Different training strategy for deep learning based AF detection

Xiangyu Zhang, Jianqing Li*, Zhipeng Cai, Li Zhang, Chengyu Liu*, *Member, IEEE,*

***Abstract*—Nowadays, deep learning-based models have been widely developed for atrial fibrillation (AF) detection in electrocardiogram (ECG) signals. However, due to the unavoidable over-fitting problem, classification accuracy of the developed models severely differed when applying on the tested independent datasets. This situation is more significant for AF detection from dynamic ECGs. In this study, we explored two potential training strategies to address the over-fitting problem in AF detection. The first one is to use the Fast Fourier transform (FFT) and Hanning-window based filter to suppress the influence from individual difference. Another is to train the model on the wearable ECG data rather than the traditional ECG data to improve the robustness of model. Wearable ECG data from 29 patients with arrhythmia were collected for at least 24 hours during the daily life, without any activity limitation. To verify the effectiveness of the training strategies, an LSTM and CNN based model was proposed and tested. We tested the model on the independent wearable ECG data, as well as in the MIT-BIH Atrial Fibrillation database and PhysioNet/Computing in Cardiology Challenge 2017 database. The model achieved 96.23%, 95.44%, and 95.28% accuracy on the three databases. Compared with the accuracy of the model on training set, when test on independent datasets, the accuracy of the model trained with training strategies only reduced by 2% while the accuracy of the model trained without training strategies reduced by about 15%. Thus, the proposed training strategies significantly enhanced the detection accuracy for the developed deep learning models. It would be one recommendation to train a deep learning based AF detector with good model robustness and generalization.**

***Index Terms*—Atrial fibrillation (AF), electrocardiogram (ECG), deep learning model, wearable ECG.**

I. Introduction

NOWADAYS, smart wearable devices have become a hot topic, especially in the measurement of physiological signals [1] . Development of wearable electrocardiogram (ECG) devices makes the real-time and continuous individual ECG monitoring available [2]. Atrial fibrillation (AF) is a very common type of cardiac arrhythmia, characterized by two features: the absence of P waves and highly irregular variation of R-R intervals [3], [4]. AF may lead to stroke and congestive heart failure (CHF) and increase the death rate for AF patients [5]. Early diagnosis allows better treatment and prevention of secondary diseases like stroke [6]. Its reliable detection is an important target of long-term bio-signal monitoring and is still an unmet challenge even for clinical ambulatory ECG [4].

Jianqing Li (e-mail: ljq@seu.edu.cn) and Chengyu Liu (e-mail: bestlcy@sdu.edu.cn) are co-corresponding authors.

Xiangyu Zhang, Zhipeng Cai, Jianqing Li, and Chengyu Liu are with the School of Instrument Science and Engineering, Southeast University, China.

Li Zhang is with Computational Intelligence Group, Northumbria University, Newcastle upon Tyne, UK

Traditional feature based AF detection methods usually extract features from time-, frequency-, or amplitude-domain [7]–[9]. Features from entropy domain were also used for AF detection [10]–[15]. Deep learning based models with unique hierarchical structure extract deep-level features brought new solutions to the analysis of biological signals [16]–[19]. There is also an increasing trend in the deep learning research for AF detection [20], [21]. Convolutional neural network (CNN) and recurrent neural network (RNN) are two mainstream architectures of supervised deep learning method.

CNN based AF detection models can be simply divided into two categories: 2D CNN model and 1D CNN model. Researchers using 2D CNN based models convert 1D signal to 2D representation. Xia et al. applied short-term Fourier transform (STFT) and stationary wavelet transform (SWT) to obtain 2D matrix input suitable for deep 2D CNN models [22]. Qayyum et al. convert ECG signals into 2D images by STFT, and used pre-trained CNN models for transfer learning [23] . Lorenz plot imaging of ECG RR intervals was also used as input images to training 2D CNN based model for AF classification [4]. 1D CNN models were usually trained using ECG signal as input directly [24], [25].

RNN is another type of network, which architectures allowing the network to retain and utilise state information of the input sequence. Thus, RNN is suitable for processing time series signals [26]. Fausta used LSTM based deep learning model and RR interval signals to detect AF [27]. Chang proposed an AF detection method exploited the spectral and temporal characteristics of AF ECG signals with a multi-lead LSTM model [28]. Maknickas proposed a three-layer LSTM deep learning model using pre-computed QRS complex features for classification [29]. Furthermore, deep learning models combined CNN with RNN were also proposed in past few years. Andersen et al. developed a CNN and LSTM combined deep learning classification model using RR interval as input [30].

All these deep learning methods had achieved good AF detection performance, but whether these models can be applied to dynamic ECG data remains unknown. Most of these deep learning methods use data from one database and test performance by cross-validation method. However, for a single data set, there is no way of knowing how reliable a particular cross validation performance estimate is [31] . In fact, it is hard to evaluate the over-fitting degree of the developed models. So, how to reduce the over-fitting and improve the generalization ability of the model is a very important research [20]. Chang et al. tested their model on separated set and obtained an accuracy of 85%, while the model achieved an accuracy of 98.3% in the training phase [28]. Model proposed by Andersen achieved an accuracy of 97.80% on training data and an accuracy of 87.40% on new recordings [30]. The over-fitting of these models mainly caused by 2 reasons:

a) the similarity between different ECG signals collected from one person, which can lead to the trained model extracted more personal features beyond AF [30].

b) the individual difference between different patients, which was the main reason of the accuracy drops sharply when model test on independent database.

Therefore, in this study, two strategies were explored to improve the robustness and generalization ability of the model. One training strategy was that an FFT and Hanning-window based filter was utilized to remove the individual difference in the ECG waveform. By so doing, ECG data collected from different patients holds a similar property of waveform characteristics after preprocessing. which means the individual waveform difference between patients was eliminated. The other training strategy was that using wearable ECG segments as training data to improve the robustness of the proposed model. Twenty-nine patients were monitored by wearable device for at least 24 hours with no activity restrictions. To verify the effectiveness of the training strategies, an LSTM and CNN based detection model was proposed and test on separate databases. ECG signals from 20 patients were used for training model and ECG signals form the rest patients were used as separate dataset. MIT-BIH Atrial Fibrillation Database (AFDB) and the PhysioNet/CinC Challenge 2017 database (CinC 2017) were also used as separate databases.

II. DATASETS

A. *wearable ECG database I & II*

Wearable ECG data was collected by a wearable ECG monitoring device developed by the authors' team and Lenovo, with a sampling frequency of 400 Hz. The ECG signals was stored in the ECG module and was uploaded to the cloud server via WiFi module. Twenty-nine patients aged 26 to 65 participated in the collection, eleven of them suffered from AF and four of them have a history of premature contractions. Twenty-four hours ECG signal was collected from each patient with no activity restrictions. Then, the collected ECG episodes was labeled to AF class and non-AF class by Cardiologists.

ECGs from 20 patients (10 with AF, 2 with premature contractions and 8 without abnormal heart rhythm) was selected as wearable dataset I. Rest ECGs from 9 patients (1 with AF, 2 with premature contractions and 6 without abnormal heart rhythm) were selected as wearable dataset II.

ECGs in the wearable dataset I was divided into ten fold without overlapping for cross validation. Each fold includes an AF patient and an non-AF patient. Because of the similarity of ECG segments within one person, only 4,000 ECG segments of each patient were randomly picked out. Thus, a total of 80,000 ECG segments were extracted. For the 10-fold cross validation, 72,000 segments were used as training and the remaining 8,000 were used as validation database for the model training. All ECG segments from the 9 patients in wearable dataset II were extracted and were used as the test.

B. *MIT-BIH atrial fibrillation database*

The MIT-BIH atrial fibrillation database (AFDB) consists of 25 long term ECG recordings of human subjects with atrial fibrillation (mostly paroxysmal) [32]. Each recording is 10 h duration, and contain two leads of ECG signals sampled at 250 Hz. The rhythm annotation files were prepared manually; these contain rhythm annotations of types (AFIB (atrial fibrillation), (AFL (atrial flutter), (J (AV junctional rhythm), and (N (used to indicate all other rhythms).

In this study, ECG signals labeled as (J and (N were regard as non-AF ECG data and signals labeled (AFIB were set as AF ECG data. And then the ECG signal parts longer than 10 seconds (4,000 points) were divided into 10 seconds segments with no overlapping, meanwhile, the last 10-second signal of each ECG signal parts was also added to make sure no data missing after dividing. After segmentation, we obtained 33,484 AF segments and 49,980 non-AF segments after segmenting.

C. *The PhysioNet/CinC Challenge 2017 database*

ECG recordings in the PhysioNet/CinC Challenge 2017 database, collected using the AliveCor device and training set contains 8,528 single lead ECG recordings lasting from 9 s to just over 60 s ECG recordings were sampled as 300 Hz and they have been band pass filtered by the AliveCor device [3]. It contains 4 different types of rhythm: AF (atrial fibrillation), normal (normal rhythm), other (used to indicate all other rhythms) and noise.

In this study, noise data was abandoned, while normal and other rhythm data were regard as non-AF class and AF rhythm data was set as AF class. By the way, the ECG signal parts longer than 10 seconds (4,000 points) were also divided into 10 seconds segments with no overlapping and the last 10 seconds data of each recording was added. In addition, the segments was relabeled by Cardiologists. Finally, a total of 1,821 AF segments and 17,399 non-AF segments were obtained.

III. METHOD

A. *preprocessing*

Wearable ECG signals usually contain different kinds of noise, band pass filter was a good method to alleviate the impact of noise. In this work, a Butterworth-bandpass filter with a passband of 0.05-45 Hz was used to remove the baseline drift and reduce the high-frequency noise.

The difference in ECG signals between different people is usually manifested in the following two aspects: The difference in the morphology difference of QRS complex and the overall amplitude range of the signal. FFT and Hanning-window based filter was used to reduce the individual differences of QRS complex. ECG segments were transformed by FFT firstly. Then, the frequency range of 1-45 Hz was intercepted by Hanning window for each FFT-transformed signal, and the Inverse Fast Fourier Transform (IFFT) was performed to obtain the filtered signal. As shown in Fig. 1, the sub-figures in the left column (A, B and C) were three ECG segments with obvious rhythm and waveform differences; the sub-figures in the right column (D, E and F) were the

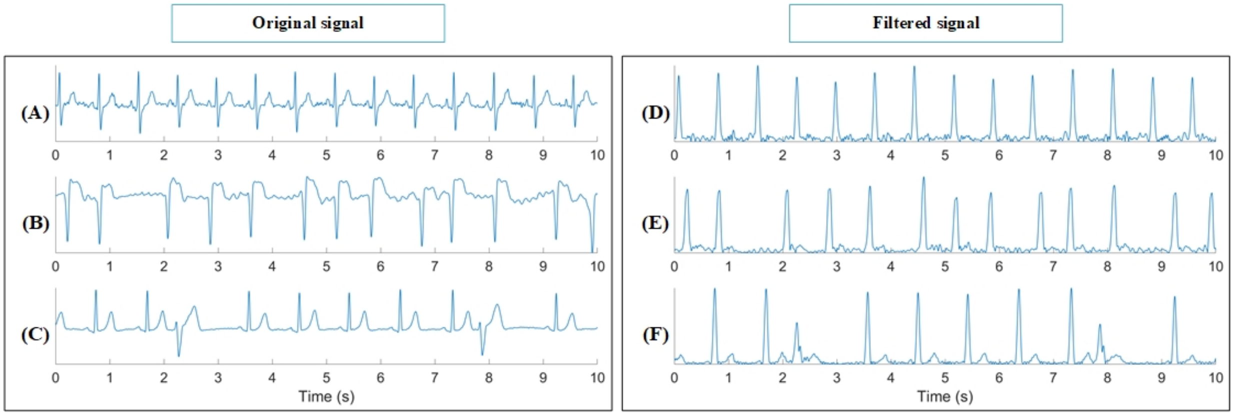

Fig. 1. ECG signals from different people and pre-processed ECG signals. A, B, C are ECG segments from different patients, A was the ECG signal of normal rhythm; B was the ECG signal of patient with atrial fibrillation; C was the ECG signal of patient with premature beats. D, E and F were the corresponding signal after preprocessing.

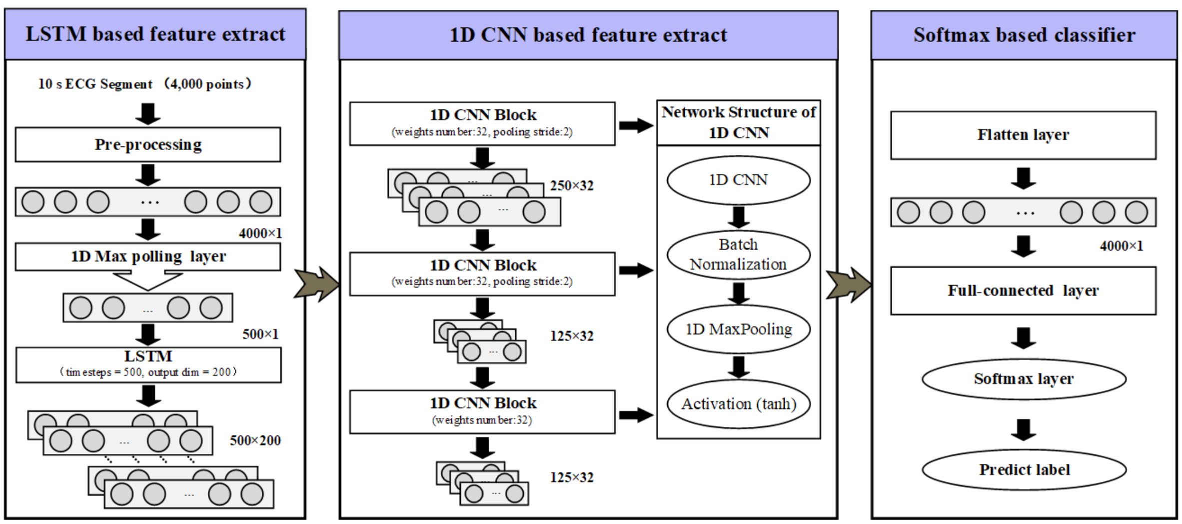

Fig. 2. The proposed classification model based on LSTM and CNN

corresponding ECG segments after eliminating the individual differences in the QRS complex.

The minimum-maximum normalization method was adopted to amplify the filtered ECG segments into the interval [0, 1]. Thereby, the FFT and Hanning-window based filter and normalization method removed the morphology difference of QRS complex and the overall amplitude difference in ECG signals between patients.

B. network structure and optimization method

As showed in Fig. 2, the classification model proposed in this paper was mainly composed of three parts: LSTM module, 1D-CNN module, and Softmax classification module.

The ECG data after preprocessed was firstly passed through a 1D-maxpooling layer with a stride size of 8 to reduce the signal processing time length. By so doing, the input of the LSTM module was transformed to a time series with a length of 500. The output dimension was set to 200, and the output at each time step was recorded as a feature extracted from the input ECG segment.

The 1D CNN module consists of three 1D CNN Blocks, and one CNN block consists of four layers: a 1D convolutional layer, a batch normalization layer, a 1D-maxpooling layer, and an activation layer with activation function of tanh. Each block uses 32 convolution kernels, and the convolution kernel length was 3. The last CNN block did not use 1D-maxpooling layer. The features extracted by 1D CNN were stretched into a feature sequence of 4,000 points by flatten layer. Then the feature sequence was classified to AF or non-AF class by SoftMax layer.

The model used a cross entropy loss function and was trained by Adam optimization method. In order to prevent over-fitting, the batch normalization layer was added after the LSTM layer, and the dropout mechanism was added after each convolution layer and fully-connected layer. Meanwhile, the L2 regularization method of the fully-connected layer was also implied to improve the model generalization ability. Early stopping training method was also used to reduce over-fitting.

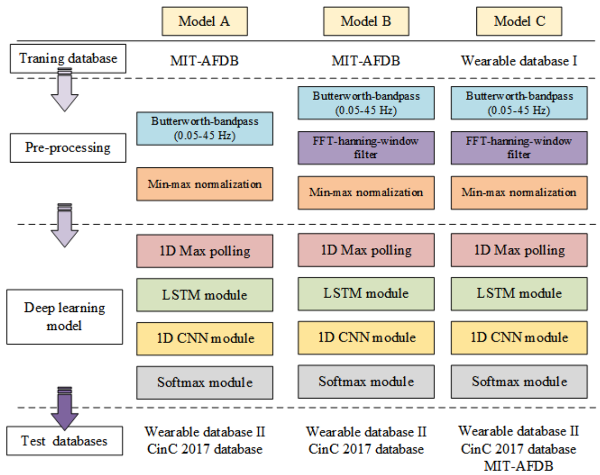

Fig. 3. Training and test flowchart of the three different model

C. *model performance evaluation method*

In order to measure the model performance, the sensitivity (Se), specificity (Sp) and accuracy (Acc) were calculated as:

$$S_e = \frac{TP}{TP + FN} \times 100\% \tag{1}$$

$$S_p = \frac{TN}{TN + FP} \times 100\% \tag{2}$$

$$A_{cc} = \frac{TN + TP}{TP + TN + FP + FN} \times 100\% \tag{3}$$

IV. RESULT

A. *result of model adopting FFT and hanning-window based filter*

To evaluate the effectiveness of the proposed training strategies, three models with same network structure were trained. As is showed in Fig. 3. In order to verify the effectiveness of the FFT and hanning-window based filter, model B was trained with filtered ECG data while model A trained with ECG data which were not filtered. The ten fold validation and test result of the model was showed in Table I.

The average validation accuracy, sensitivity and specificity of model A were 98.91%, 98.50% and 99.42%. And the average validation accuracy, sensitivity and specificity of model B were 98.92%, 98.87% and 99.12%. However, when test on wearable datasets, the test accuracy of model B was 89.99% on the CinC 2017 challenge database and 88.10% on the wearable database II. Model A only achieved an accuracy of 84.67% on the CinC 2017 challenge database and 83.38% on the wearable database II. Compared with the results of cross-validation on the training set, the test accuracy reduced by about 10% for model B and 15% for model A.

B. *result of model using wearable ECG as training data*

In order to verify the effectiveness of using wearable ECG as training data, model B was trained on AFDB while model C was trained on wearable database I. Both two models adopted the FFT and hanning-window based filter in preprocessing. In the test phase, model B was test on CinC 2017 challenge database and wearable database II. Model C was test on wearable database II, AFDB and CinC 2017 challenge database.

Model C achieved an average validation accuracy of 97.71%, sensitivity of 98.13% and specificity of 97.29% in the training database, and also got an excellent performance on test databases. The test accuracy of model C was 96.23% on the CinC 2017 challenge database and 95.44% on the wearable database II. The model C was also tested on AFDB with an accuracy of 95.28%. There was only about 2% decrease in accuracy when model C test on separate databases. The test result shows that although the model C got the worst validation accuracy, model C maintained a good generalization ability.

V. DISCUSSION

A. *effectiveness of training strategies*

In this study, we proposed two training strategies to improve the robustness and the generalization capability of the trained model. One strategy was that a FFT and Hanning-window based filter was utilized to remove the individual difference in the ECG waveform. Model A and B were trained with same ECG segments, but the ECG data of model B were filtered in preprocessing. Model A and B obtained approximate cross validation result on the training database with the average validation accuracy of 98.91% and 98.92%. However, there was a nearly 5% accuracy improvement when adopting the proposed filter in preprocessing. Therefore, the proposed filter plays an important role in overcoming over-fitting.

The other training strategy was that using wearable ECG segments as training data to improve the robustness of the proposed model. Model B and C were trained with same preprocessing method but different ECG segments from different databases. Model C was trained with wearable ECG data while model B was trained by ECG from AFDB. The test accuracy of model C on two wearable datasets were 96.23%, 95.44%, which were close to the validation accuracy. However, the test accuracy of model B on wearable databases II and CinC 2017 database draw-down about 10% compared to the validation accuracy of model B. Therefore, it can be conclude that the model trained with wearable ECG data was more robust than model trained with ECG signals from traditional databases.

B. *classification capability of different deep learning models*

In this work, the proposed LSTM-CNN based model shows a relatively similar result on AFDB compared to other models proposed by Zhou [25], Xia [22], Wang [24], Andersen [30], and Faust [27]. The accuracy of model input with 4 s ECG was 97.4% [24] while the model proposed by Xia achieved the accuracy of 98.29% with input of 5 s ECG [22]. The accuracy of our model with input of 10 s ECG was 98.92%, and the model proposed by Chang [28] achieved 98.50%. The best accuracy of 99% was achieved by the model with input of 30 s ECG [25]. It demonstrates a trend that the accuracy of the model becomes higher when the input length of the model increases.

TABLE I
PERFORMANCE OF THREE DIFFERENT MODELS ON VALIDATION AND TEST DATABASES

models	method	training database	validation performance			test database	test performance		
			S_e (%)	S_p (%)	A_{cc} (%)		S_e (%)	S_p (%)	A_{cc} (%)
model A	LSTM+CNN	AFDB	98.50	99.42	98.91	CinC 2017 database	75.67	85.61	84.67
						wearable database II	92.64	82.37	83.38
model B	Filter+LSTM+CNN	AFDB	98.87	99.12	98.92	CinC 2017 database	79.79	91.06	89.99
						wearable database II	91.40	87.75	88.10
model C	Filter+LSTM+CNN	wearable database I	98.13	97.29	97.71	CinC 2017 database	92.09	96.66	96.23
						wearable database II	97.73	95.19	95.44
						AFDB	96.46	94.49	95.28

TABLE II
THE PERFORMANCE OF METHODS ON INDEPENDENT DATA OR SEPARATE DATABASE

author	method	training database	validation performance			test database	test performance		
			S_e (%)	S_p (%)	A_{cc} (%)		S_e (%)	S_p (%)	A_{cc} (%)
Limam [33]	CNN+LSTM	CinC 2017 (85%)	82.5	98.70	90.60	CinC 2017 (15%)	72.70	98.60	85.60
Aderson [30]	30 RR+LSTM+CNN	AFDB	98.98	96.95	97.80	MITDB NSRDB	98.96	86.04	87.40
Chang [28]	STFT+LSTM	six databases	97.80	99.20	98.50	separate data	86.88	79.55	83.21
						CinC 2017	70.17	–	75.60
Proposed	Filter+LSTM+CNN	wearable database I	98.13	97.29	97.71	CinC 2017	92.09	96.66	96.23
						wearable database II	97.73	95.19	95.44
						AFDB	96.46	94.49	95.28

RR interval based methods proposed by Andersen [30] and Faust [27] with a slightly worse accuracy mainly because of the QRS detection error. The RR interval based models also requires longer ECG segments, which means that these models maybe not suitable for real-time analysis. Compared to these methods of converting ECG segments to spectrograms as 2D matrix, our method used 1D ECG segments as input directly reduced the input size and computational complexity. Compared to other 1D CNN models proposed by Wang [24] and Zhou [25], our model used 1D max-pooling method to reduce the input length. The input size of our model was only 500*1 which contains information of 10s ECG, and achieved a close result to the model proposed by Zhou [25]. Therefore, our model achieved the second best result with the shortest input length.

C. generalization ability of different deep learning models

As shown in Table II, some previously proposed deep learning models were also tested on separate databases or independent ECG data. Limam [33] used 15% ECG signals from CinC 2017 database as independent test data, and the accuracy of the model on the test set is 85.60% which was about 5% lower than the result on training set. Aderson [30] tested model on two independent databases, and the accuracy dropped by about 10%. The model proposed by Chang [28] got an accuracy of 83.21% on separate data from six databases and 75.60% on CinC 2017 database while the model achieved an validation accuracy of 98.50% in training phase. The test results of the model trained with proposed training strategies only reduced by 1% or 2% on separate databases. Thus, the proposed training strategies shows its ability of improving the robustness and generalization ability of the model.

VI. CONCLUSION

In this work, we proposed two training strategies to overcome the over-fitting in deep learning models. FFT and hanning-window based filter was used to remove the individual difference in ECG waveform. Wearable ECG signals which contained complex noise were used to improve the robustness of the proposed model. The model was test on three different ECG database and the result on all test databases reflects that the proposed model maintained good generalization capability when test on separate databases. The comparative experiments also illustrate that the selected training strategies can effectively improve the robustness and generalization ability of deep learning based AF detection models. The selected training strategies would be regarded as one reference method of eliminating over-fitting for future deep learning based AF detection models.

ACKNOWLEDGMENT

This study was supported by the Distinguished Young Scholars of Jiangsu Province (BK20190014), the National Natural Science Foundation of China (81871444), the Primary Research & Development Plan of Jiangsu Province (BE2017735) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX_0068) .

REFERENCES

[1] H. Atoui, J. Fayn, and P. Rubel, "A novel neural-network model for deriving standard 12-lead ecgs from serial three-lead ecgs: application to self-care," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 883–890, 2010.

[2] C. Liu, X. Zhang, L. Zhao, F. Liu, X. Chen, Y. Yao, and J. Li, "Signal quality assessment and lightweight qrs detection for wearable ecg smartvest system," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1363–1374, 2019.

[3] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. Johnson, and R. G. Mark, "AF classification from a short single lead ecg recording: the physionet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[4] J. Hayano, M. Kisohara, Y. Masuda, and E. Yuda, "Detection of paroxysmal atrial fibrillation by lorenz plot imaging of ecg rr intervals," in *International Forum on Medical Imaging in Asia 2019*, F. Lin, H. Fujita, and J. H. Kim, Eds., vol. 11050, International Society for Optics and Photonics. SPIE, 2019, pp. 297–301.

[5] A. M. Gillis, A. D. Krahn, A. C. Skanes, and S. Nattel, "Management of atrial fibrillation in the year 2033: new concepts, tools, and applications leading to personalized medicine," *Canadian Journal of Cardiology*, vol. 29, no. 10, pp. 1141–1146, 2013.

[6] G. D. Giebel and C. Gissel, "Accuracy of mhealth devices for atrial fibrillation screening: Systematic review," *JMIR mHealth and uHealth*, vol. 7, no. 6, p. e13641, 2019.

[7] A. Petrėnas, L. Sörnmo, A. Lukoševičius, and V. Marozas, "Detection of occult paroxysmal atrial fibrillation," *Medical & Biological Engineering & Computing*, vol. 53, no. 4, pp. 287–297, 2015.

[8] C. Bruser, J. Diesel, M. D. Zink, S. Winter, P. Schauerte, and S. Leonhardt, "Automatic detection of atrial fibrillation in cardiac vibration signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 162–171, 2012.

[9] N. Nuryani, B. Harjito, I. Yahya, and A. Lestari, "Atrial fibrillation detection using support vector machine," in *Proceedings of the Joint International Conference on Electric Vehicular Technology and Industrial, Mechanical, Electrical and Chemical Engineering (ICEVT & IMECE)*. IEEE, 2015, pp. 215–218.

[10] M. Kumar, R. B. Pachori, and U. R. Acharya, "Automated diagnosis of atrial fibrillation ecg signals using entropy features extracted from flexible analytic wavelet transform," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 3, pp. 564–573, 2018.

[11] J. Lee, B. A. Reyes, D. D. McManus, O. Maitas, and K. H. Chon, "Atrial fibrillation detection using an iphone 4s," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 203–206, 2012.

[12] C. Liu, J. Oster, E. Reinertsen, Q. Li, L. Zhao, S. Nemati, and G. D. Clifford, "A comparison of entropy approaches for AF discrimination," *Physiological Measurement*, vol. 39, no. 7, p. 074002, 2018.

[13] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, and Y. Zhang, "Automatic online detection of atrial fibrillation based on symbolic dynamics and shannon entropy," *Biomedical Engineering Online*, vol. 13, no. 1, p. 18, 2014.

[14] D. Dharmaprani, L. Dykes, A. D. McGavigan, P. Kuklik, K. Pope, and A. N. Ganesan, "Information theory and atrial fibrillation (AF): A review," *Frontiers in Physiology*, vol. 9, 2018.

[15] W. Xiong, L. Faes, and P. C. Ivanov, "Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of artifacts, nonstationarity, and long-range correlations," *Physical Review E*, vol. 95, no. 6, p. 062114, 2017.

[16] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, p. 65, 2019.

[17] G. Sannino and G. D. Pietro, "A deep learning approach for ecg-based heartbeat classification for arrhythmia detection," *Future Generation Computer Systems*, vol. 86, pp. 446–455, 2018.

[18] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ecg features for screening paroxysmal atrial fibrillation patients," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2095–2104, 2018.

[19] M. Kim and S. B. Pan, "Deep learning based on 1-d ensemble networks using ecg for real-time user recognition," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5656–5663, 2019.

[20] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning: A review," *Journal of Electrocardiology*, vol. 57, pp. S70–S74, 2019.

[21] S. Sahoo, M. Dash, S. Behera, and S. Sabut, "Machine learning approach to detect cardiac arrhythmias in ecg signals: A survey," *IRBM*, 2020.

[22] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Computers in Biology and Medicine*, vol. 93, pp. 84–92, 2018.

[23] A. Qayyum, F. Meriaudeau, and G. C. Chan, "Classification of atrial fibrillation with pre-trained convolutional neural network models," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, 2018, pp. 594–599.

[24] J. Wang, "A deep learning approach for atrial fibrillation signals classification based on convolutional and modified elman neural network," *Future Generation Computer Systems*, vol. 102, pp. 670–679, 2020.

[25] X. Zhou, X. Zhu, K. Nakamura, and M. Noro, "Atrial fibrillation detection using convolutional neural networks," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*. IEEE, 2018, pp. 84–89.

[26] F. Zhu, Y. Fei, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional lstm-cnn generative adversarial network," *Scientific Reports*, vol. 9, 2019.

[27] O. Faust, A. Shenfield, M. Kareem, T. R. San, H. Fujita, and U. R. Acharya, "Automated detection of atrial fibrillation using long short-term memory network with rr interval signals," *Computers in Biology and Medicine*, vol. 102, pp. 327–335, 2018.

[28] Y. Chang, S. Wu, L. Tseng, H. Chao, and C. Ko, "AF detection by exploiting the spectral and temporal characteristics of ecg signals with the lstm model," in *2018 Computing in Cardiology Conference (CinC)*, vol. 45, 2018, pp. 1–4.

[29] V. Maknickas and A. Maknickas, "Atrial fibrillation classification using qrs complex features and lstm," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[30] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Systems with Applications*, vol. 115, pp. 465–473, 2019.

[31] A. Isaksson, M. Wallman, H. Goransson, and M. G. Gustafsson, "Cross-validation and bootstrapping are unreliable in small sample classification," *Pattern Recognition Letters*, vol. 29, no. 14, pp. 1960–1965, 2008.

[32] G. B. Moody and R. R. Mark, "New method for detecting atrial fibrillation using r-r intervals," *Computers in Cardiology*, pp. 227–230, 1983.

[33] M. Limam and F. Precioso, "Atrial fibrillation detection and ecg classification based on convolutional recurrent neural network," in *2017 Computing in Cardiology (CinC)*, 2017.