Computers in Biology and Medicine 127 (2020) 104057

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

HAN-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks☆

Sajad Mousavi [a,*], Fatemeh Afghah [a], U. Rajendra Acharya [b,c,d]

[a] School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, 86011, USA
[b] School of Engineering, Ngee Ann Polytechnic, Singapore
[c] School of Science and Technology, Singapore University of Social Sciences, 463 Clementi Road, 599494, Singapore
[d] Department Bioinformatics and Medical Engineering, Asia University, Taiwan

ARTICLE INFO

Keywords:
Atrial fibrillation detection
Heart arrhythmia
Interpretability
Attention mechanism
Bidirectional recurrent neural networks

ABSTRACT

Atrial fibrillation (AF) is one of the most prevalent cardiac arrhythmias that affects the lives of many people around the world and is associated with a five-fold increased risk of stroke and mortality. Like other problems in the healthcare domain, artificial intelligence (AI)-based models have been used to detect AF from patients' ECG signals. The cardiologist level performance in detecting this arrhythmia is often achieved by deep learning-based methods, however, they suffer from the lack of interpretability. In other words, these approaches are unable to explain the reasons behind their decisions. The lack of interpretability is a common challenge toward a wide application of machine learning (ML)-based approaches in the healthcare which limits the trust of clinicians in such methods. To address this challenge, we propose *HAN-ECG*, an interpretable bidirectional-recurrent-neural-network-based approach for the AF detection task. The *HAN-ECG* employs three attention mechanism levels to provide a multi-resolution analysis of the patterns in ECG leading to AF. The detected patterns by this hierarchical attention model facilitate the interpretation of the neural network decision process in identifying the patterns in the signal which contributed the most to the final detection. Experimental results on two AF databases demonstrate that our proposed model performs better than the existing algorithms. Visualization of these attention layers illustrates that our proposed model decides upon the important waves and heartbeats which are clinically meaningful in the detection task (e.g., absence of P-waves, and irregular R-R intervals for the AF detection task).

1. Introduction

Atrial fibrillation (AF) is a common cardiac arrhythmia that can lead to various heart-related complications such as stroke, heart failure, and atrial thrombosis [6]. Electrocardiography is a test that measures the electrical activity of the heart over a specific period. The test output is an electrocardiogram (ECG) signal that is a plot of voltage against time. A common non-invasive diagnosis way for the AF detection is the process of the recorded electrocardiogram (ECG) signal visually by a cardiologist or medical practitioner. However, this is a time-consuming process and subject to human error.

Therefore, dozens of computer-aided methods have been developed for automatic detection of atrial fibrillation and other heart arrhythmias. The existing ML-based methods include handcrafted feature-based and automatic-extracted feature-based approaches in their solutions [1,2,8,13,14,21,31,34,34,35]. Among them, the methods that extract features automatically have gained more attention because they could learn the ECG signal representations efficiently and achieve state-of-the-art results.

Deep learning models with the capability of the automatic feature extracting provide significant performance in the AF detection task noting their ability to detect complex patterns in the ECG signals [26,33]. Nevertheless, they work as black boxes that make it hard to understand the reasons behind their decisions. Interpretability and transparency are key required factors in AI-based decision making in healthcare to enable and encourage physicians who are held

☆ This study 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.

* Corresponding author.
E-mail addresses: sm3276@nau.edu, SajadMousavi@nau.edu (S. Mousavi), Fatemeh.Afghah@nau.edu (F. Afghah), aru@np.edu.sg (U.R. Acharya).

https://doi.org/10.1016/j.compbiomed.2020.104057
Received 7 June 2020; Received in revised form 10 October 2020; Accepted 11 October 2020
Available online 15 October 2020

accountable for medical decisions to trust the recommendations of these algorithms. One way to make deep learning models interpretable is to incorporate an attention mechanism in the model that learns the relationship between the input data samples and the given task [16].

In this study, to provide an interpretable method with high performance for automatic detection of atrial fibrillation, we propose a deep learning model powered by hierarchical attention networks. The proposed method is composed of three parts in which each part contains a stacked bidirectional recurrent neural networks (BiRNN) followed by an attention model. The first part learns a wave level representation of the ECG signal, the second part learns a heartbeat level representation of the ECG signal and the third part learns a window-based (i.e., contains multiple heartbeats) level representation of the ECG signal. All learned representations at each level are interpretable and are able to show which parts of the input signal are the reasons to trigger an AF event.

The hierarchical attention model was first proposed in Ref. [32] in the content of document classification task, as a novel hierarchical attention architecture that matches the hierarchical nature of a document, meaning words make sentences and sentences make the document. Since in the ECG analysis application, we deal with a similar notion of hierarchical input where the ECG signal includes multiple levels of resolution (waves, beats, and windows), the proposed hierarchical attention model can mirror the physicians' decision-making process. For instance, in order to detect AF, they, first, look for some important windows (a sequence of continuous heartbeats), next, they look at the important heartbeats of the windows and then focus on the heartbeat waves.

The main contributions of this study are summarized as follows:

- We propose an end-to-end hierarchical attention model that achieves the state-of-the-art performance with the capability of the interpretability.
- The proposed model provides multi-level resolution interpretability (i.e., window by window (multiple heartbeats), heartbeat by heartbeat, and wave by wave levels).
- We empirically demonstrate that the important parts of the ECG signal for the model in triggering the AF are clinically meaningful.
- The proposed approach can be used to recognize new potential patterns leading to trigger heart arrhythmias.

The rest of this paper is organized as follows. Section 2 gives a review of the related work. Section 3 provides a detailed description of the proposed approach. Section 4 presents the experimental setup, the used databases to evaluate the proposed model, and compares the performance of the proposed model to the existing methods following by the interpretability analysis. Section 5 discuss the results and describes the limitations of the proposed method. Finally, Section 6 concludes the paper.

2. Related work

Heart arrhythmia classification and prediction tasks are very important research problems in machine learning for the healthcare area. Recent advances of deep learning algorithms have impacted on achieving great performance in the machine learning-oriented healthcare problems. Deep convolutional neural networks have been used to improve the performance of ECG heartbeat classification task [1,11,12,33]. Recurrent neural networks (RNNs) and sequence to sequence models were employed to perform automatic heartbeat annotations [7,17,27]. Deep learning models have also been utilized to detect false arrhythmia alarms. In the paper by Lehman et al. [15], authors applied a supervised denoising autoencoder (SDAE) to ECG signals to classify ventricular tachycardia alarms. In the paper by Mousavi et al. [19], authors used an attention-based convolutional and recurrent neural networks to suppress false arrhythmia alarms in the ICU.

Atrial fibrillation (AF) is one of the most common types of arrhythmias in patients with heart diseases and challenging arrhythmias to detect. The research papers [5,8,31] aimed to use deep convolutional neural networks for the atrial fibrillation arrhythmia detection task and achieved good arrhythmia detection performance. In the paper by Shashikumar et al. [29], authors applied an attention mechanism to detect the atrial fibrillation arrhythmia. Authors employed a deep recurrent neural network on 30-s ECG windows' inputs, and also fed some time series covariates to the network. These covariates are hand-crafted features and include the standard deviation and sample entropy of the beat-to-beat interval time series. Although they have used an attention mechanism in the architecture of their model, their proposed method was not an interpretable detective model. The single-level attention applied to fixed-length 30s ECG windows, which contains several heartbeats only improves the detective performance. Our previous work named ECGNET [18] is an interpretable atrial fibrillation detective model, which uses a deep visual attention mechanism to automatically extract features and focus on different parts of the heartbeats of the input ECG signal. The ECGNET has suggested an interpretable AF detection with a single-level attention using the wavelet power spectrum as input, however, this study proposes a hierarchical attention network having raw ECG signals as input.

Unlike the aforementioned AF detective models, the proposed model provides a high resolution interpretable predictive model (i.e., window (i.e., multiple heartbeats) by window, beat by beat, and wave by wave levels) using the hierarchical bidirectional recurrent neural networks and attention networks. The proposed model improved the detection performance and explained the reasons behind model decisions simultaneously.

3. Methodology

In this section, we first describe the pre-processing steps needed to prepare the data to be fed into the proposed model. Then, we explain the main components of the proposed method in detail.

3.1. Pre-processing

The input of the proposed method is a sequence of ECG heartbeats in which each heartbeat contains a sequence of building waves (P-wave, QRS complex, T-wave, etc.). To prepare this structure of ECG signals, we perform a few pre-processing steps on them as follows:

1. Removing the baseline wander and power-line interference noises in the ECG signal. To this end, the ECG signal was passed through a band-pass Butterworth filter with a filter order of 10 and passband frequency ranges of 0.5–50 Hz.
2. Transforming the given ECG signal to have a zero mean and a unit standard deviation (i.e., standardization).
3. Detecting the R-peaks of given ECG signal or the QRS complexes using an algorithm that considers the consensus of multiple algorithms including the Pan-Tompkins algorithm [20] and gqrs package provided by Ref. [22].
4. Dividing the continuous ECG signal into a sequence of heartbeats and split the heartbeats into distinct units named waves. The waves in the ECG signal are extracted with respect to the detected R-peaks and using adaptive searching windows (with fixed length windows). Indeed, the output of the search algorithm is a one-dimensional vector in which each element corresponds to start/end locations of the waves (i.e., P, QRS, and T-waves). Also, a heartbeat is defined from the onset of the current P-wave to the offset of consecutive T-wave. Fig. 1 depicts a segmented ECG signal annotated with the R-peaks, P, QRS and T-waves.

After doing the above pre-processing steps, each ECG signal becomes a sequence of B heartbeats in which each heartbeat, $Beat_i$ contains T_i

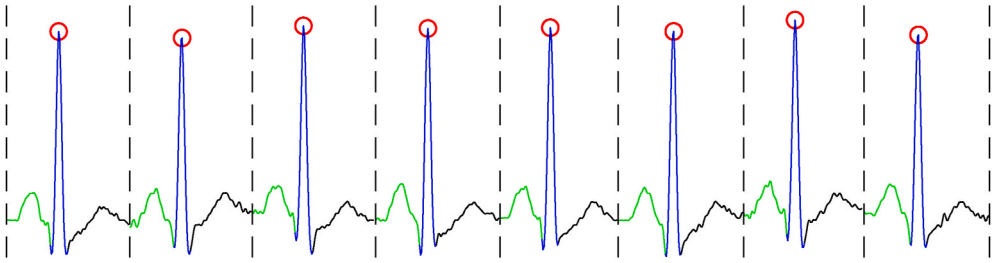


Fig. 1. Illustration of an ECG signal; The red circles indicate R peaks; green, blue and black curves illustrate P, QRS and T-waves, respectively.

waves, where $wave_{it}$ represents the t^{th} ($t \in [1, T_i]$ wave in the i^{th} ($i \in [1,B]$) heartbeat, $Beat_i$. It is worth mentioning that we have resized all waves to have fixed-length vectors (herein, 75 samples). For example, for a given ECG signal, i, we have an array with a size of (B_i, T_i, 75) where B_i is the number of heartbeats, T_i is the number of waves, and 75 is the number of samples for each wave.

3.2. Model

The goal of the proposed method is to detect atrial fibrillation arrhythmia in an explainable way. Fig. 2 presents the network architecture of the proposed method. A sequence of waves of an ECG signal is fed into a stacked bidirectional recurrent neural networks (BiRNN) followed by an attention model. The stacked BiRNNs are used to extract a vector representation for each input wave and the attention model is used to focus on those waves that are the best representatives of a heartbeat. Next, the vector representations of the waves are integrated to represent a heartbeat vector. Then, the heartbeat vectors of the previous step are introduced to other stacked BiRNNs followed by another attention model. Similarly, the attention model puts more emphasis on important heartbeats and produces heartbeat context vectors. After that, a sequence of windows in which each window contains multiple heartbeat context vectors is computed and the same procedure is applied to the windows, and a summarized vector that includes all information of the ECG signal is extracted. Finally, the summarized vector can be used for the atrial fibrillation detection task. Overall, the model architecture is composed of three main parts: a wave encoder along with a wave attention, a beat encoder along with beat attention, and a window encoder along with window attention. In the following sections, we explain each part of the proposed model in detail.

3.2.1. Bidirectional recurrent neural networks

Bidirectional recurrent neural network (BiRNN) is more efficient than the RNN while the length of the sequence is very large [28]. The reason is that standard RNNs are unidirectional so they are restricted to only use the previous input state. However, the BiRNN can process data in both forward and backward directions. Therefore, the current state has access to previous and next input information simultaneously. To improve the detection performance, the BiRNNs were employed in the model for encoding the sequences of wave and beat vectors.

The BiRNN consists of forward and backward networks. The forward network takes in a sequence of waves/beats in a regular order, from $t = 1$ to $t = T_i$, as input and computes forward hidden state, $\overrightarrow{h_t}$ and the backward network takes in wave/beat sequence in a reverse order, from $t = T_i$ to $t = 1$, as input and calculates backward hidden state, $\overleftarrow{h_t}$. Then, the output of the BiRNN is considered as a weighted sum over the concatenation of the forward hidden state, $\overrightarrow{h_t}$, and the backward one, $\overleftarrow{h_t}$. The BiRNN can be defined mathematically as follows:

$$\overrightarrow{h_t} = \tanh\left(\overrightarrow{W}x_t + \overrightarrow{V}\overrightarrow{h}_{t-1} + \overrightarrow{b}\right) \tag{1}$$

$$\overleftarrow{h_t} = \tanh\left(\overleftarrow{W}x_t + \overleftarrow{V}\overleftarrow{h}_{t+1} + \overleftarrow{b}\right) \tag{2}$$

$$y_t = \left(U\left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] + b_y\right), \tag{3}$$

where ($\overrightarrow{h_t}$, $\overrightarrow{b}$) are the hidden state and the bias of the froward network, and ($\overleftarrow{h_t}$, $\overleftarrow{b}$) are the hidden state and the bias of the backward one. Besides, x_t and y_t are the input and the output of the BiRNN, respectively.

3.2.2. Wave encoder and wave attention

A sequence of waves, $wave_{it}$ $t \in [1, T_i]$, for i^{th} heartbeat, is fed into a bidirectional recurrent neural network (BiRNN) to encode the wave sequence. The forward network of the BiRNN gets the heartbeat, i in a normal time order of waves from $wave_{i1}$ to $wave_{iT}$ and the backward network gets the heartbeat, i in a reverse time order of waves from $wave_{iT}$ to $wave_{i1}$. Then, the BiRNN outputs, h_{it} representing a low dimensional latent vector representation of the heartbeat, i.

Similar to the words of a sentence in which necessarily all words are not important to give the meaning of the sentence [32], herein all waves of a heartbeat do not have the same weights in representing the heartbeat. Therefore, an attention mechanism can extract the relevant waves of the heartbeat that contribute more to the meaning of the heartbeat. The attention mechanism is a shallow neural network that takes the BiRNN output, h_{it} as input and computes a probability vector, α_{it} corresponding to the importance of each wave vector. Then, it calculates a wave context vector, b_i which is a weighted sum over h_{it} with the weight vector α_{it} (as shown in Fig. 3). Indeed,

$$\alpha_{it} = softmax(V_w \tanh(W_w h_{it} + b_w)) \tag{4}$$

$$b_i = \sum_t \alpha_{it} h_{it}, \tag{5}$$

where (W_w, b_w, V_w) are the parameters to be learned and $softmax(.)$ is a function that squeezes its input, which is a vector of real numbers, in values between 0 and 1.

3.2.3. Beat encoder and beat attention

Similar to the wave encoder part, the BiRNN of the beat encoder part takes a sequence of wave context vectors, b_i ($i \in [1, B]$) as input and produces vectors, h_i ($i \in [1, B]$) which are latent representations of the

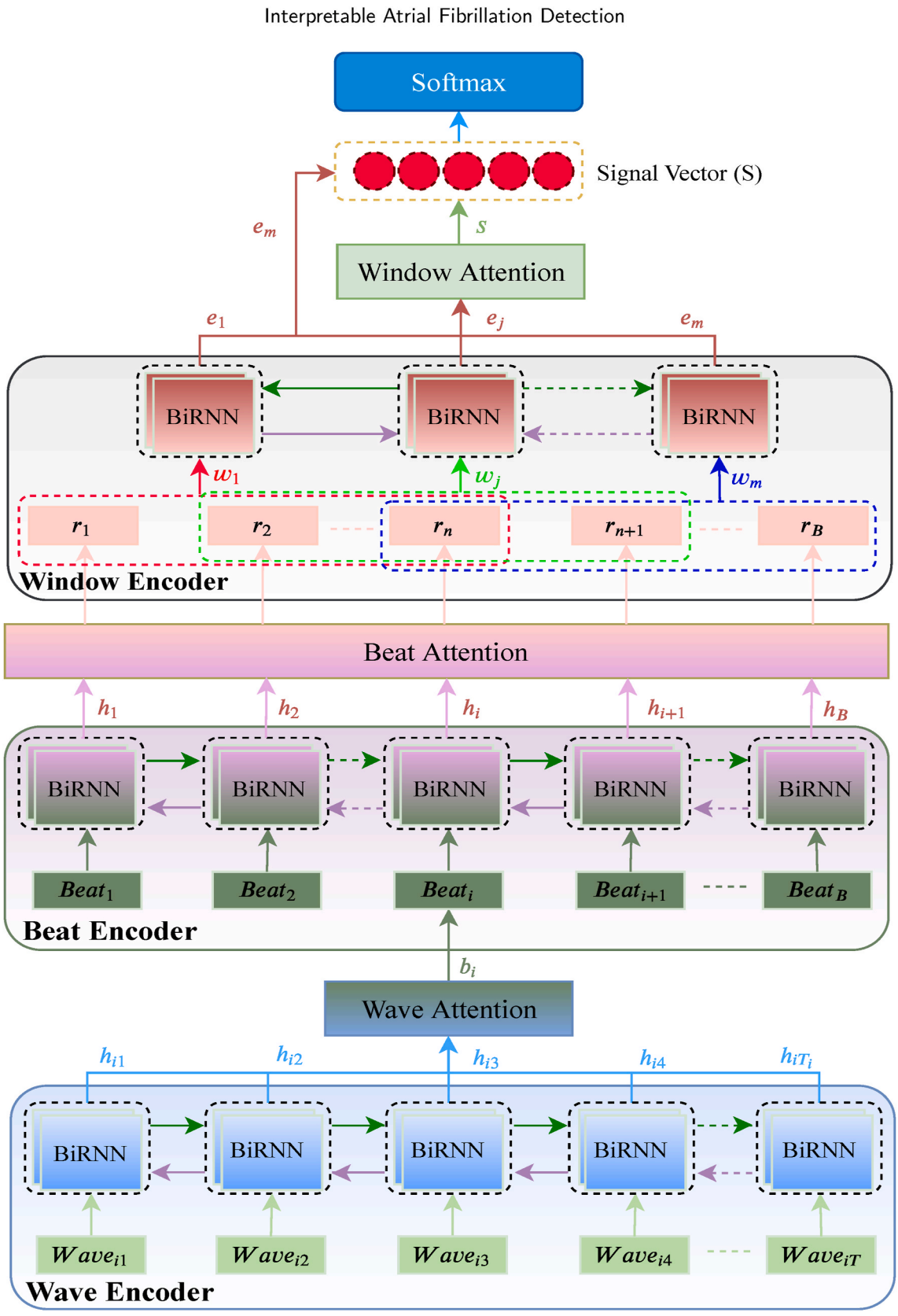


Fig. 2. Architecture of hierarchical attention network (HAN) for Atrial Fibrillation Detection.

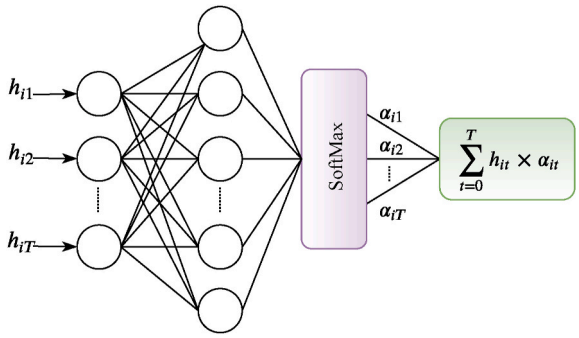


Fig. 3. Schematic diagram of the attention mechanism.

input heartbeats. To emphasize the more important heartbeats in triggering the arrhythmia, another attention mechanism is used on the heartbeat level. Therefore,

$$\alpha_i = softmax(V_b \tanh(W_b h_i + b_b)) \tag{6}$$

$$r = \alpha_i \circ h_i, \tag{7}$$

where (W_b, b_b, V_b) are the parameters to be learned, α_i is the attention weight vector of the heartbeats, and r $(r_1, r_2, \ldots, r_B)$ is the heartbeat context vectors which is an element-wise product of the hidden states, h_i and the importance of each heartbeat, α_i.

3.2.4. Window encoder and window attention

In addition to the wave and heartbeat level encoding modules, we also consider a window level encoding module in which a window contains multiple heartbeats. The heartbeat context vectors, $(r_1, r_2, \ldots, r_B)$ are converted to a sequence of windows, w_j $(j = 1, 2, \ldots, m)$ by sliding a predefined-fixed-length window with a predefined-fixed hop size in the heartbeats over the heartbeat context vectors (as shown in Fig. 2). For example, if we consider a window with $n = 3$ heartbeats and the hop size be 1, the extracted sequence of windows becomes $(w_1 = [r_1; r_2; r_3],\ w_2 = [r_2; r_3; r_4],\ \ldots,\ w_m = [r_{B-2}; r_{B-1}; r_B])$ where ; is a simple concatenation operation.

Analogous to the previous steps, a BiRNN is used to encode the windows, w_j $(j = 1, 2, \ldots, m)$ and again an attention mechanism is employed to measure the importance of the windows. Specifically,

$$\gamma_j = softmax\left(V_v \tanh\left(W_v e_j + b_v\right)\right) \tag{8}$$

$$s = \sum_j \gamma_j e_j, \tag{9}$$

where (W_v, b_v, V_v) are the parameters to be learned, γ_j is the attention weight vector of the windows, e_j $(j = 1, 2, \ldots, m)$ are the hidden states of the BiRNN, and s is the window context vector that encompasses the whole information of the windows, containing multiple heartbeats, of the input ECG signal.

3.2.5. Detection

We concatenate the window context vector, s, and the last hidden state, e_m, to obtain combined information of both vectors and then feed it into a shallow network followed by a softmax layer to produce a probability vector, p in which each element determines the probability of the input signal belonging to each class of interest (AF or non-AF). Specifically,

$$S = \tanh(W_c[s; e_m]) \tag{10}$$

$$p = softmax(WsS + b_s), \tag{11}$$

where (W_c, Ws, b_s) are the parameters to be learned.

Finally, we use a cross-entropy loss to calculate the training loss as follows:

$$L = -y \cdot \log p, \tag{12}$$

where $(\cdot)$ is the vector dot product operator and y is the ground truth vector.

3.2.6. Interpretation

Typically artificial intelligence (AI)-based models that both give good performance and are interpretable, are preferable to apply to real medical practice. Therefore, having machine learning models that explain the reasons behind their decisions are very important in medical applications. The proposed method has three levels of the attention mechanism, the first level (i.e., the wave level) produces the wave weights, α_{it} $(t = 1, 2, \ldots, T_i)$ representing the importance of the waves in a heartbeat, the second level (i.e., the heartbeat level) computes the heartbeat weights, α_i $(i = 1, 2, \ldots, B)$ showing the amount of the influence of each heartbeat on the occurrence of an arrhythmia, and third level (i.e., the window level) produces the window weights, γ_j $(j = 1, 2, \ldots, m)$ demonstrating the importance of the combinations of the heartbeats. In Section 4.3, we provide visualized examples of some ECG signals with the AF and non-AF arrhythmias where the focused portions of the signals determined by the proposed attention mechanism are highlighted.

4. Experiments

In this section, we describe the two atrial fibrillation datasets used for the quantitative and qualitative analyses of the proposed method. Then, we compare its performance against the existing algorithms for the atrial fibrillation detection task and show how explainable the proposed model is in detecting atrial fibrillation arrhythmia.

4.1. Data description

To evaluate the proposed method, we used two datasets including the MIT-BIH AFIB database [24] and the PhysioNet Computing in Cardiology Challenge 2017 dataset [25].

4.1.1. MIT-BIH AFIB dataset

This dataset contains 23 long-term ECG recordings of human subjects with mostly atrial fibrillation arrhythmia. Each patient of the MIT-BIH AFIB includes two 10-h long ECG recordings (ECG1 and ECG2). The ECG signals are sampled at 250 Hz with a 12-bit resolution over a range of ±10 millivolts. In this study, we divided each ECG signal into 5-s segments and labeled each based on a threshold parameter, p. To perform the segment labeling, we followed the approach reported in Ref. [2,31]. A 5-s segment is labeled as AF if the percentage of annotated AF beats of the segment is greater than or equal to p, otherwise, it is determined as a non-AF arrhythmia. We chose $p = 50\%$ to be consistent with the previous research work. In our experiments, we used the ECG1 recordings (to be consistent with the existing methods) and extracted a total of 167,422 5-s data segments in which the number of AF segments was 66, 939 and the number of non-AF segments was 100, 483. As it is clear, the data segments are imbalanced. To deal with this problem and be able to compare our proposed model to the other existing algorithms, we randomly drew the same number of samples for both AF and non-AF classes (considered 66, 939 samples for each class). However, we tested the proposed method on the original imbalanced dataset as well.

4.1.2. PhysioNet challenge AFIB dataset

The goal of the challenge is to build the models to classify a single short ECG lead recording (30–60s in length) to normal sinus rhythm, atrial fibrillation (AF), an alternative rhythm, or too noisy classes. The training set includes 8528 single-lead ECG recordings and the test set contains 3658 ECG recordings. The test set has not been publicly available yet, therefore we use the training set for both test and training

phases. The ECG recordings were recorded by AliveCor devices, sampled as 300 Hz, and filtered by a bandpass filter. In this study, we considered only two classes including the normal sinus rhythm (N) and atrial fibrillation (AF) and discarded the remaining groups. Also, unlike the MIT-BIH AFDB, the ECG signals were not split into 5-s segments, indeed, the whole signal (i.e., 30–60s length) was considered.

4.2. Experimental setup

The proposed approach is based on hierarchical attention networks and has employed three levels of attention. To show the performance of this proposed model, in our experiments, we consider the model without the attention mechanism (denoted as RNN containing just the BiRNNs), one- (denoted as HAN-ECG1), two-(denoted as HAN-ECG2) and three-levels (denoted as HAN-ECG3) of the attention mechanism.

We applied a 10-fold cross-validation approach to evaluate the model. Indeed, we split the dataset into 10 folds. At each round of the cross-validation, 9 folds were used for training the model and the remaining fold (1 fold) was used for evaluating the model. In the end, we combined all the evaluation results.

The models were trained with a maximum of 25 epochs and mini-batches of size 64. The Adam optimizer was used to minimize the loss, L with a learning rate of $\alpha = 0.001$. We also used a L_2 regularization with a coefficient $\beta = 1e-5$ and a drop-out technique with a probability of dropping of 0.5 to reduce the effect of the overfitting problem during the training. The number of layers for the BiRNNs ware set to 2. The window and the hop sizes for the last attention layer were set to (2,2) and (5,5) for the MIT-BIH AFIB and AFDB17 databases, respectively. We utilized Python programming language and Google Tensorflow deep learning library to implement the proposed model. We ran the 10-fold cross-validation on a machine with 8 CPUs (Intel(R) Xeon(R) CPU @ 3.60 GHz), 32 GB memory and Ubuntu 18.04. In all experiments, the best performance was reported.

4.3. Results

4.3.1. Quantitative analysis

Table 1 shows the performance of the proposed method with different numbers of employed attention mechanisms against the state-of-the-art algorithms on the MIT-BIH AFIB database with the ECG segment of size 5-s. It can be seen from the table that the proposed method with one, two, and three hierarchies achieved quite better performance against other methods listed in the table. In Table 1, we can observe that the accuracy of the proposed method with two levels of attention is slightly higher than the one with three levels. The reason might be that the ECG segment of size 5-s (as input) has approximately 6 heartbeats in which almost all heartbeats contains the AF arrhythmia. Therefore, the heartbeats windowing at level three makes no significant improvement in the model performance.

Table 1
Comparison of performance of the proposed model against other algorithms on the MIT-BIH AFIB database with the ECG segment of size 5-s.

Method	Database	Best Performance (%)			
		Sensitivity	*Specificity*	*Accuracy*	*AUC*
HAN-ECG$_3$	AFDB	**99.08**	98.54	98.81	**99.86**
HAN-ECG$_2$	AFDB	98.88	**98.78**	**98.83**	99.85
HAN-ECG$_1$	AFDB	98.87	98.62	98.74	99.84
RNN	AFDB	98.72	98.44	98.58	99.80
HAN-ECG$_{2f}$	AFDB	98.68	98.36	98.52	99.79
Xia et al. (2018) [31]	AFDB	98.79	97.87	98.63	–
Asgari et al. (2015) [2]	AFDB	97.00	97.10	–	–
Lee et al. (2013) [14]	AFDB	98.20	97.70	–	–
Jiang et al. (2012) [10]	AFDB	98.20	97.50	–	–
Huang et al. (2011) [9]	AFDB	96.10	98.10	–	–
Babaeizadeh et al. (2009) [3]	AFDB	92.00	95.50	–	–
Dash et al. (2009) [4]	AFDB	94.40	95.10	–	–
Tateno et al. (2001) [30]	AFDB	94.40	97.20	–	–
Petrenas et al. (2015) [21]	AFDB	97.1	98.3	–	–
Zhou et al. (2014) [34]	AFDB	96.89	98.25	97.62	–

The row number 5 (i.e., the method named HAN-ECG2f) in Table 1 shows the evaluation results of the proposed method while the input ECG signals are split into fixed-size portions (here 180 samples for each portion as a heartbeat) and the portions are divided into fixed-size parts (here 6 parts for each portion and each part is considered as a distinct T-wave). It can be seen that the RNN method can perform as good as the method provided by Xia et al. [31] which is a deep convolutional neural network with the stationary wavelet transform (SWT) coefficient time series as input. In addition, Fig. 4 illustrates the confusion matrices' plots to describe a summary of how well the proposed model is performing given all folds.

It is worth mentioning that the last two methods, Petrenas et al. [21] and Zhou et al. [34], listed in Table 1, have developed their methods based on the Long Term Atrial Fibrillation database (LTAFDB) [23] and evaluated them using AFDB. However, our method and other methods in Table 1 have used the AFDB for both developing and evaluating the models.

The reported results by our proposed method and other listed methods in Table 1 (except the last two methods in the table) are based on balancing the dataset before training the models, in which the same number of non-AF data samples as the AF data samples are selected randomly. In addition, the selection of the 5-s data segments is from all combined data segments extracted from all individuals. Therefore, the training and evaluation sets can include data segments from the same subjects which is a data leakage problem. To have a more realistic evaluation mechanism, we considered another scenario in which the test and training data segments came from different individuals, and left the dataset imbalanced. Table 2 presents the performance of the proposed AF detectors with the new evaluation scenario on the MIT-BIH AFIB database. Since we could not find any research paper that followed the aforementioned scenario, we just reported our results without any comparison in Table 2. From Table 2, we can again note that the models with more attention layers yield higher accuracy and better performance.

Table 3 shows the experimental results on the PhysioNet Computing in Cardiology Challenge 2017 dataset. The overall performance of the proposed models with more than one attention layer (i.e., HAN-ECG2 and HAN-ECG2) is better than other methods, demonstrating the hierarchical attention networks work better for the AF detection task. Since, in this experiment, we considered a two-class problem (AF and normal classes), there was not any work in the literature to report a comparison.

4.3.2. Qualitative analysis

Understanding the cause of the model decision is very important in healthcare applications. In order to validate that the decisions made by our model are interpretable, we demonstrate through visualizing the hierarchical attention layers that the proposed method is considering clinically important heartbeats and waves in detecting the AF arrhythmia. Figs. 5 and 6 illustrate a few ECG signals containing the AF and non-AF categories. The top plots of the figures show the original ECG signals and the bottom plots present the informative heartbeats and waves in the detection of a class of interest (AF and non-AF). In the figures, the red segments denote the heartbeat weights and darker ones show more important heartbeats on the network's decision, and the blue strips and the yellow circles denote the locations of the important waves of the heartbeats in which the darker blue ones show more influence on

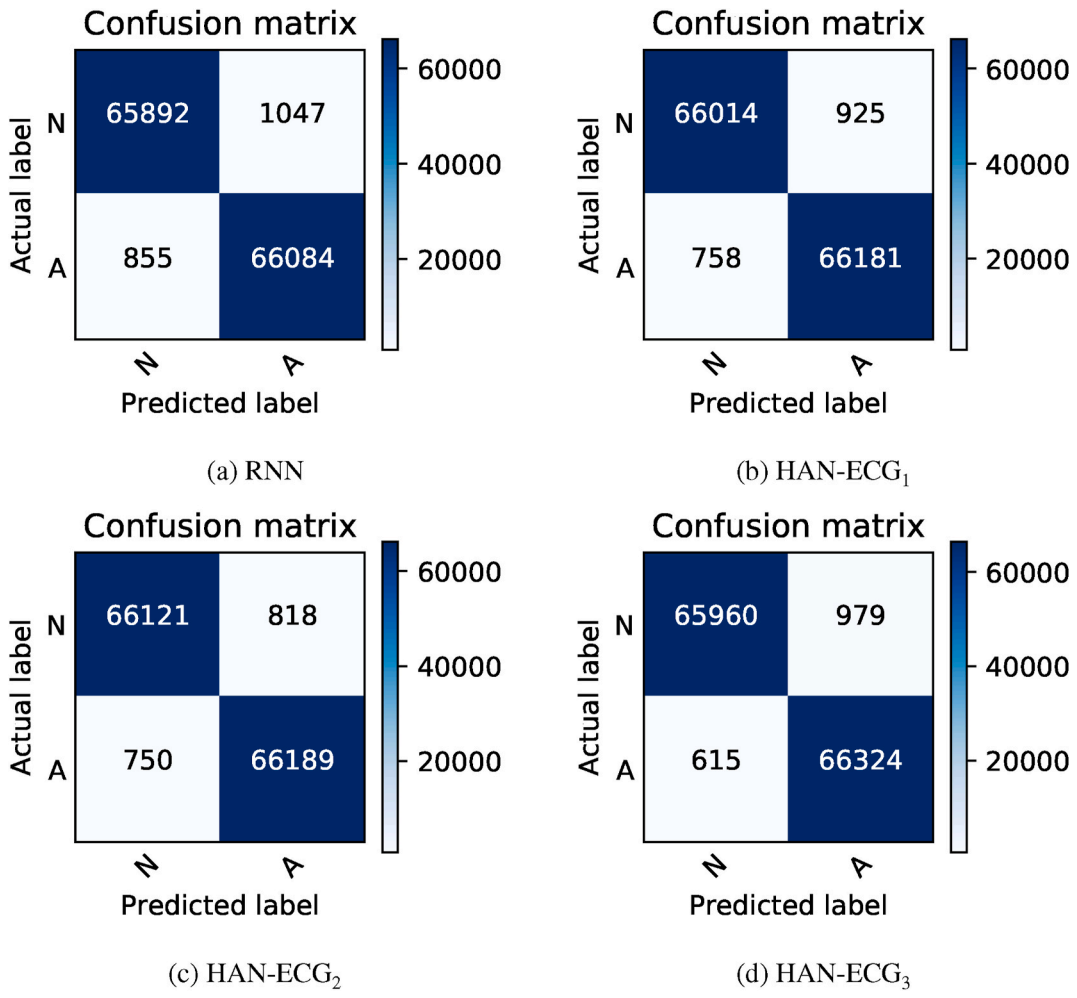


Fig. 4. Confusion matrices achieved by all the proposed method variants on the MIT-BIH AFIB database.

Table 2
Performance of the proposed model for the AF classification task on the MIT-BIH AFIB database while the database is not balanced and the data segments for the training and test phases come from different ECG recordings.

Method	Database	Best Performance (%)			
		Sensitivity	*Specificity*	*Accuracy*	*AUC*
HAN-ECG$_3$	AFDB	90.53	79.54	82.41	89.46
HAN-ECG$_2$	AFDB	89.86	77.49	81.58	88.65
HAN-ECG$_1$	AFDB	89.47	75.15	79.96	85.94
RNN	AFDB	89.20	74.38	79.55	85.88

Table 3
Performance of the proposed model for the AF classification task on the PhysioNet Computing in Cardiology Challenge 2017 dataset (AFDB17).

Method	Database	Best Performance (%)			
		Sensitivity	*Specificity*	*Accuracy*	*AUC*
HAN-ECG$_3$	AFDB17	86.02	98.62	96.98	98.46
HAN-ECG$_2$	AFDB17	86.15	98.50	96.90	98.41
HAN-ECG$_1$	AFDB17	84.30	98.48	96.64	98.44
RNN	AFDB17	80.52	97.40	95.18	97.18

the detection process.

There are two essential visual features in the patient ECG signals of that the practitioners use to identify the atrial fibrillation: (i) the absence of P-waves that occasionally are replaced by a series of small waves called fibrillation waves, and (ii) the irregular R-R intervals in which the heartbeat intervals are not rhythmic. Fig. 5 visualizes the important regions of the ECG signal while it contains AF arrhythmia. From the figure, we can see the importance of the heartbeats (i.e., through the intensity of the red segments), and that the proposed method pays attention to the irregularity of R-R intervals and emphasizes on the absence of P-waves which are the clinical features in recognizing the atrial fibrillation.

In addition, the proposed hierarchical attention mechanism considers the normal heartbeat rhythms for the detection of the non-AF class as shown in Fig. 6. In order to label an ECG signal as the AF, from Fig. 6 we can again observe that the model is interested in the parts of the ECG signal in which the P-waves are absent (replaced with a series of low-amplitude oscillations). Since all the heartbeats of the 5-s data segments in Fig. 6 and are either the normal heartbeats or the atrial fibrillation heartbeats, the importance of all the heartbeats approximately is the same (i.e., the same intensity for the red segments). Generally, in all aforementioned figures, our proposed model considers the clinically meaningful waves and their corresponding heartbeats in its decision-making process.

5. Discussion

One of the main challenges of building AF detection methods is that the number of AF samples is limited compared to the normal samples. Therefore, it results in performance degradation of machine learning

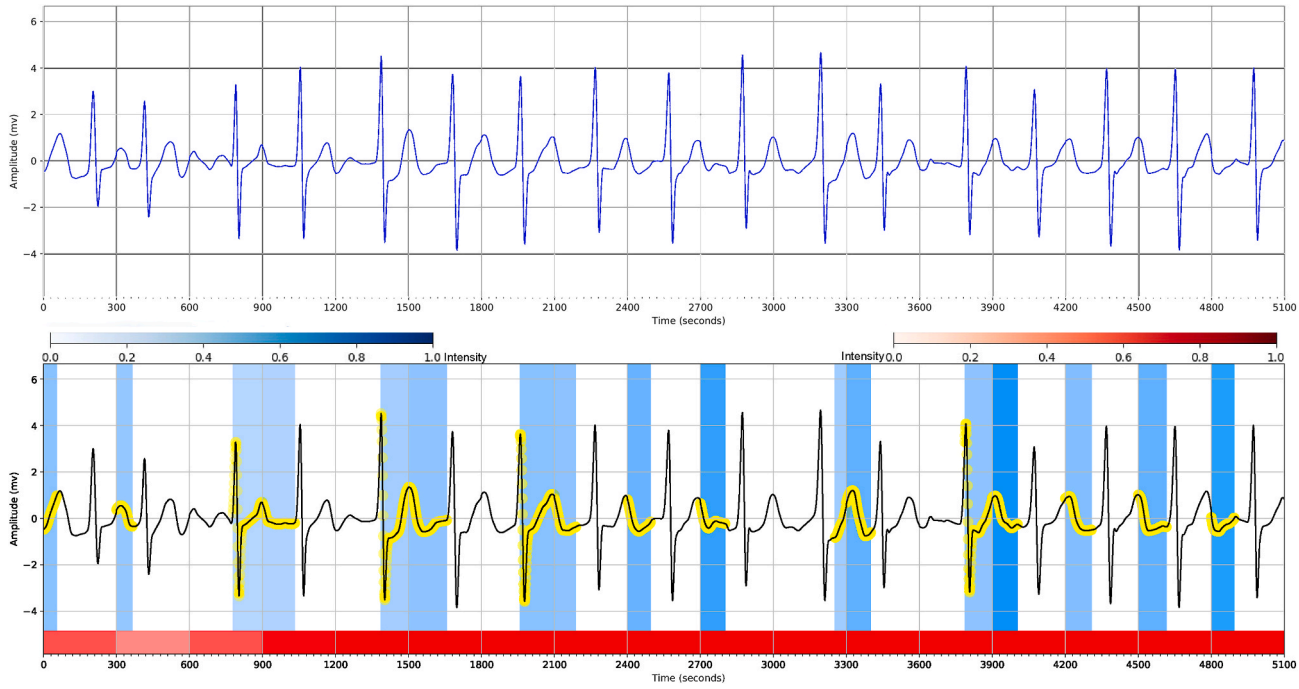
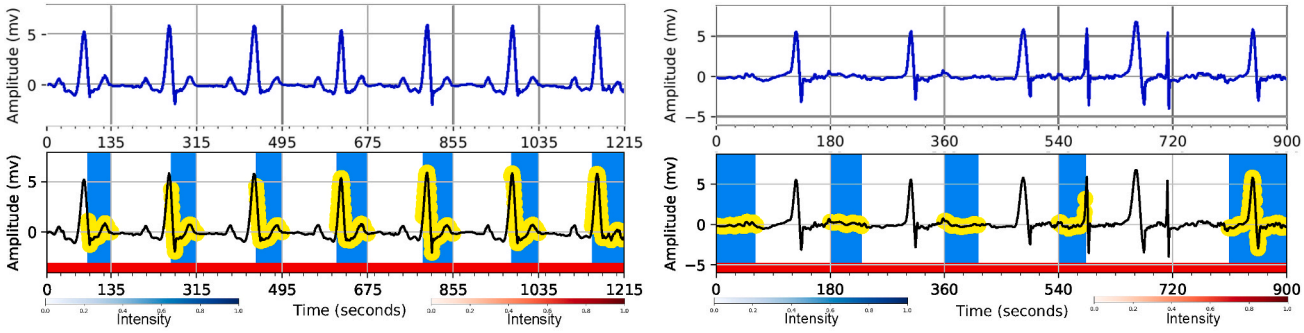


Fig. 5. Hierarchical attention visualization of a subject with the AF arrhythmia from the PhysioNet Computing in Cardiology Challenge 2017 dataset. The red segments depict the heartbeat weights and darker ones present more important heartbeats on the network's decision, and the blue strips and the yellow circles depict the locations of the important waves of the heartbeats in which the darker blue ones show more influence on the detection process.



(a) Subject without the AF arrhythmia (b) Subject with the AF arrhythmia

Fig. 6. Hierarchical attention visualization of two subjects from the MIT-BIH AFIB database. The first row shows the original ECGs and the second row shows the highlighted portions with attention. The blue/yellow parts depict the wave level attention and the red parts show the beat level attention. The darker blue ones show more influence on the detection process.

(ML)-based AF detection methods dealing with imbalanced datasets. Although such ML-based methods (i.e., deep learning methods) can benefit from good artificially generated AF samples by generative adversarial networks (GANs) [36], our main goal is to show the potential of hierarchical attention mechanisms to provide an interpretable arrhythmia detection method. In other words, our proposed method can interpret the reason behind the algorithm decisions.

Our AF detector with the power of interpretability can help physicians verify the performance of the algorithm, trust the results, find new patterns, etc.

Our proposed model have a few limitations including, first, the model performance is dependent on the pre-processing step where we extract R-peaks and split the ECG signals into small waves (P-, QRS- and T waves). For example, from Table 1, it is clear that the proposed method using the fixed-size heartbeats and the waves as input results in the lowest performance among all the proposed method variants. Hence, we can conclude that the pre-processing step in our methodology, as shown in Section 3.1, is necessary to obtain better performance. Second, regarding Table 1, we used just one dataset (i.e., AFDB) to build and test the model. Although it would be more realistic if the model development and evaluation were done using different databases, we employed AFDB for the two phases to be consistent with other existing methods in literature (i.e., Table 1).

In addition, a future research direction is to apply the proposed method to other ECG leads and other arrhythmias to extract new patterns that might be reasonable clinical features in the detection of an arrhythmia.

6. Conclusions

In this study, we proposed a hierarchical attention mechanism to accomplish the detection of atrial fibrillation using a single-lead ECG signal. The attention mechanisms allow us to interpret the detection results with the high resolution. The experiment results on two different databases reveal that our method achieves state-of-the-art performance and outperforms the existing algorithms. Furthermore, via

visualizations, we demonstrated that the pointed artifacts of signals by the model were clinically meaningful.

References

[1] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R. San Tan, A deep convolutional neural network model to classify heartbeats, Comput. Biol. Med. 89 (2017) 389–396.

[2] S. Asgari, A. Mehrnia, M. Moussavi, Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine, Comput. Biol. Med. 60 (2015) 132–142.

[3] S. Babaeizadeh, R.E. Gregg, E.D. Helfenbein, J.M. Lindauer, S.H. Zhou, Improvements in atrial fibrillation detection for real-time monitoring, J. Electrocardiol. 42 (2009) 522–526.

[4] S. Dash, K. Chon, S. Lu, E. Raeder, Automatic real time detection of atrial fibrillation, Ann. Biomed. Eng. 37 (2009) 1701–1709.

[5] H. Fujita, D. Cimr, Computer aided detection for fibrillations and flutters using deep convolutional neural network, Inf. Sci. 486 (2019) 231–239.

[6] C.D. Furberg, B.M. Psaty, T.A. Manolio, J.M. Gardin, V.E. Smith, P.M. Rautaharju, C.C.R. Group, et al., Prevalence of atrial fibrillation in elderly subjects (the cardiovascular health study), Am. J. Cardiol. 74 (1994) 236–241.

[7] J. Gao, H. Zhang, P. Lu, Z. Wang, An effective lstm recurrent network to detect arrhythmia on imbalanced ecg dataset, Journal of healthcare engineering 2019 (2019) 10. Article ID 6320651, https://doi.org/10.1155/2019/6320651.

[8] A. Ghaffari, N. Madani, Atrial fibrillation identification based on a deep transfer learning approach, Biomedical Physics & Engineering Express 5 (2019), 035015.

[9] C. Huang, S. Ye, H. Chen, D. Li, F. He, Y. Tu, A novel method for detection of the transition between atrial fibrillation and sinus rhythm, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 58 (2011) 1113–1119.

[10] K. Jiang, C. Huang, S.m. Ye, H. Chen, High accuracy in automatic detection of atrial fibrillation for holter monitoring, J. Zhejiang Univ. - Sci. B 13 (2012) 751–756.

[11] T.J. Jun, H.M. Nguyen, D. Kang, D. Kim, D. Kim, Y.H. Kim, Ecg Arrhythmia Classification Using a 2-d Convolutional Neural Network, 2018 arXiv preprint arXiv:1804.06812.

[12] M. Kachuee, S. Fazeli, M. Sarrafzadeh, Ecg heartbeat classification: a deep transferable representation, in: 2018 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2018, pp. 443–444.

[13] D.E. Lake, J.R. Moorman, Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices, Am. J. Physiol. Heart Circ. Physiol. 300 (2010) H319–H325.

[14] J. Lee, B.A. Reyes, D.D. McManus, O. Maitas, K.H. Chon, Atrial fibrillation detection using an iphone 4s, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 60 (2013) 203–206.

[15] E.P. Lehman, R.G. Krishnan, X. Zhao, R.G. Mark, L.W.H. Lehman, Representation learning approaches to detect false arrhythmia alarms from ecg dynamics, Proceedings of machine learning research 85 (2018) 571.

[16] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, J. Gao, Dipole: diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2017, pp. 1903–1911.

[17] S. Mousavi, F. Afghah, Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 1308–1312.

[18] S. Mousavi, F. Afghah, A. Razi, U.R. Acharya, Ecgnet: learning where to attend for detection of atrial fibrillation with deep visual attention, in: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, 2019, pp. 1–4.

[19] S. Mousavi, A. Fotoohinasab, F. Afghah, Single-modal and multi-modal false arrhythmia alarm reduction using attention-based convolutional and recurrent neural networks, PloS One 15 (2020), e0226990.

[20] J. Pan, W.J. Tompkins, A real-time qrs detection algorithm, IEEE Trans. Biomed. Eng. 32 (1985) 230–236.

[21] A. Petrėnas, V. Marozas, L. Sörnmo, Low-complexity detection of atrial fibrillation in continuous long-term monitoring, Comput. Biol. Med. 65 (2015) 184–191.

[22] PhysioNet, PhysioNet community. https://www.physionet.org/, 2000.

[23] PhysioNet, PhysioNet community. http://physionet.org/physiobank/database/, 2000.

[24] PhysioNet, PhysioNet MIT-BIH atrial fibrillation database. https://physionet.org/content/afdb/1.0.0/, 2000.

[25] PhysioNet, AF classification from a short single lead ECG recording - the PhysioNet computing in Cardiology challenge 2017. https://physionet.org/physiobank/database/mitdb/, 2001.

[26] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng, Cardiologist-level Arrhythmia Detection with Convolutional Neural Networks, 2017 arXiv preprint arXiv:1707.01836.

[27] S. Saadatnejad, M. Oveisi, M. Hashemi, Lstm-based ecg classification for continuous monitoring on personal wearable devices, IEEE J. Biomed. Health Inform. 24 (2) (2019) 515–523. IEEE.

[28] M. Schuster, K.K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (1997) 2673–2681.

[29] S.P. Shashikumar, A.J. Shah, G.D. Clifford, S. Nemati, Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, 2018, pp. 715–723.

[30] K. Tateno, L. Glass, Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of rr and δrr intervals, Med. Biol. Eng. Comput. 39 (2001) 664–671.

[31] Y. Xia, N. Wulan, K. Wang, H. Zhang, Detecting atrial fibrillation by deep convolutional neural networks, Comput. Biol. Med. 93 (2018) 84–92.

[32] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical attention networks for document classification, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 1480–1489.

[33] Ö. Yıldırım, P. Pławiak, R.S. Tan, U.R. Acharya, Arrhythmia detection using deep convolutional neural network with long duration ecg signals, Comput. Biol. Med. 102 (2018) 411–420.

[34] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, Y. Zhang, Automatic online detection of atrial fibrillation based on symbolic dynamics and shannon entropy, Biomed. Eng. Online 13 (2014) 18.

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

[36] F. Zhu, F. Ye, Y. Fu, Q. Liu, B. Shen, Electrocardiogram generation with a bidirectional lstm-cnn generative adversarial network, Sci. Rep. 9 (2019) 1–11.

Sajad Mousavi received a B.Sc. degree in computer engineering from Zanjan University (ZNU), Zanjan, Iran, in 2010, and an M. Sc. degree in artificial intelligence and robotics from the Iran University of Science and Technology (IUST), Tehran, Iran, in 2012. He is currently pursuing the Ph.D. degree with the School of Informatics, Computing and Cyber Systems, Northern Arizona University. He worked as a Research Assistant in machine learning and deep learning with the National University of Ireland Galway, Galway, Ireland, from 2015 to 2017. His main research interests include machine learning, deep learning, computer vision, multiagent systems, and task allocation.

Fatemeh Afghah was an Assistant Professor with the Electrical and Computer Engineering Department, North Carolina A\&T State University, Greensboro, NC, USA, from 2013 to 2015. She is an Assistant Professor with the School of Informatics, Computing and Cyber Systems, Northern Arizona University (NAU), Flagstaff, AZ, USA, where she is the Director of Wireless Networking and Information Processing (WiNIP) Laboratory. Her research interests include wireless communication networks, decision making in multiagent systems, radio spectrum management, hardware-based security, and artificial intelligence in healthcare. She received the Air Force Office of Scientific Research Young Investigator Award, in 2019 and the NSF CRII Award, in 2017.

U. R. Acharya, MTech, PhD, DEng, DSc is a senior faculty member at Ngee Ann Polytechnic, Singapore. He is also (i) Adjunct Professor at University of Malaya, Malaysia, (ii) Adjunct Professor at Asia University, Taiwan, (iii) Associate faculty at Singapore University of Social Sciences, Singapore and (iv) Adjunct Professor at University of Southern Queensland, Australia. He received his Ph.D. from National Institute of Technology Karnataka (Surathkal, India), DEng from Chiba University (Japan) and DSc from AGH University of Science and Technology, Poland. He has published more than 500 papers, in refereed international SCI-IF journals (345), international conference proceedings (42), books (17) with more than 34,000 citations in Google Scholar (with h-index of 97). He has worked on various funded projects, with grants worth more than 5 million SGD. He is ranked in the top 1% of the Highly Cited Researchers for the last five consecutive years (2016 to 2020) in Computer Science according to the Essential Science Indicators of Thomson. He is one among the World's Top 100,000 Scientists in 2019. He is on the editorial board of many journals and has served as Guest Editor for many journals.