



# Prediction to Atrial Fibrillation Using Deep Convolutional Neural Networks

Jungrae Cho<sup>1</sup>, Yoonnyun Kim<sup>2</sup>, and Minho Lee<sup>1</sup>(✉)

<sup>1</sup> School of Electronics Engineering, Kyungpook National University,  
Daegu, Republic of Korea

zzemb6@gmail.com, mhlee@gmail.com

<sup>2</sup> Department of Internal Medicine (Cardiology),  
Keimyung University Dongsan Medical Center, Daegu, Republic of Korea  
ynkim@dsmc.or.kr

**Abstract.** Recently, many researchers have attempted to apply deep neural networks to detect Atrial Fibrillation (AF). In this paper, we propose an approach for prediction of AF instead of detection using Deep Convolutional Neural Networks (DCNN). This is done by classifying electrocardiogram (ECG) before AF into normal and abnormal states, which is hard for the cardiologists to distinguish from the normal sinus rhythm. ECG is transformed into spectrogram and trained using VGG16 networks to predict normal and abnormal signals. By changing the time length of abnormal signals and making up their own datasets for preprocessing, we investigate the changes in F1-score for each dataset to explore the right time to alert the occurrence of AF.

**Keywords:** Atrial fibrillation · Arrhythmia · Deep neural networks  
Convolutional neural networks

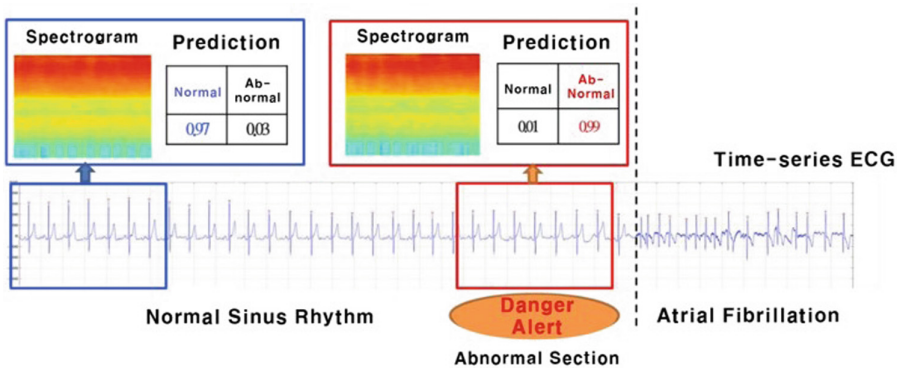
## 1 Introduction

Atrial Fibrillation (AF) is one of the common cardiac arrhythmia in patients, which is usually accompanied by other serious symptoms such as stroke [1]. This irregular heart rhythm also increases the chance of occurrence of heart failures and mortality in arrhythmia patients [2]. Then, it is crucial to detect and predict AF as early as possible. Many machine learning approaches have been used to classify normal sinus rhythm and cardiac arrhythmias from electrocardiogram (ECG) [3–5]. Recently, deep learning has reported its successful contributions to various areas such as image classification [6–9]. Many researchers have applied deep neural networks to monitor the occurrence of AF. Among which, convolutional neural networks (CNN) and recurrent neural networks (RNN) have been popular in feature extraction to detect AF and other arrhythmias [10–13]. However, they have been focused on detection and not prediction of AF. A few researches have challenged to predict AF before it happens using machine learning and neural networks [14, 15]. In this paper, we propose a new AF prediction algorithm to explore the prelude of AF that is difficult for the cardiologists to identify using Deep Convolutional Neural Networks (DCNN). The ECG signals before AF are divided into normal and abnormal signals, and the time length of each of these

two classes are different. Then, we train VGG16 networks [16] and measure F1-scores for each different label case to check the dynamics of ECG before AF.

## 2 Scenario of Prediction to AF

Figure 1 shows the scenario of predicting normal and abnormal states from ECG. To predict prelude of AF, we divide the ECG signals (before the occurrence of AF) into normal signal and abnormal signal. Normal signal is the same as regular sinus rhythm, however, abnormal signal is difficult to be distinguished from normal signal with human eye. The goal of the scenario is to alert arrhythmia patients 4–5 min ahead the possibility of occurrence of AF by monitoring their ECG continuously.



**Fig. 1.** Scenario for prediction of AF. ECG signal is divided into normal and abnormal signals before occurrence of AF. DCNN learns and predicts if the signal is in normal state or in abnormal state.

## 3 Data Preprocessing

### 3.1 Dataset

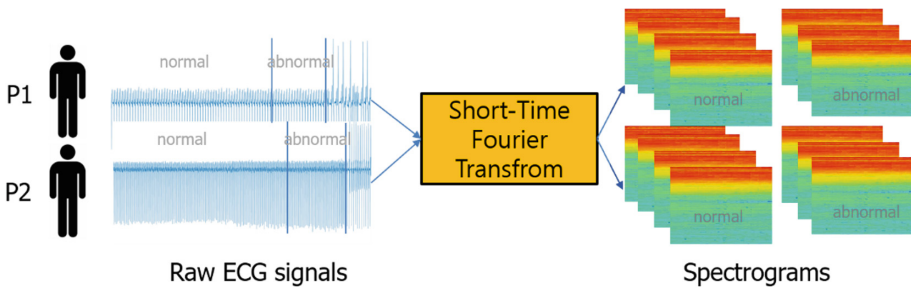
We use single-lead ECG dataset provided by Keimyung University Dongsan Medical Center (KUDMC), which is private and anonymized. We use ECG signals that are about 10 min long to predict normal and abnormal signals before AF to predict normal and abnormal signals as mentioned in Fig. 1. We choose to train each patient's data separately because hemodynamic response i.e. the average rhythm and characteristic of the heart beat is unique to each patient [2]. So, the dataset for patient-dependent model requires long sequences of each patient which contains both normal sinus rhythm and AF. From the restricted condition, we choose ECG signals of three patients because ECG records have very few cases of the continuation of normal and AF signals. To provide more experiments, we additionally use two records of paroxysmal AF (PAF) from PhysioNet dataset [15, 17]. Table 1 illustrates the number of obtained spectrograms from ECG signals, duration of ECG, and data source.

**Table 1.** Dataset information

Patients	# of spectrograms	Duration of ECG [minutes]	Source of data
Patient 1	1879	23	KUDMC
Patient 2	272	9	KUDMC
Patient 3	363	10	KUDMC
Patient 4	299	30	PhysioNet
Patient 5	299	30	PhysioNet

### 3.2 Preprocessing

Figure 2 shows the preprocessing of ECG signals. Since ECG is a time-series data, we transform ECG into spectrograms to be used as the input for CNN. Spectrograms are generated using short-time Fourier transform of every 30 s ECG signals with an overlap of 1 s [14].



**Fig. 2.** Preprocessing of the ECG signals convert to spectrograms. P1, P2 are patients. From their ECG records, we divide normal and abnormal periods. The divided ECG signals are transformed in to spectrograms to use them as input to CNN. Each patient's dataset is used to train its own subject specific model.

### 3.3 VGG16

To classify very similar but different signals, i.e. normal and abnormal states, we utilizes DCNN because it is considered as the powerful algorithm in image processing field for its successful performance to image classification in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [7]. VGG16 network is one of the DCNN which consists of 13 convolution and pooling layers and 3 fully-connected layers [16]. In this paper, VGG16 is considered to predict normal and abnormal states since it has a deep architecture with good performance for image classification. We use VGG16 networks implemented with Keras [18]. To overcome the paucity of data samples, pretrained weights from ImageNet [7] are considered and fine-tuned for normal and abnormal classes. The architecture of VGG16 is as shown in Fig. 3.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 128, 128, 3)	0
block1_conv1 (Conv2D)	(None, 128, 128, 64)	1792
block1_conv2 (Conv2D)	(None, 128, 128, 64)	36928
block1_pool (MaxPooling2D)	(None, 64, 64, 64)	0
block2_conv1 (Conv2D)	(None, 64, 64, 128)	73856
block2_conv2 (Conv2D)	(None, 64, 64, 128)	147584
block2_pool (MaxPooling2D)	(None, 32, 32, 128)	0
block3_conv1 (Conv2D)	(None, 32, 32, 256)	295168
block3_conv2 (Conv2D)	(None, 32, 32, 256)	590080
block3_conv3 (Conv2D)	(None, 32, 32, 256)	590080
block3_pool (MaxPooling2D)	(None, 16, 16, 256)	0
block4_conv1 (Conv2D)	(None, 16, 16, 512)	1180160
block4_conv2 (Conv2D)	(None, 16, 16, 512)	2359808
block4_conv3 (Conv2D)	(None, 16, 16, 512)	2359808
block4_pool (MaxPooling2D)	(None, 8, 8, 512)	0
block5_conv1 (Conv2D)	(None, 8, 8, 512)	2359808
block5_conv2 (Conv2D)	(None, 8, 8, 512)	2359808
block5_conv3 (Conv2D)	(None, 8, 8, 512)	2359808
block5_pool (MaxPooling2D)	(None, 4, 4, 512)	0
global_average_pooling2d_1 (	(None, 512)	0
dropout_1 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 128)	65664
dropout_2 (Dropout)	(None, 128)	0
batch_normalization_1 (Batch	(None, 128)	512
dense_2 (Dense)	(None, 2)	258
Total params: 14,781,122		
Trainable params: 14,780,866		
Non-trainable params: 256		

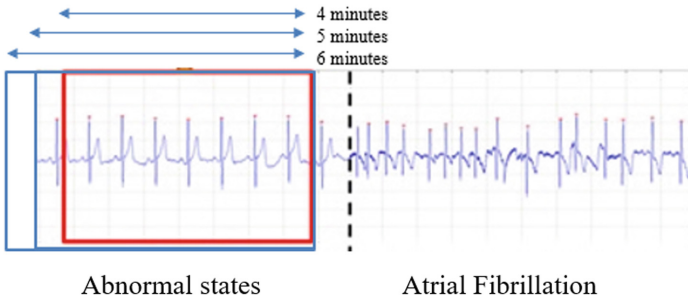
**Fig. 3.** Architecture of VGG16 network. To solve AF prediction problem, we modified fully connected layers and the dimension of the output layer.

### 3.4 Training Configuration

The input of VGG16 is spectrograms obtained from ECG of size  $256 \times 256$ . The output is two classes with normal ([1 0]) and abnormal ([0 1]) signals. Optimizer is ADAM [19]. We applied dropout [20] and batch normalization [21] to fully-connected layers to stabilize training and validation loss.

## 4 Experiments

Not only to predict normal and abnormal ECG signals but also to find proper time to alert the occurrence of AF, we make up abnormal states as 4, 5, and 6 min before AF. By varying abnormal periods, we obtain 3 cases of datasets for each patient as shown in Fig. 4.



**Fig. 4.** Varying the length of abnormal states to explore proper alert time for AF.

We train VGG16 network for each dataset and every patient. We measure the F1-scores after the training process and test the accuracies to investigate the discrimination of each case of abnormal states. Each F1-score is the average of F1-scores calculated from three times experiments with the same training and test set. We additionally prepare several baseline models to compare them to VGG16. Since Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) are not suitable for high dimensional data i.e. spectrograms, we consider standard CNN and Long-Short Term Memory (LSTM) networks [22] as baseline models. The simple CNN consists of two convolution layers, one max pooling layer, and fully connected layers. The LSTM model has 1 LSTM cell and fully connected layers.

### 4.1 Model Performance

For each case of abnormal state, we measure F1-scores of each model as shown in Table 2. VGG16 reports higher F1-scores compared to standard CNN and LSTM. LSTM fails to learn the data for patient 3, 4 and 5 with low F1-scores as mentioned in Table 2, whereas standard CNN and VGG16 converge their train losses. The better F1-scores of VGG16 shows that the deeper CNN architecture is good for learning our datasets.

As shown in the table, every patient shows different F1-scores. The diversity in dynamics of F1-score is considered since patients can have their own hemodynamic consequences in their ECG records [2]. The lower F1-score indicates that normal and abnormal states are hard to believe the prediction results by the DCNN. Whereas, the higher F1-score indicates that normal and abnormal states are easy to distinguish with high reliability. In Fig. 4, Patient 1 and 5 show a monotonic decrease of F1-scores.

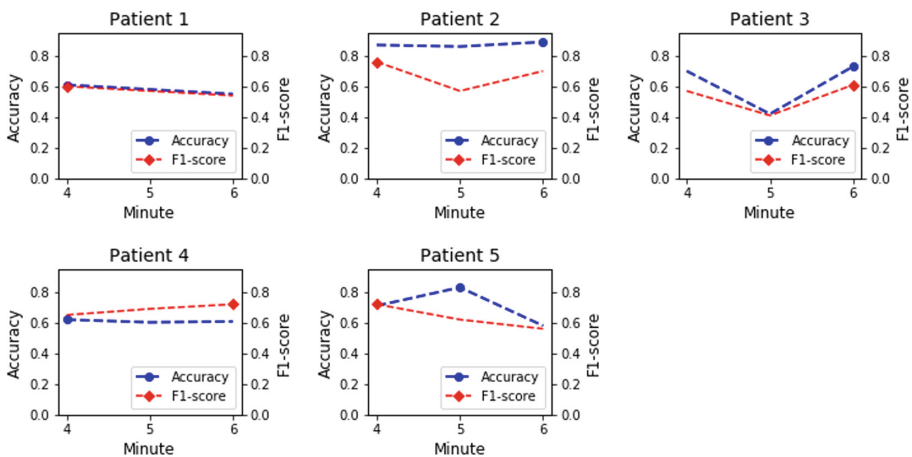
It indicates that the 4-min case has higher discrimination compared to 5 and 6-min cases. Patients 4 is the opposite case of Patient 1 and 5, which means that the prediction accuracy is more believable through time. For Patients 2 and 3, the prediction of AF is more believable in 4 and 6 min. Those results show that the change in F1-scores implies the relative reliability of prediction results for existence of abnormal state before AF happens.

**Table 2.** F1-scores of VGG16 and baseline models (standard CNN and LSTM) for various case of abnormal states

Model	Case of Abnormal states [minutes]	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
VGG16	4	<b>0.6</b>	<b>0.76</b>	0.57	0.65	<b>0.72</b>
	5	0.57	0.57	0.41	0.69	0.62
	6	0.54	0.70	<b>0.61</b>	<b>0.72</b>	0.56
CNN	4	<b>0.5</b>	0.56	0.36	0.42	<b>0.57</b>
	5	0.4	0.54	<b>0.46</b>	<b>0.55</b>	0.53
	6	0.44	<b>0.59</b>	0.39	0.4	0.42
LSTM	4	0.46	<b>0.6</b>	0.33	0.33	0.33
	5	0.36	0.52	<b>0.38</b>	0.33	0.33
	6	<b>0.6</b>	0.49	0.33	0.33	0.33

### 4.2 Application to Alert the Occurrence of AF

Figure 5 shows the change of test accuracies and F1-scores for datasets which have different length of abnormal states for each patient. In each figure, the markers indicate the highest accuracy and F1-score to predict test data as normal or abnormal signals.



**Fig. 5.** Change of test accuracies and F1-scores for 4, 5, 6 minute-dataset for patients. The Markers in each figure indicate the highest accuracy and F1-score for its abnormal section.

Since F1-score considers both precision and recall unlike the accuracy, the dynamics between accuracies and F1-scores looks different for each patient as shown in Fig. 5. By considering the abnormal periods from the highest accuracy verified by its F1-score as the confidence for the prelude of AF, it has chance for application to alert the occurrence of AF to both doctors and patients to determine if AF could happen.

## 5 Conclusion and Future Works

Prediction of AF is a crucial task to save the patient's life because AF could lead to fatal diseases. In this paper, we attempted to predict the prelude of AF using DCNN. We trained VGG16 network to predict normal sinus rhythm and abnormal signals, and measured F1-scores for different length of abnormal states. The F1-score range of patients showed that it had subject specific discriminant patterns. DCNN suggests that there are some abnormal signals before AF that are difficult to distinguish from normal signals.

For future work, it is possible to analyze wider range of abnormal sections to explore F1-score dynamics and additional experiments with more patients can be considered. Also, we are considering a new DCNN learning algorithm to detect very sensitive variation of AF signals in normal and abnormal conditions. Finally, the research to automate the exploration of proper alert time is recommended.

**Acknowledgements.** This work was partly supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding) (30%), Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2018-2-00861, Intelligent SW Technology Development for Medical Data Analysis) (40%) and the Ministry of Trade, Industry & Energy (MOTIE, Korea) under Industrial Technology Innovation Program. No.1063553, 'Self-directed portable safety kits and application based living environment service system' (30%).

## References

1. Wolf, P.A., Abbott, R.D., Kannel, W.B.: Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* **22**(8), 983–988 (1991)
2. Dries, D., Exner, D., Gersh, B., Domanski, M., Waclawiw, M., Stevenson, L.: Atrial fibrillation is associated with an increased risk for mortality and heart failure progression in patients with asymptomatic and symptomatic left ventricular systolic dysfunction: a retrospective analysis of the SOLVD trials. *J. Am. Coll. Cardiol.* **32**(3), 695–703 (1998)
3. Polat, K., Güneş, S.: Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine. *Appl. Math. Comput.* **186**(1), 898–906 (2007)
4. Bost, R., Popa, R.A., Tu, S., Goldwasser, S.: Machine learning classification over encrypted data. In: NDSS, p. 4325 (2015)
5. Li, Q., Rajagopalan, C., Clifford, G.D.: Ventricular fibrillation and tachycardia classification using a machine learning approach. *IEEE Trans. Biomed. Eng.* **61**(6), 1607–1613 (2014)

6. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436 (2015)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
8. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision* **115**(3), 211–252 (2015)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
10. Rajpurkar, P., Hannun, A.Y., Haghpanahi, M., Bourn, C., Ng, A.Y.: Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:170701836* (2017)
11. Poh, M.-Z., et al.: Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms. *Heart:heartjnl-2018-313147* (2018)
12. Oh, S.L., Ng, E.Y., San Tan, R., Acharya, U.R.: Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput. Biol. Med.* (2018)
13. Xia, Y., Wulan, N., Wang, K., Zhang, H.: Detecting atrial fibrillation by deep convolutional neural networks. *Comput. Biol. Med.* **93**, 84–92 (2018)
14. Kim, J., Sangjun, O., Kim, Y., Lee, M.: Convolutional neural network with biologically inspired retinal structure. *Procedia Comput. Sci.* **88**, 145–154 (2016)
15. Bianchi, F.M., Livi, L., Ferrante, A., Milosevic, J., Malek, M.: Time series kernel similarities for predicting Paroxysmal Atrial Fibrillation from ECGs. *arXiv preprint arXiv:180106845* (2018)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556* (2014)
17. PhysioNet: The PAF Prediction Challenge Database (2001). <https://physionet.org/physiobank/database/afpdb/>
18. Chollet, F.: Keras (2015)
19. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXiv:14126980* (2014)
20. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
21. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:150203167* (2015)
22. Sak, H., Senior, A., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: *Fifteenth Annual Conference of the International Speech Communication Association* (2014)