A Novel Recurrent Neural Network Architecture for Classification of Atrial Fibrillation Using Single-lead ECG

Rohan Banerjee, Avik Ghose and Sundeep Khandelwal
Research and Innovation, Tata Consultancy Services
{rohan.banerjee, avik.ghose, sundeep.khandelwal}@tcs.com

***Abstract*—Atrial Fibrillation (AF) is a type of abnormal heart rhythm which may lead to a stroke or cardiac arrest. In spite of numerous research works, developing an automatic mechanism for accurate detection of AF remains a popular yet unsolved problem. In this paper, we propose a deep neural network architecture for classification of AF using single-lead Electrocardiogram (ECG) signals of short duration. We define a novel Recurrent Neural Network (RNN) structure, comprising two Long-Short Term Memory (LSTM) networks for temporal analysis of RR intervals and PR intervals in an ECG recording. Output states of the two LSTMs are merged at the dense layer along with a set of hand-crafted statistical features, related to the measurement of heart rate variability (HRV). The proposed architecture is proven on the open access PhysioNet Challenge 2017 dataset, containing more than 8500 single-lead ECG recordings. Results show that our methodology yields sensitivity of 0.93, specificity of 0.98 and F1-score of 0.89 in classifying AF, which is better than the existing accuracy scores, reported on the dataset.**

***Index Terms*—Atrial Fibrillation, Electrocardiogram, Long-Short Term Memory, Classification**

I. Introduction

Atrial Fibrillation (AF) is an electrophysiological disorder, caused when abnormal electrical impulses suddenly start firing in the atria. The heart's normal rhythm goes awry, resulting in an abnormally fast heart rate with an enhanced risk of stroke and heart attack. Being one of the most common type of arrhythmias, AF is associated with significant mortality and morbidity. Presence of AF affects the morphology of the Electrocardiogram (ECG) and can be visually identified by an expert. However, manual detection of intermittent AF episodes from prolonged ECG recordings is challenging and often impractical. Hence, automatic detection of AF becomes an important area of research. As shown in Fig. 1 [1], a complete ECG cycle of a normal subject comprises three major components, the P wave followed by the QRS complex and the T wave. The P wave represents the depolarization of the atria, the QRS complex represents the depolarization of the ventricles and the T wave represents the repolarization of the ventricles. The PR interval (also known as PQ interval) measures the time from the initial depolarization of the atria to the initial depolarization of the ventricles and the QT interval measures the time in which the ventricles depolarize and repolarize. The distance between successive R peaks (RR interval) is used to measure the instantaneous heart rate. Automatic AF detectors, available in literature broadly belong

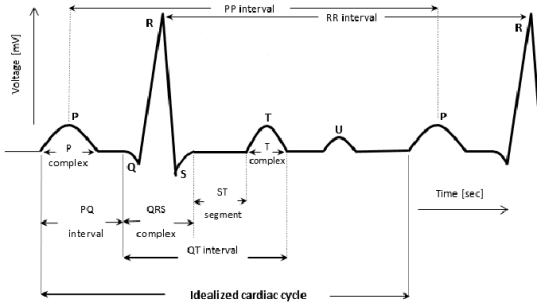

Fig. 1. Complete cycle of a normal ECG waveform, source of image: [1]

to two categories, 1) atrial analysis based approaches and 2) ventricular response based approaches. Atrial analysis based approaches look for the absence of P waves or the presence of fibrillatory f-waves in ECG during depolarization of atria for an AF event. Ventricular response based approaches analyse the irregularities in heart rate for AF. Atrial analysis based approaches are known to be more accurate but are vulnerable to background noise.

Features, derived from scatter plot of successive RR intervals using Poincaré [2] and Lorenz [3] plots are found in prior literature for classifying AF. Statistical features are extracted from RR intervals in [4] to measure the irregularity in heart rate variability (HRV) due to AF. A number of recently developed algorithms can be found to classify normal, AF and other abnormal heart rhythms using the short single-lead ECG recordings, provided during the PhysioNet Challenge 2017 [5]. Promising accuracy scores are reported on the challenge dataset by Datta *et al.* [6] and Zabihi *et al.* [7] using novel statistical and domain specific features. Novel architectures of Convolution Neural Networks (CNN), Recurrent Neural Networks (RNN) and their effective combinations are proposed in [8], [9] and [10]. A combination of classical and deep learning approaches can be found in [11]. In spite of numerous works, designing of an accurate AF detector using single-lead ECG is still an unsolved problem. Available deep learning based approaches apply the raw ECG signals directly to the network to learn the desired pattern automatically. In this paper, we propose a composite RNN structure, combining domain knowledge and classical signal processing approaches for an improved AF classifier. Our key contributions are:

1) Novel neural network architecture, based on domain

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

knowledge for temporal analysis of RR intervals and PR intervals in ECG using a pair of independent Long-Short Term Memory (LSTM) networks.

2) Combining the output states of the two LSTMs with a set of hand-crafted features, related to the measurement of HRV to design a composite network structure for AF classification.

We prove the efficacy of our proposed methodology on the large open access dataset of PhysioNet Challenge 2017. Our reported accuracies are found better than state-of-the-art approaches, reported on the dataset. In Section II, we describe the architecture of our proposed network, detailing its input-output structure. Our experimental dataset, different network parameters used to create the training model and experimental results are provided in Section III, followed by conclusion in Section IV.

II. Proposed Network Architecture

Typical morphological differences in single-lead ECG waveforms between a non-AF and an AF subject are shown in Fig. 2. As shown, there are two important clinical markers for identifying an AF event, 1) absence of P waves or presence of fibrillatory f-waves before the QRS complex and 2) irregular RR intervals [12], [13]. Most of the available techniques aim at detecting any one of these two markers for an AF episode. Our proposed methodology is capable of combining both of them in a single network structure. This is done by a pair of LSTM networks for temporal analysis of RR intervals and a series of P wave regions, extracted from ECG. Fig. 3 shows our proposed composite network architecture along with tensor dimensions at the output of each layer. The inputs to the LSTMs are shown in the format of [batch size, number of time steps, data dimension]. In sequence models, RNNs use their internal memory to process an input time series for extraction of temporal patterns that can be used for classification or prediction. LSTM is a class of RNN that can effectively learn a longer pattern of unknown length as it can deal with the exploding and the vanishing gradient problems, faced by RNNs during training. LSTM does this because of its unique cell structure that enables deleting less important information from memory. For an input sequence $x_t = \{x_1, x_2, ...x_T\}$ of length T, an LSTM cell with one forget gate computes a hidden vector sequence $h_t = \{h_1, h_2, ...h_T\}$ by iterating the following equations over t.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (1)$$
$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (2)$$
$$c_t = f_t * c_{t-1} + i_t * tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (4)$$
$$h_t = o_t * tanh(c_t) \quad (5)$$

W_x, W_h and b represent the weight matrices of the input, the recurrent connections and the bias terms; σ represents the logistic sigmoid function; $*$ denotes element-wise product operation. Input gate, forget gate, output gate and cell activation vectors are represented by i, f, o and c. Hyperbolic tangent $(tanh)$ is used as the input activation function.

Output states of the two LSTMs are merged at the dense layer along with a set of hand-crafted statistical features, quantifying short term HRV. After concatenation, the output of the dense layer is applied to a pair of fully connected layers and softmax function for binary classification. A detail description of different inputs to the proposed composite network are provided subsequently.

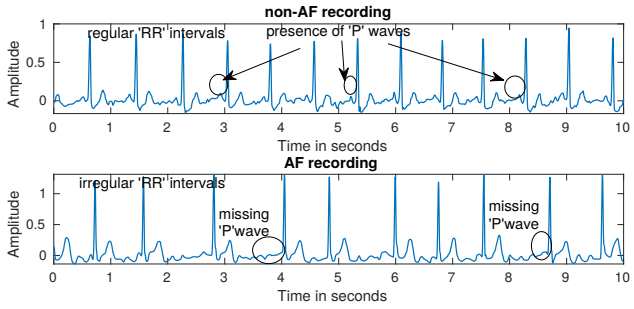

Fig. 2. Typical ECG waveform for non-AF and AF

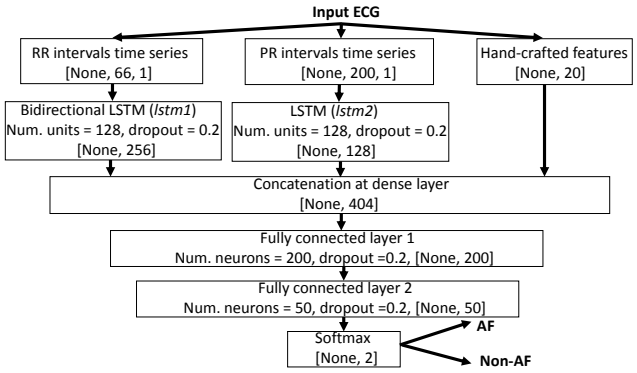

Fig. 3. Proposoed composite network architecture for classifying AF

A. RR Intervals Time Series

Irregular HRV is considered as one of the common symptoms of AF. Although the instantaneous heart rate of a healthy non-cardiac person changes with time, the HRV pattern is different than a cardiac patient [13]. However, quantifying the exact pattern of HRV, caused due to AF is challenging. HRV directly affects the ECG waveform, as the R peaks do not repeat after a fixed interval. In our proposed approach, we design a Bidirectional LSTM (BiLSTM) network for measuring the long term temporal dependencies in successive RR intervals in an ECG recording. Both previous and future context of a time series can be effectively utilized in BiLSTM, as it processes the input sequence in both forward and backward direction. The open source implementation of Behar's algorithm [14] is applied on our experimental dataset [5] for locating the R peaks in an ECG recording. Irregular sampling rate of the

extracted RR intervals time series due to HRV is fixed to 2 Hz using cubic-spline interpolation technique. The time series is mapped to the range of 0 to 1. Duration of an ECG recording is selected as 33 seconds in our architecture for analysing RR intervals and the explanation is provided in Section III-B.

B. PR Intervals Time Series

Irregular HRV occurs due to a number of pathological conditions, both cardiac and non-cardiac in nature [12], [13]. Hence, HRV is not always the sufficient marker for detecting an AF event. Atrial activities of the heart behave chaotically during AF. This affects the PR intervals in the ECG in terms of few missing P waves or presence of f-waves. Hence, the morphology of PR intervals in successive cardiac cycles looks different than a normal ECG. An accurate segmentation of P waves using signal processing techniques is difficult due to their varying morphology and vulnerability to background noise. Here, we design a second LSTM network for temporal analysis of a series of PR interval regions in successive cardiac cycles. The input to the LSTM is formed by selecting a window before the QRS complex, where the P wave is likely to be located and stacking multiple such windows on time axis. Typical duration of PR intervals lies between 120-200 ms [13]. Hence, the window length is selected as 200 ms and each window is selected from a location before the QRS complex so that it ends 33 ms before the reference R peak. Since our experimental dataset is sampled at 300 Hz [5], each window contains 60 data points. A total of 10 such windows are selected from consecutive cardiac cycles and merged to construct an input sequence. This is further normalized to the range of 0 to 1 and down-sampled by a factor of 3 to reduce the subsequent processing load.

C. Hand-crafted Features

Several statistical parameters exist in literature to define the randomness in a time series. An LSTM is not always guaranteed to learn the desired pattern from a time series, containing multiple independent patterns. We propose an optimum set of 20 features, extracted from RR intervals time series $RR_t = \{RR_1, RR_2, ...RR_m\}$ to mathematically quantify the extent of HRV, where m represents the number of RR intervals in a recording. Sampling rate of the time series is set to 2 Hz. The features are selected from a larger set of features, used in various applications. Fisher score is a popular tool that aims to find a subset of features so that the distances between points in different classes are as large as possible, while the distances between points in the same class are as small as possible. Fisher score for i^{th} feature (S_i) is calculated as: $S_i = \frac{\Sigma n_j(\mu_{ij}-\mu_i)^2}{\Sigma n_j * \rho_{ij}^2}$, where μ_{ij} and ρ_{ij} are the mean and the standard deviation of the i^{th} feature in the j^{th} class, n_j is the number of instances in the j^{th} class and μ_i is the mean of the i^{th} feature. Initially, we rank all features on a labelled training dataset using Fisher score. The optimum features are selected from the space of ranked features using brute-force search, so that sensitivity is maximized on a validation set in classifying AF with minimum features. The training set is used to create the learning model. A neural network having two hidden layers, similar to the configuration of the fully connected layers in Fig. 3 is used for classification. For a largely unbiased dataset, the minority class is oversampled using Synthetic Minority Oversampling Technique (SMOTE) before training, so that the features are not biased to the majority class. The selected 20 features are defined below.

Approximate entropy, $ApEn(RR_t, q, r)$ quantifies the irregularity and complexity of RR_t in terms of a predefined pattern length q and a similarity criterion parameter r. A sequence of vectors $\{x_q(1), x_q(2), ...x_q(m-q+1)\}$ in real q-dimensional space is defined from RR_t, such that $x_q(i) = \{RR_i, RR_{i+1}, RR_{i+2}, ...RR_{i+q-1}\}$. Two such vectors $x_q(i)$ and $x_q(j)$ are similar if $|RR_{i+k} - RR_{j+k}| < r,\ for\ 0 < k < q$. We define $C_{iq}(r)$, where $C_{iq}(r) = (number\ of\ x_q(j)\ similar\ to\ x_q(i))/(m-q+1)$. If $C_q(r)$ indicates the mean of all $C_{iq}(r)$ for $i \in 1...m-q+1$, $ApEn$ is defined as:

$$ApEn(RR_t, q, r) = ln[\frac{C_q(r)}{C_{q+1}(r)}] \tag{6}$$

Sample Entropy, $SampEn(RR_t, q, r)$ is a modification of approximate entropy for measuring the dynamics of a time series and is defined as:

$$SampEn(RR_t, q, r) = -ln[\frac{A}{B}] \tag{7}$$

Here, A = number of vector pairs where $|x_{q+1}(i) - x_{q+1}(j)| < r$, B = number of vector pairs where $|x_q(i) - x_q(j)| < r$. Shannon entropy of RR_t is measured as:

$$E_{sh} = -\sum_{b=1}^{N} pr_b \log pr_b \tag{8}$$

A normalized histogram of N bins is created from RR_t. Empirical probability of b^{th} histogram bin is denoted by pr_b. Here, $b \in 1...N$ and $\sum_{b=1}^{N} pr_b = 1$.

Other features are mean, median, variance, maximum, minimum, range, kurtosis and skewness of RR_t. The fraction of RR interval pairs in a recording differ by more than 20 ms ($pNN20$) and 50 ms ($pNN50$) [6] are two more features. The remaining features are derived from Poincaré [2] and Lorenz plot [3] of RR_t. The selected features are normalized to zero mean and unit variance before merging with the output states of the two LSTM networks.

III. Data Analysis

A. Dataset Description

The PhysioNet Challenge 2017 dataset [5] is an open access dataset, comprising a total of 8528 single-lead ECG data, recorded by a commercially available sensor. As per available annotations, 5154 recordings of the dataset are normal sinus rhythms, 771 are AF, 2557 are non-AF but abnormal rhythms and remaining 46 recordings are noisy. Since our proposed algorithm is designed for detecting AF, we re-label the dataset for binary classification by merging all non-AF recordings into a single class. Thus, the modified dataset contains 771 AF and 7757 non-AF recordings. This makes the dataset quite imbalanced with a ratio of majority (non-AF) to minority (AF) class is close to $10:1$. The recordings are sampled at 300 Hz

and bandpass filtered with cut off frequencies of 0.5 Hz and 40 Hz in the dataset. We partition the entire dataset into three portions based on random selection, maintaining the skewness of the original dataset. 60% of the total dataset is selected for training, 20% for internal validation and the remaining 20% for testing. The training and the validation sets are used for optimization of different network parameters, selection of the hand-crafted features and creation of the training model. The final evaluation is done on the test set.

B. Selection of Network Parameters

Durations of the recordings in the dataset vary from 9 seconds to 61 seconds. However, an LSTM structure requires a fixed number of time steps for all its input instances. The histogram in Fig. 4 shows that most of the recordings have a duration close to 30 seconds with a mean duration of 32.5 seconds. Hence, we select 33 seconds as the optimum duration for analysing RR intervals to learn the HRV pattern effectively by the BiLSTM network. Since the sampling rate of the RR intervals is fixed to 2 Hz, the network analyses 66 data points for every input. The recordings with shorter durations are repeated from beginning to achieve the desired length. For the

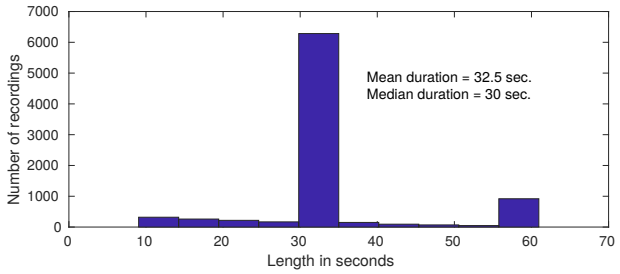

Fig. 4. Historgram of data length in PhysioNet dataset

analysis of PR intervals by the second LSTM, 10 consecutive cycles are selected from a random location in a recording. This is done based on the shortest available recording in the dataset. This supports the perception of the clinicians, as an event of missing P waves or presence of f-waves during 10 consecutive cardiac cycles can be marked as AF [12]. Being independent to the duration of the input time series, the hand-crafted features are calculated from its entire length.

The proposed composite network is trained on an Intel® Xeon(R) 16-core processor having 64 GB of RAM. The implementation is done using Keras API with TensorFlow. Dropout is applied in the LSTM layers and the fully connected layers to mitigate the chance of over-fitting. The neurons in the fully connected layers are activated using Rectified Linear Unit (ReLU) function and their initial weights are set using Xavier initialization [15]. Here, initial weights of the neurons are randomly assigned from a Gaussian distribution of zero mean and a finite variance $var = \frac{2}{n_{in}+n_{out}}$, where n_{in} and n_{out} are the number of input and output neurons in the layer. The cross entropy loss is minimized during training using Adam optimizer with 100 epochs, mini-batch size of 64 and learning rate of 0.001. In order to overcome the class imbalance problem in the dataset, the minority class is assigned 10 times more class-weight than the majority class. Hence, the gradient computed from the instances of the minority class becomes 10 times larger than the majority class, paying more attention to AF. Fig. 5 shows the cross entropy loss for different epochs, obtained on the training and the validation data by our proposed network. It can be observed that the loss on both the training and the validation data gets minimized and saturates after 80 epochs. The final training model is created using 100 epochs.

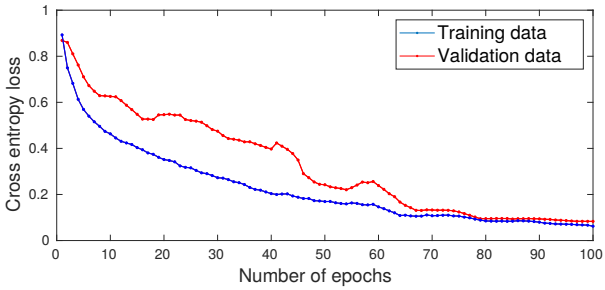

Fig. 5. Cross entropy loss vs. number of epochs by the proposed network

C. Experimental Results

Classification performance of the proposed methodology is reported based on sensitivity (Se), specificity (Sp) and positive predictive value (PPV) of detecting AF. These metrics are defined in terms of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) as:

$$Se = \frac{TP}{TP+FN},\ Sp = \frac{TN}{TN+FP},\ PPV = \frac{TP}{TP+FP} \quad (9)$$

For a comprehensive analysis, the composite network is split into four smaller networks based on individual inputs, 1) the RR intervals sequences fed to the BiLSTM ($lstm1$ in Fig. 3), 2) the PR intervals sequences fed to the second LSTM ($lstm2$ in Fig. 3), 3) only the hand-crafted features are used for classification and 4) $lstm1$ and $lstm2$ are merged at dense layer ($lstm3$) without hand-crafted features. All of them are trained individually and compared with the proposed composite network on the validation and test sets. In all cases, the configuration of the fully connected layers for binary classification are kept identical to the actual composite network. The comparative study is shown in Table I. It can be observed that, $lstm1$ yields a very high specificity but a relatively lower sensitivity as it sometimes fails to detect the desired HRV pattern, specific to AF from a short ECG recording. The hand-crafted features are selected in such a way that a high sensitivity score is achieved. However, specificity and PPV are compromised. It is observed that $lstm2$ does not produce a promising accuracy individually. A relatively short portion of data (only 10 cardiac cycles) is analysed by $lstm2$, where the signature of missing P waves may not be always present. However, $lstm2$ significantly improves the sensitivity of $lstm1$ when the two networks are merged in $lstm3$ as it combines two independent AF markers in a

single network. However, the hand-crafted features alone yield higher sensitivity than $lstm1$, $lstm2$ and $lstm3$. Thus, in our proposed structure, we merge these features with the output of $lstm3$ at the dense layer. This significantly improves the sensitivity and PPV of $lstm3$ without affecting its specificity, resulting in, an optimum classification performance.

Performance of the proposed composite network is compared with top four joint winning entries in the PhysioNet Challenge 2017 [5]. The submitted entries in the challenge were ranked based on average F1-score of detecting different classes on a separate hidden test dataset. Since we propose a binary AF classifier, F1-score of classifying AF as reported by the prior algorithms on the open access portion of the dataset is considered here for comparison. F1-score is defined by combining precision (pr) and recall (re) in a single metric:

$$F1 = 2 * \frac{pr * re}{pr + re},\ where\ pr = \frac{TP}{TP + FP},\ re = \frac{TP}{TP + FN} \tag{10}$$

Existing algorithms, successfully evaluated on the dataset applied both classical and deep learning approaches. Table II shows that our methodology outperforms the top scoring entries in the challenge, in terms of F1-score of classifying AF. This proves the efficacy of our proposed approach. Due to unavailability in public domain, our methodology has not yet been evaluated on the hidden data, used in the challenge for ranking the submitted entries.

TABLE I
PERFORMANCE OF INDIVIDUAL COMPONENTS OF THE PROPOSED COMPOSITE NETWORK IN AF CLASSIFICATION

Network Structure	Validation set			Test set		
	Se	Sp	PPV	Se	Sp	PPV
RR intervals based BiLSTM ($lstm1$)	0.86	0.98	0.73	0.84	0.98	0.72
PR intervals based LSTM ($lstm2$)	0.82	0.89	0.42	0.79	0.91	0.46
Hand-crafted 20 features	0.96	0.90	0.53	0.94	0.91	0.52
$lstm1$ and $lstm2$ merged at dense layer ($lstm3$)	0.89	0.97	0.74	0.88	0.96	0.70
The composite network	**0.93**	**0.98**	**0.85**	**0.93**	**0.98**	**0.85**

TABLE II
COMPARISON WITH EXISTING APPROACHES IN AF CLASSIFICATION

Authors	Methodology used	F1-score
Teijeiro *et al.* [8]	Combination of RNN and XGBoost	0.85
Datta *et al.* [6]	Domain specific features followed by cascaded binary AdaBoost classifier	0.80
Zabihi *et al.* [7]	150 features in time, frequency and time-frequency domain are fed to a random forest classifier	0.79
Hong *et al.* [16]	Ensemble of DNN and domain specific features	0.85
Our proposed method	**Merged LSTM and hand-crafted features**	**0.89**

IV. CONCLUSION

An accurate AF detector from single-lead ECG is an unsolved research problem. In this paper, we propose a novel architecture of an AF detector, combining a pair of LSTM networks for temporal analysis of ECG time series and a set of statistical parameters, related to HRV. The proposed methodology is evaluated on a large open access ECG dataset, outperforming the reported accuracy scores. The current architecture is specifically designed to identify AF. We are working on to enhance it to detect other cardiac diseases to design a complex cardiac anomaly detection platform.

REFERENCES

[1] A. Szczepański and K. Saeed, "A mobile device system for early warning of ecg anomalies," *Sensors*, vol. 14, no. 6, pp. 11 031–11 044, 2014.

[2] J. Park, S. Lee, and M. Jeon, "Atrial fibrillation detection by heart rate variability in poincare plot," *Biomedical engineering online*, vol. 8, no. 1, p. 38, 2009.

[3] S. Sarkar, D. Ritscher, and R. Mehra, "A detector for a chronic implantable atrial tachyarrhythmia monitor," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1219–1224, 2008.

[4] D. Lake and J. R. M. Moorman, "Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 300, no. 1, 2011.

[5] G. Clifford *et al.*, "Af classification from a short single lead ecg recording: The physionet computing in cardiology challenge 2017," *Proceedings of Computing in Cardiology*, vol. 44, p. 1, 2017.

[6] Datta *et al.*, "Identifying normal, af and other abnormal ecg rhythms using a cascaded binary classifier," *Computing*, vol. 44, p. 1, 2017.

[7] M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection of atrial fibrillation in ecg hand-held devices using a random forest classifier," *Computing*, vol. 44, p. 1, 2017.

[8] T. Teijeiro, C. A. García, D. Castro, and P. Félix, "Arrhythmia classification from the abductive interpretation of short single-lead ecg records," *Comput. Cardiol*, vol. 44, pp. 1–4, 2017.

[9] P. Warrick and M. N. Homsi, "Cardiac arrhythmia detection from ecg combining convolutional and long short-term memory networks," *Computing*, vol. 44, p. 1, 2017.

[10] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," *Computing*, vol. 44, p. 1, 2017.

[11] F. Plesinger, P. Nejedly, I. Viscor, J. Halamek, and P. Jurak, "Parallel use of a convolutional neural network and bagged tree ensemble for the classification of holter ecg," *Physiological measurement*, vol. 39, no. 9, p. 094002, 2018.

[12] A. John, P. Kirchhof, G. Y. Lip, U. Schotten, I. Savelieva, S. Ernst, I. C. Van Gelder *et al.*, "Guidelines for the management of atrial fibrillation: the task force for the management of atrial fibrillation of the european society of cardiology (esc)," *European heart journal*, vol. 31, no. 19, pp. 2369–2429, 2010.

[13] M. Shenasa, M. E. Josephson, and N. M. Estes III, *The ECG handbook of contemporary challenges*. Cardiotext Publishing, 2015.

[14] A. E. Johnson, J. Behar, F. Andreotti, G. D. Clifford, and J. Oster, "R-peak estimation using multimodal lead switching," in *Computing in Cardiology 2014*, 2014, pp. 281–284.

[15] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[16] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, "Encase: An ensemble classifier for ecg classification using expert features and deep neural networks," in *Computing in Cardiology (CinC), 2017*. IEEE, 2017, pp. 1–4.