG segment, 5-s long, to detect AF without the need of tuning any parameters in the model.

## REFERENCES

- [1] Véronique L. Roger et al., “Heart disease and stroke statistics—2012 update,” *Circulation*, vol. 125, no. 1, pp. e2–e220, 2012.
- [2] Fatemeh Afghah, Abolfazl Razi, SM Reza Soroushmehr, Somayeh Molaie, Hamid Ghanbari, and Kayvan Najarian, “A game theoretic predictive modeling approach to reduction of false alarm,” in *International Conference on Smart Health*. Springer, 2015, pp. 118–130.
- [3] M. Zaeri-Amirani, F. Afghah, and S. Mousavi, “A feature selection method based on shapley value to false alarm reduction in icus a genetic-algorithm approach,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2018, pp. 319–323.
- [4] Shadnaz Asgari, Alireza Mehrnia, and Maryam Moussavi, “Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine,” *Computers in biology and medicine*, vol. 60, pp. 132–142, 2015.
- [5] Philip de Chazal and Richard B Reilly, “A patient-adapting heartbeat classifier using eeg morphology and heartbeat interval features,” *IEEE transactions on biomedical engineering*, vol. 53, no. 12, pp. 2535–2543, 2006.
- [6] Supreeth P Shashikumar, Amit J Shah, Gari D Clifford, and Shamim Nemati, “Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks,” *arXiv preprint arXiv:1805.09133*, 2018.
- [7] PhysioNet, “MIT-BIH AF Database (AFDB),” 2000.
- [8] Yong Xia, Naren Wulan, Kuanquan Wang, and Henggui Zhang, “Detecting atrial fibrillation by deep convolutional neural networks,” *Computers in biology and medicine*, vol. 93, pp. 84–92, 2018.
- [9] Sajad Mousavi, Michael Schukat, Enda Howley, Ali Borji, and Nasser Mozayani, “Learning to predict where to look in interactive environments using deep recurrent q-learning,” *arXiv preprint arXiv:1612.05753*, 2016.
- [10] Lonca Wyse, “Audio spectrogram representations for processing with convolutional neural networks,” *arXiv preprint arXiv:1706.09559*, 2017.
- [11] Jinseok Lee, Bersain A Reyes, David D McManus, Oscar Maitas, and Ki H Chon, “Atrial fibrillation detection using an iphone 4s,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 203–206, 2013.
- [12] Kai Jiang, Chao Huang, Shu-ming Ye, and Hang Chen, “High accuracy in automatic detection of atrial fibrillation for holter monitoring,” *Journal of Zhejiang University SCIENCE B*, vol. 13, no. 9, pp. 751–756, 2012.
- [13] Chao Huang, Shuming Ye, Hang Chen, Dingli Li, Fangtian He, and Yüwen Tu, “A novel method for detection of the transition between atrial fibrillation and sinus rhythm,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 4, pp. 1113–1119, 2011.
- [14] Saeed Babaeizadeh, Richard E Gregg, Eric D Helfenbein, James M Lindauer, and Sophia H Zhou, “Improvements in atrial fibrillation detection for real-time monitoring,” *Journal of electrocardiology*, vol. 42, no. 6, pp. 522–526, 2009.
- [15] S Dash, KH Chon, S Lu, and EA Raeder, “Automatic real time detection of atrial fibrillation,” *Annals of biomedical engineering*, vol. 37, no. 9, pp. 1701–1709, 2009.
- [16] K Tateno and L Glass, “Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of rr and  $\delta rr$  intervals,” *Medical and Biological Engineering and Computing*, vol. 39, no. 6, pp. 664–671, 2001.
- [17] Xingran Cui, Emily Chang, Wen-Hung Yang, Bernard C Jiang, Albert C Yang, and Chung-Kang Peng, “Automated detection of paroxysmal atrial fibrillation using an information-based similarity approach,” *Entropy*, vol. 19, no. 12, pp. 677, 2017.

<sup>1</sup>[https://drive.google.com/file/d/1VMhoVPzUVsXjP3JSge9\\_PJGAUImZTinN/view?usp=sharing](https://drive.google.com/file/d/1VMhoVPzUVsXjP3JSge9_PJGAUImZTinN/view?usp=sharing)