

# Deep Learning Based Atrial Fibrillation Detection Using Wearable Photoplethysmography Sensor

Alireza Aliamiri<sup>1\*</sup> and Yichen Shen<sup>1</sup>

**Abstract**—Atrial Fibrillation (AFib) is one of the most common cardiac arrhythmia and can potentially progress to serious illness. Early detection and prevention of AFib can bring huge benefits to AFib affected population. With the gain in popularity of wearable devices such as smartwatches, it is possible to utilize the devices build-in photo-plethysmography (PPG) sensor and computational power to provide a portable, non-intrusive and low-cost solution for AFib monitoring and detection to general population. However, utilizing signal data collected from wearable device is challenging due to various types of noise affecting the signal quality. We cope with this challenge by proposing an end-to-end deep learning AFib detection system that can filter out poor quality signals and make reliable predictions. Our models achieves over 95% AUC in quality assessment task and over 99% AUC in AFib detection task while having a reasonable size.

## I. INTRODUCTION

Atrial Fibrillation (AFib) is one of the most common cardiac arrhythmia, with a prevalence of approximately 2% of the general population in developed world [1]. The presence of AFib could potentially lead to major health risks. Traditionally, AFib is detected by electrocardiogram (ECG). While an ECG based approach achieves very high accuracy in AFib detection, monitoring ECG signal requires sophisticated devices that have multiple electrodes installed and the process requires active human participation. Recently, photoplethysmography (PPG) has become a promising alternative to ECG in AFib detection algorithm. The advantage of PPG-based solution is that PPG signals can be easily recorded and monitored from consumer-level wearable devices with no active effort from participants. This advantage, together with affordable wearable devices and smart phones, can make daily-basis, user friendly AFib monitoring and detection available to general public, which may lead to huge beneficial impact on AFib-affected population.

Conventional PPG-based AFib detection uses manual hand-crafted features. With successful applications in computer vision and speech recognition fields[2][3], deep learning has been proved very powerful in feature extraction capability and scalability to larger datasets. Recently deep learning applications in biomedical signal processing and classification has also shown its effectiveness [4][5][6]. In [7], authors managed to utilize convolutional neural network (CNN) as a feature extractor on wavelet transformed PPG signals to predict AFib, however, the proposed method

requires 8 channel PPG data and an intermediate transformation to two dimensional space which makes it computationally expensive, also signal quality assessment is based on periodicity measurement which doesn't consider the intricacy and morphology of the PPG waveform.

In this paper, we would like to introduce a light-weight AFib detection algorithm that directly applies deep neural-networks on raw signals from a sensor with minimum pre-processing or transformation.

Utilizing PPG signals collected from wearable devices imposes many challenges. The signal quality can be drastically affected by quality of sensors and motion from users. The artifact noise from both sensors and user movements can mask any signals useful for detecting AFib. Previous studies on PPG signals collected from wearable devices[6][7] tried to tackle this problem by using handcrafted signal quality index calculated by PPG data or accelerometer data. However in reality those manually engineered features are likely to suffer from high bias and poor generalizability. The unfiltered low quality signal will likely deteriorate the performance in AFib detection in real use cases.

To cope with this challenge, we trained an auxiliary CNN to detect good quality signals from signal segments collected from the sensor, without manually constructing quality measure. This network ensures our detection algorithm is fed with good signal quality and can make sensible predictions. So the workflow of our algorithm is two-fold: first the quality assessment network which acts as a gate keeper and identifies segments of raw PPG signal with good quality suitable for AFib detection; second the AFib detector network, which is another trained neural network, predicts the probability of AFib presence in the provided signals from quality assessment network. Our method achieves high accuracy in AFib detection comparable to state-of-art ECG based approaches, yet having a relatively small structure using only one channel PPG signal.

## II. METHODOLOGY

### A. system overview

The core part of our system consists of a quality assessment network and an AFib detection network. The system takes 30-second segments of the raw PPG and accelerometer signal from a wearable device and after performing pre-processing (see section II-C), it feeds the processed signal to the quality assessment network. Motion significance (see section II-C) is also calculated which is used in subsequent networks. The quality assessment network filters out segments with poor qualities and only sends good quality segments to the

<sup>1</sup>Samsung Strategy and Innovation Center, 3655 N 1st St, San Jose, CA 95134

\*Corresponding author, Email: alireza.a@samsung.com

AFib detector to make reliable predictions. The overall system is shown in Fig. 1.

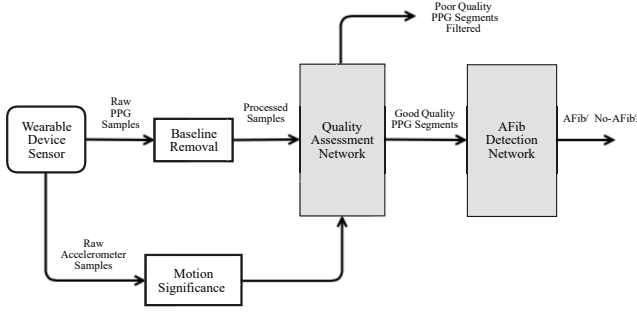


Fig. 1. Overall workflow of the AFib detection System.

### B. Data Collection

PPG data was collected from 19 patients using Samsung gear device. Total of 1443 PPG segments were collected, each PPG segment contained 30 seconds of data with a sampling rate of 100 Hz resulting in 3000 samples per segment. In addition, each PPG segment have accelerometer readings recorded simultaneously at 100Hz. ECG samples were also collected and annotated by cardiologist for presence of AFib episode, this annotation was used as ground truth in our AFib-detection experiments. Among all the PPG samples, 1101 have AFib and 342 had normal sinus rhythm. PPG signal quality was assessed by human expert and a quality index range from 1 (lowest quality) to 10 (highest quality) was assigned to each segment. The quality of signal was assessed based on morphology, periodicity, presence of dirotic notch, consistency of beats, etc. Fig. 2 shows examples of different PPG segments.

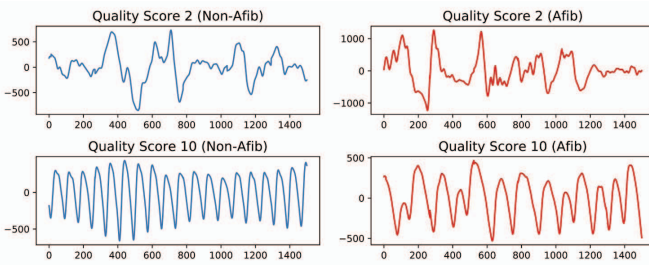


Fig. 2. Examples of PPG signals with different quality scores. Signal segments from left column are labeled as non-AFib and those from right column are labeled as AFib. When quality of signals are low, it is impossible to distinguish AFib signals from those are not AFib.

### C. Data Preprocessing

Wavelet decomposition with 8 levels and Daubechies wavelet was applied on each segment of PPG and approximation channel was removed in wavelet reconstruction to get rid the baseline. A set of motion significance measures are calculated for each segment based on corresponding magnitude (calculated as average of magnitudes in three

axis:  $Acc = \sqrt{x^2 + y^2 + z^2}$ ) of accelerometer signal to capture the degree of motion during the recording of this signal. Specifically, each motion segment contains results in two motion indicator, motion moderate and motion vigorous, which are calculated by thresholding the percentage of standard deviations on half-second window over 30-second PPG segment. The thresholds were tuned by an expert knowledgeable of the device sensor. Furthermore, we marked all the signal segments having quality index higher than 6 to be good quality segments (labeled as 0) and the rest as poor quality segments (labeled as 1), resulting in 716 good quality segments and 727 bad quality segments. We believe this binarization can reduce the human subjectivity.

### D. Quality Assessment Network

A multimodal neural network is designed which accepts two inputs, the preprocessed PPG signal with length of 30 seconds and the vigorous motion measure mentioned above, and outputs the probability of a signal being good quality or not. The motion variable is injected at the last stage along with learned features of PPG from convolutional layers. The model was trained by binary quality labels explained in Section II-C.

The Quality Assessment Network shown in Fig. 3 consists of three convolutional layer with max-pooling, batch-normalization and rectified linear unit (ReLU) as activation function, followed by one fully-connected layer and one output layer. There are 8, 8 and 4 filters with sizes of 32, 8 and 2 along the temporal dimension in first, second and third layer. The pooling size of 1D max pooling layers are 12, 8 and 2 respectively. We do not pad samples for filters. The output from convolutional layers are flattened and concatenated with the motion significance measure for that signal and is fed into one fully connected layer with ReLU activation, followed by the output layer with a sigmoid activation function which outputs the probability of segment having poor-quality.

The model is regularized to prevent overfitting. With L2 constraint imposed on the weights of all convolutional filters and a drop-out regularization on the fully connected layer. The parameter for L2 constraint is 0.05 and rate for dropout is 50%. Total number of parameters of this model is 2,529.

### E. AFib Detection Network

A second neural network is designed which accepts the preprocessed PPG signal that is predicted as good signal by the quality assessment network. The model was trained with AFib annotation by cardiologist on corresponding 30-second ECG signal as ground truth. To effectively extract both local structures of waveforms and their temporal progression as features for predicting AFib, we developed CRNN which is a hybrid model of CNN and recurrent neural network (RNN). We also tested a CNN model without recurrent structures as a baseline.

1) *CNN Baseline Model*: The architecture of the baseline model is a 3-layer CNN very similar to the quality assessment network (see Fig. 3). The difference is in number of

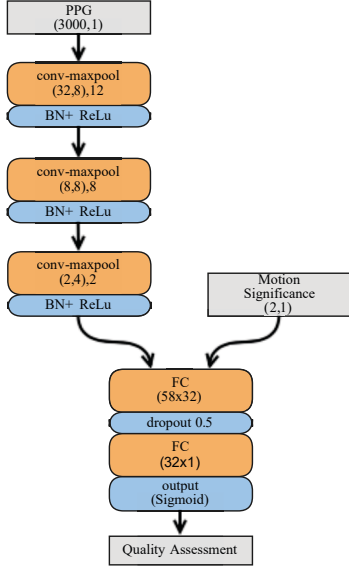


Fig. 3. Quality Assessment Network Architecture

filters in each convolution layer, where this model has 32, 16 and 16 filters. We hypothesized that more filters are needed to extract intricate waveform structures that are invariant in AFib PPG segments. In addition, there is no motion measure concatenated as a second input as there is very few reason to believe that this feature is helpful for predicting AFib. The total number of parameters of this model is 13,121.

2) *Convolution-Recurrent Hybrid Model (CRNN)*: To better model the complex time dependencies of PPG signal (such as irregular patterns between fixed time segments), we propose an RNN structure with signals from consecutive time windows as input. In this model, we split the 30-second segment into 10 three-second-window segments and each segment was fed into a CNN feature extractor GRU cells in time order.

To efficiently extract waveform structures from each time window, we apply a feature extractor containing three convolutional layers to each window. The number of filters are 32, 16 and 16 at the first, second and third layer of CNN. The size of filters along the temporal dimension are 32, 16 and 8 and the size of 1D max pooling layer are 4, 4, and 2 at each level, respectively. We pad no samples for filters. Weights of filters in CNN are shared among all temporal steps, with the premise that features regarding waveform structures are invariant along time dimension. ReLU is used for activation function for all convolutional layers.

The extracted representations of each window were then fed into Gated Recurrent Units (GRU), at each temporal step (3 seconds). The size of hidden layer in each GRU cell is 16. Dropout (50%) is applied on the hidden layer of the recurrent units. The network architecture is depicted in Fig. 4.

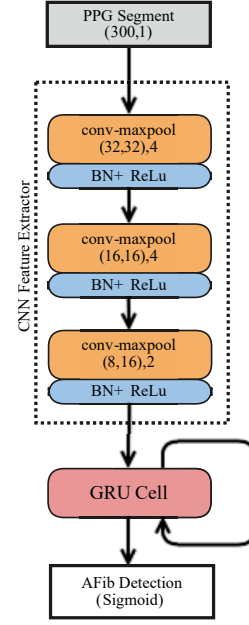


Fig. 4. AFib Detection Network Architecture (CRNN)

### F. Optimization

The weights of each model were optimized by Adam optimizer (learning rate of 0.0005) via back-propagation, using binary cross entropy as loss function. To address the imbalanced class sizes in AFib detection tasks, we attribute a higher class weight to the underrepresented class (segments without AFib) of 0.7 and the other class 0.3 when training AFib detection models. We trained models with mini-batches of 64 examples. We stop training at 150 epochs or whenever the validation error stops improving for 10 consecutive iterations to prevent overfitting.

## III. RESULTS

### A. Quality Assessment

We utilized a five-fold cross validation to assess the performance of our model. We use two metrics to evaluate the performance: mean accuracy (ACC) and pooled area-under-curve (AUC) among test sets in all folds. We compared our model to a baseline model which measures signal quality based on thresholding the motion significance measure. The accuracy is calculated using the threshold having the highest gini-index in quality labels from two separated groups. Fig. 5 and Table I shows the performance comparison of our model and the baseline model.

It is clear that our deep learning model has much higher accuracy and AUC than the baseline model.

### B. AFib Detection

In order to evaluate the AFib detection system independently, we use all signals that are labeled to be good quality (with quality score  $> 6$ ) to train and validate AFib detection

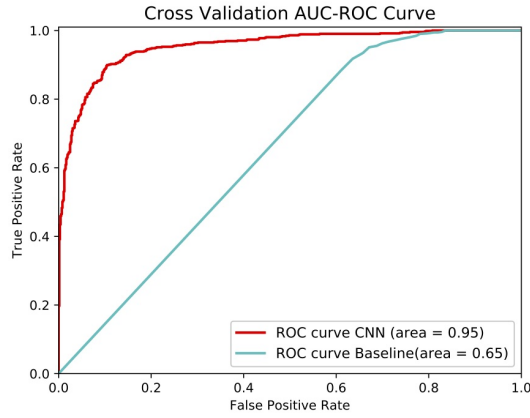


Fig. 5. ROC-AUC comparison of CNN with motion significance measure and rule-based baseline model on quality assessment task

TABLE I

SIGNAL QUALITY ASSESSMENT 5 FOLD CROSS-VALIDATION RESULTS

| Models   | ACC           | AUC           |
|----------|---------------|---------------|
| Baseline | 63.75%        | 65.09%        |
| CNN      | <b>90.02%</b> | <b>95.21%</b> |

models. This will result in 557 AFib segments and 159 non-AFib segments. In order to train more robust models, we attributed a higher weight to the underrepresented class in loss functions of the models.

We test two models, including above mentioned two architectures: CNN and CRNN. Both models have accuracy higher than 97% and AUC higher than 99%. The hybrid model achieved best performance overall, showing that modeling time-dependency information is helpful for predicting AFib. The result shows that a relatively small network structure with number of parameters in 10,000 range can achieve very high performance. The cross validation results are summarized in Table II.

TABLE II

AFIB DETECTION 5 FOLD CROSS-VALIDATION RESULTS.

| Models                | ACC           | AUC           | # of parameters |
|-----------------------|---------------|---------------|-----------------|
| CNN-base              | 97.20%        | 99.34%        | <b>13,121</b>   |
| <b>CNN-RNN-Hybrid</b> | <b>98.19%</b> | <b>99.67%</b> | 14,674          |

#### IV. CONCLUSION

Our deep learning AFib detection system using PPG signal showed very high performance, with minimum manual feature engineering. By training a quality assessment network, the system is capable of accurately filtering out bad signals and thus ensures robust prediction by the AFib detection algorithm. This will be very effective when signals come from a wearable device and are subject to large variation in quality. Moreover, we observe that even using a small convolutional network can predict AFib with high accuracy.

Due to limitation of our resources, we only have PPG signals collected from 19 patients, which is relatively small comparing to similar studies. Future studies should look for ways to cope with dearth of both data and labels. In addition, training the model on a large dataset will validate the proposed algorithm on a larger scale.

#### REFERENCES

- [1] A. J. Camm, G. Y. Lip, R. De Caterina, I. Savelieva, D. Atar, S. H. Hohnloser, G. Hindricks, P. Kirchhof, J. J. Bax, H. Baumgartner *et al.*, "2012 focused update of the esc guidelines for the management of atrial fibrillation: an update of the 2010 esc guidelines for the management of atrial fibrillation developed with the special contribution of the european heart rhythm association," *European heart journal*, vol. 33, no. 21, pp. 2719–2747, 2012.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [3] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [4] Y. Ren and Y. Wu, "Convolutional deep belief networks for feature extraction of eeg signal," in *Neural Networks (IJCNN), 2014 International Joint Conference on*. IEEE, 2014, pp. 2850–2853.
- [5] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.
- [6] S. Nemati, M. M. Ghassemi, V. Ambai, N. Isakadze, O. Levantsevych, A. Shah, and G. D. Clifford, "Monitoring and detecting atrial fibrillation using wearable technology," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016, pp. 3394–3397.
- [7] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on*. IEEE, 2017, pp. 141–144.