

Classification of Atrial Fibrillation with Pre-Trained Convolutional Neural Network Models

Abdul Qayyum

*Centre for Intelligent Signal and Imaging
Research (CISIR), Universiti Teknologi
PETRONAS, Perak, Malaysia*

engr.qayyum@gmail.com

Fabrice Meriaudeau

*Centre for Intelligent Signal and Imaging
Research (CISIR), Universiti Teknologi
PETRONAS, Perak, Malaysia*

Genevieve C Y Chan

*Centre for Intelligent Signal and Imaging
Research (CISIR), Universiti Teknologi
PETRONAS, Perak, Malaysia*

Abstract— Atrial Fibrillation (AF) is the most common chronic arrhythmia. Effective detection of the AF would avoid serious consequences like stroke. Conventional AF detection methods need heuristic or hand-crafted features extraction. Recently, deep learning (DL) techniques with massive data have been used on image, voice and other field widely with impressive results. The ECG rhythms such as AF, normal, Noisy and other rhythms have been segmented into 3 segments per rhythm, converted into 2D images using short time Fourier transform (STFT) and fed into pre-trained models. The pre-trained CNN models are used for transfer learning or are fine tuned for the detection and classification of the AF rhythm. The features extracted from the last layer of the pre-trained models are used as input to classical classification algorithms such as Ensemble classifier and support vector machine (SVM) for AF detection. The proposed approach would have great potential on real-time monitoring of atrial fibrillation signal in electrocardiogram. Overall, our approaches achieved accuracy, sensitivity and specificity as of 97.89%, 97.12% and 96.99% similar to the latest state of the art techniques but with more flexibility.

Keywords— ECG, AF detection, CNN models, Feature extraction, Deep learning.

I. INTRODUCTION

Atrial fibrillation (AF) is the most common cardiac arrhythmia, with an occurrence of 1.5 to 2 percent in the developed countries [1]. 2.2 million people in the United States are affected with six to eight percent affecting people [3] over 65 years old [2], whereas other developed countries like the European Union [3] has 4.5 million people suffering from AF. The presence of AF indicates three-fold risk of developing heart failure (HF) [4] and five-fold risk of having stroke [8]. The extensive utilization of health resources and high hospitalization rates due to AF prediction may increase day by day and impose a great economic and clinic burden on society [5]. Different sensors are used for continuous screening and monitoring of AF. These methods are based on photoplethysmograph (PPG) and ECG signals. Generally, based on ECG, the AF signal wave is very different from the normal heart rhythm signal [2]. A simple, cost effective, portable and continuous monitoring system requires continuous screening and monitoring of AF due to large number of patients. A new AF detection technology

would reduce the mortality rate and morbidity and it could also be used to assist for clinical treatment of the AF patients. Well-trained professionals with sufficient skills and knowledge are required to interpret the ECG signal accurately. Therefore, to lower the cost and increase the efficiency, an effective, reliable and accurate AF diagnosis solution based on algorithms is needed.

Currently, various algorithms used in AF detection has been proposed and these are relying on the characteristics of the ECG signal. These characteristics are 1) irregularity of R-R intervals and absence of P-waves. In RR interval based approach, first R waves are identified and the irregularity between RR intervals from the ECG signal is checked. Jie Lian presented RR interval as Markov process and computed likelihood of RR pairs for AF identification [6]. Logan et. al. proposed variance of RR intervals [7] for AF detection. Tatento and Glass proposed Kolmogorov Smirnov test of RR intervals [8] for AF detection and identification. Cerutti et. al. presented autoregressive moving average (ARMA) model to extract time-variant features to identify the difference between AF and normal sinus rhythm.

They also used corrected conditional entropy [9] based on non-linear features for AF detection. Sakar et. al. extracted features based on RR intervals using Lorenz distribution [10]. Lake and Moorman proposed nonlinear features such as sample entropy and its variants extracted from RR intervals to detect AF [11]. Lee et. al. used coherence features to check similarity between adjacent segments of RR intervals to identify the difference between AF and normal sinus rhythm [12]. Petrenase et. al. proposed an hybrid method for AF detection which provides high efficient AF detector with low complexity [13]. The methods uses R-R interval can be identify AF episodes for the data segment of length (50-100 beats). The limitation come arises for short AF episodes usually less than a minute.

Consequently, the performance of these methods depends on the detection of P or R peak. If the peaks are detected mistakenly or missed, the performance of these methods may decrease or degrade. Also the noise present in the ECG signal may generate false detections of P-waves [13] and lower performances [14]. Slocum used the distribution of power of waveform analysis for detection of AF and identify P-wave to check the presence of AF [14]. The features extracted from f-wave and P-wave may

be not feasible as compared to the R peak features and some induced noise produced similar behaviour in f-wave and P-wave which create disturbance; it is a big hurdle for AF detection from ECG signal. Asgari and others [15] presented different approach for AF detection that did not rely on P-peak or R-Peak detection method. They used peak-to-average power ratio and log-energy entropy extracted from wavelet transform for AF detection. These hand-crafted features do not provide optimal characteristics of the signal. Moreover, these handcrafted feature could not be used broadly in clinical application due to challenges of extracting reliable features from the ECG signal and the accuracy may differ significantly based on large data sets [16].

Nowadays, deep learning based feature extraction and detection methods have been extensively used for automatic feature extraction and classification in various domains [17]. Deep learning-based techniques have been used in a small scale in ECG signal classification due to small training dataset and specificity of ECGs. The authors used the Convolutional Neural Networks for the supraventricular ectopic beats [18] and ventricular ectopic beats. Other authors used unsupervised auto-encoder method for ECG signal enhancement [19]. However, the performance obtained were not satisfactory as compared to speech recognition and object detection approaches based on deep learning.

In this paper, deep learning approaches based on pre-trained models have been used to address the limitations in existing methods (P-wave and R-wave) and avoid the necessity of relying on manual feature extraction and classification methods. A deep neural network is capable of extracting features from the original input data, it is robust to variances like translation, rotation, and noise. The short time Fourier transform is used to convert 1D signals into 2D images that are then fed into pretrained models for detection and classification of AF rhythm.

II. MATERIAL AND METHODS

ECGs are used to record the electrical activity of a person's heart for a certain time period. Physicians use ECG signals to visually check the heartbeat of person whether it is regular or irregular. The dataset PhysioNet/Computing in Cardiology [20] used in this study, has been divided into four classes denoted as Normal (N), Atrial Fibrillation (AF), Other Rhythm (O), and Noisy Recording (NI). The ECG AliveCor device was used to collect this dataset. The total number of training samples are 4450 samples using single ECG recording for each class. The ECG recorded at 300 Hz samples using band pass filtering method based on the AliveCor device [20]. Fig.1 shows the ECG rhythms based on normal, AF, noisy and other recordings. These number of ECG waves has been segmented using 1 second and we extracted 15 seconds in each ECG wave based on four classes.

Each sample has been converted into time-frequency domain using short time Fourier transform technique (see

Fig. 3) and fed into the pre-trained CNN models for classifying AF and other ECG wave rhythms. Each sample of all classes contains 15 segments and one segment represent one image. In our case, each sample produced (15x4450) 66750 images per class. We used only the first three segment for each class in this study. Thus 13350 (3*4450) images for each class. The evaluation is done on a five folds cross validation (80/20). The complete process diagram is shown in Fig. 2. The pre-trained models are AlexNet, VGG16, VGG19, GoogLeNet and ResNet and they are used either in transfer learning (as feature extractors) or in fine tuning mode by removing the last FC layers and changing it to four classes. For transfer learning, the extracted features are fed into classical classifiers such as SVM and ensemble method.

A. PreTrained Models as Transfer Learning

The major breakthrough happened in the field of deep learning in 2012, when the AlexNet deep CNN model invented using ImageNet dataset and won the Large-Scale Visual Recognition Challenge (ILSVRC) with excellent results as compared to other models. In 2013, ZFNet [21] won the ILSVRC competition by optimizing the various parameters using a similar AlexNet architecture. In 2014, two CNN models were proposed, GoogLeNet [22] that consisted of 22 number of layers with inception module and VGG16 model which has simple architecture and excellent performance on the ImageNet dataset. ResNet, with residual blocks [23] and 152 layers won this challenge in 2015. Since then, various models including the inception module and the residual block have been proposed.

The training parameters used for fine-tuning are the initial learning rate, number of epochs, mini-batch size and L2 regularization. The stochastic gradient descent algorithm is used for optimization. The initial learning rate used was 0.001, learning drop rate period 8 and L2 regularization was 0.004. The total number of 100 epochs and mini-batch-size was 32 have been used in our design experiment. The NVIDIA GPU used in experiment has capacity 3GB RAM with K-4000 series.

B. Performance Metrics

The accuracy, sensitivity and specificity have been used as a performance metrics to evaluate CNN models efficiency. These performance metrics are shown in (1-3).

$$\text{Sensitivity} = TP / (TP + FN) \quad (1)$$

$$\text{specificity} = TN / (TN + FP) \quad (2)$$

$$\text{accuracy} = (TP+TN) / (TP+TN+FP+FN) \quad (3)$$

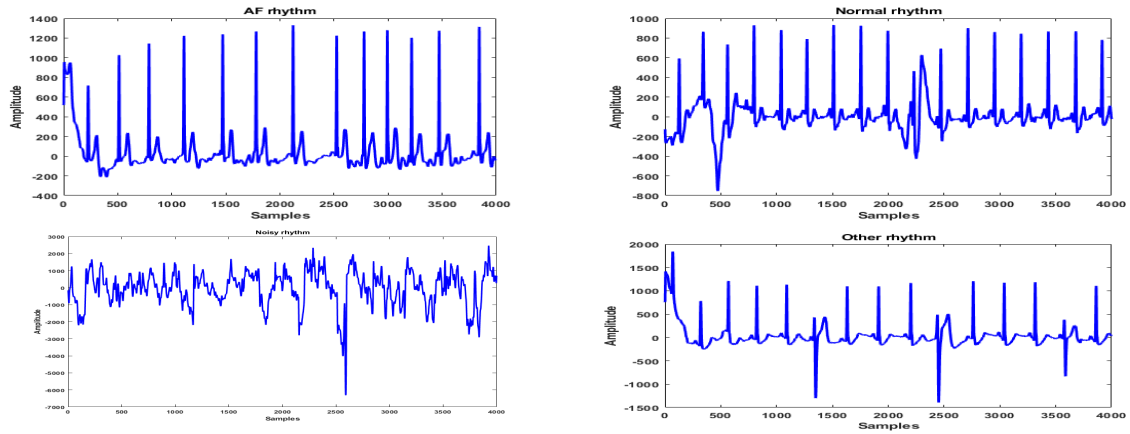


Fig. 1. The ECG dataset have four number of classes (AF, Normal, Noisy and Other rhythms).

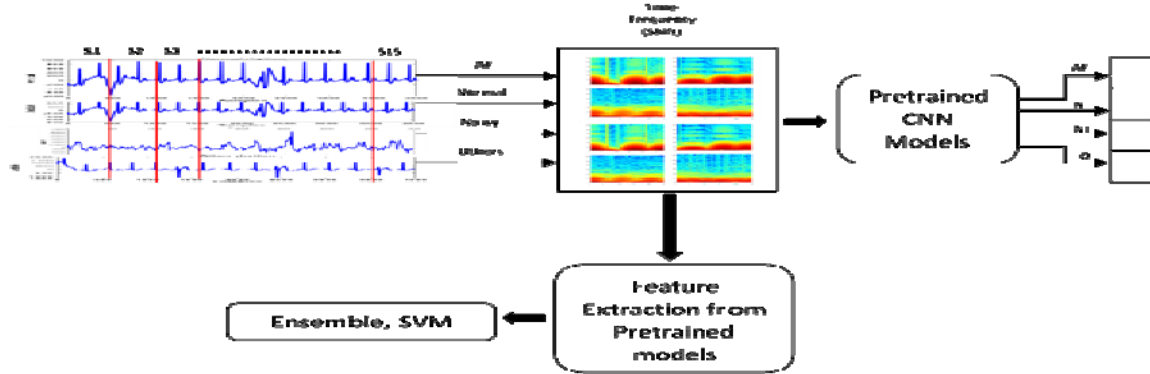


Fig. 2. The proposed methodology for classifying ECG signal based on pretrained models used as feature extraction and fine-tuning approach.

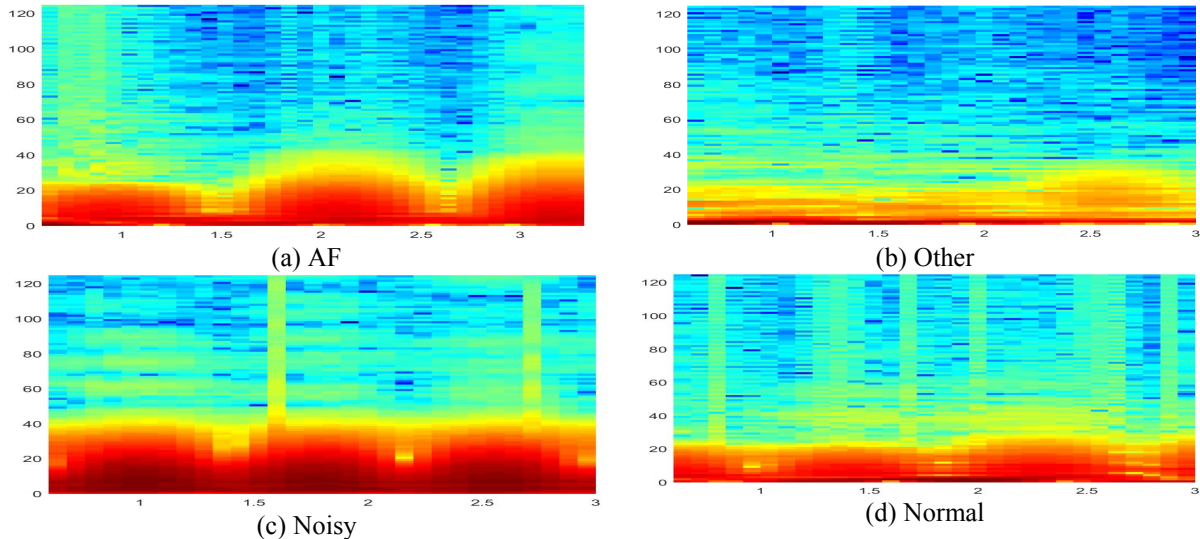


Fig. 3. Time-frequency analysis for (a) AF, (b) Other, (c) Noisy and (d) Normal rhythms.

III. RESULTS AND ANALYSIS

The approach used for validation is based on a fivefold cross validation. Accuracy, sensitivity as well as specificity are reported for different models. For the

models based from transfer learning SVM with quadratic kernel was used. The features are extracted from fully connected (FC-7) using AlexNet and VGG (16-19) whereas they are extracted from FC1000 in ResNet and GoogleNet.

The fine tuning approach on Res-Net-101 produced high score as compared to the transfer learning models.

The classes name are denoted as AF for atrial fibrillation, N for normal, NI for Noisy and O for others. The term TL is used for Transfer Learning in the Table.1. ResNet-101 retrained with four classes and with softmax classifier produced excellent performance for all performance metrics (accuracy, sensitivity, specificity). The overall performance of ResNet model is good in both transfer learning and fine tuning approach. The results obtained in Table. 1 used the PhysioNet/Computing in Cardiology dataset.

Table 1. Comparison of performance metric using transfer learning and fine tuning based models

Pretrained Models	Accuracy	Sensitivity	Specificity
ResNet-101-TL (Ensemble)	98.76	97.88	96.98
ResNet-18-TL (Ensemble)	98.69	98.12	97.63
ResNet-18- TL (SVM)	96.23	95.56	95.22
ResNet-50-TL (SVM)	96.99	96.92	94.49
ResNet-101-TL (SVM)	97.34	96.34	95.98
ResNet-101	97.89	97.12	96.99

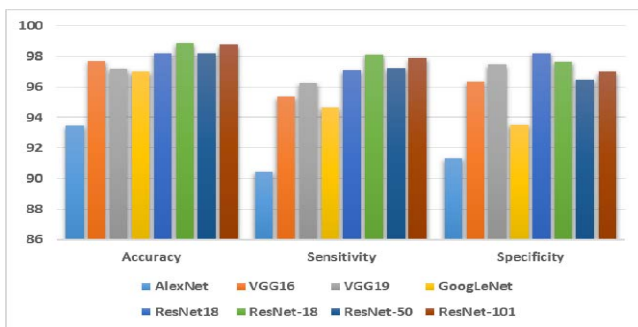


Fig. 4 . The comparison between performance metrics using pre-trained models based on SVM classifier

The performance metrics for pre-trained models using SVM classifier are shown in Fig. 4. The ResNet-101 model produced highest accuracy and ResNet-18 showed highest sensitivity. The AF detection is the preform prediction of the stroke and sensitivity is a very important metric to detect the AF form heathy and normal person.

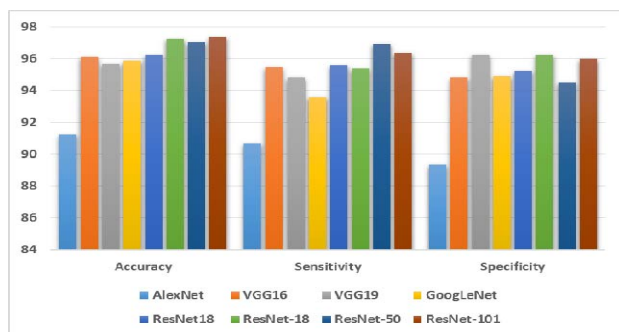


Fig. 5. The performance comparison between pretrained networks based on ensemble classifier

The ensemble classifier with subspace discriminate kernel function used CNN models features for classification of ECG rhythm and shows highest sensitivity metric for ResNet-101 model is shown in Fig. 5. This technique could be used for AF detection because high sensitivity values shows identification of AF rhythm with other ECG rhythms.

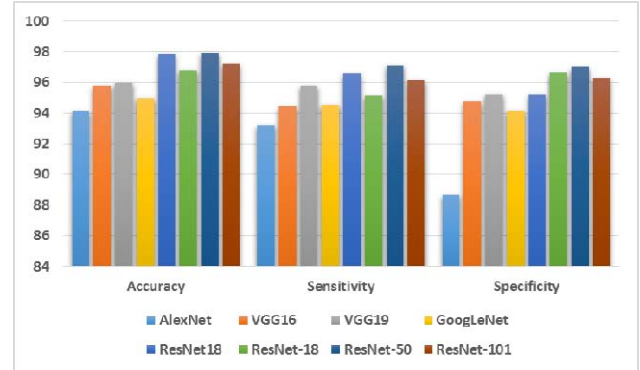


Fig. 6. The performance comparison of pretrained fine-tune models using softmax classifier

The different CNN based pretrained models show performance metric in Fig. 6. The ResNet again produced highest performance with softmax classifier and all performance metrics provided consistent results. Although the performance of the presented method is good, the AF detection problem is far from being solved. The comparison of our proposed method with existing RR-based features and deep learning approaches is shown in Table. 2.

Table.2 Comparison with existing methods related to used ECG dataset.

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
Acharya et.al., [24]	93.53	93.71	92.83
Zhou et. al. [25]	97.83	98.44	97.99
Petrenase et. al., [13]	--	97.1	98.3
Lee et. al. [26]	97.91	98.22	97.68
Acharya et al., 2016 [27]	98.80	99.45	96.27
Sharma et al., [28]	96	93	99
Zhenjie et al., [29]	98.18	98.22	98.11
Proposed (ResNet-101)	97.89	97.12	96.99
ResNet-101-FE (ESD)	98.76	97.88	96.98

Acharya [24] used 1D convolutional neural network for AF classification and detection. His method produced lower performance metrics as compared to our proposed methods. The same author [27] achieved highest accuracy and sensitivity using k-nearest neighbors algorithm (KNN) based machine learning algorithm with handcrafted features however, this method totally relies only on the particular RR-based features and did not provide generalized solution for any ECG AF based rhythms.

Zhenjie Yao [29] proposed multiscale scratched based CNN model and achieved better results as compared to our method, again this method might not generalize well for other ECG dataset. Sharma method performed better as comparison with all methods presented in Table. 2 due to high Specificity but Accuracy and Sensitivity did not perform well. Our proposed method could be better due to automatic feature detection and classification for ECG data.

There are still quite a lot of problems worth exploring. Based on the literatures, the bigeminy, trigeminy and sinus arrhythmia are easy to be erroneous for AF, which would lead to false alarm. It is inflexible to differentiate AF and part of these arrhythmias. Further consideration should be waged to these problems. The conventional deep learning AF methods are less computational complex as compared to deep learning methods. The devices with limited power supply and capacity have difficulty to incorporate the deep learning methods in real time environment. The most ECG only last in 10 second as compared to 128 heartbeats that last in 1 minute and a half second. Alternative other deep learning model could be used for 10 second last AF detection. They are the Recurrent Neural Networks (RNN), which are designed for time sequence classification. DL algorithms have been successfully used for AF detection and further investigation could be needed to improve the results.

IV. CONCLUSION

AF detection in an automatic way is a big issue to handle the very big ECG based nonstationary or semi-stationary signals. The AF algorithms could not perform well due to diverse feature space and repetition of the AF wave rhythms. These AF detection based algorithms may be not suitable for accurately assessing the health monitoring applications. DL algorithms potentially perform well in big data analysis as well as AF detection for modern computer science, internet of things (IoT) and biomedical imaging applications. We implement deep learning based models (transfer learning and fine tuning) in AF detection and classification. The obtained results are comparable to the recent start-of-the-art approaches but further experimentations should be done on larger datasets as well as on new models such as RNN.

REFERENCES

- [1] V. Fuster *et al.*, "2011 ACCF/AHA/HRS Focused Updates Incorporated Into the ACC/AHA/ESC 2006 Guidelines for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidel," *Circulation*, vol. 123, no. 10, pp. e269–e367, Mar. 2011.
- [2] S. Colilla, A. Crow, W. Petkun, D. E. Singer, T. Simon, and X. Liu, "Estimates of Current and Future Incidence and Prevalence of Atrial Fibrillation in the U.S. Adult Population," *Am. J. Cardiol.*, vol. 112, no. 8, pp. 1142–1147, Oct. 2013.
- [3] M. Lainscak, N. Dagres, G. S. Filippatos, S. D. Anker, and D. T. Kremastinos, "Atrial fibrillation in chronic non-cardiac disease: Where do we stand?," *Int. J. Cardiol.*, vol. 128, no. 3, pp. 311–315, Aug. 2008.
- [4] A. J. Camm *et al.*, "2012 focused update of the ESC Guidelines for the management of atrial fibrillation," *Eur. Heart J.*, vol. 33, no. 21, pp. 2719–2747, Nov. 2012.
- [5] N. E. Andrew, A. G. Thrift, and D. A. Cadilhac, "The Prevalence, Impact and Economic Implications of Atrial Fibrillation in Stroke: What Progress Has Been Made?," *Neuroepidemiology*, vol. 40, no. 4, pp. 227–239, 2013.
- [6] J. Lian, L. Wang, and D. Muessig, "A Simple Method to Detect Atrial Fibrillation Using RR Intervals," *Am. J. Cardiol.*, vol. 107, no. 10, pp. 1494–1497, May 2011.
- [7] B. Logan and J. Healey, "Robust detection of atrial fibrillation for a long term telemonitoring system," in *Computers in Cardiology, 2005*, 2005, pp. 619–622.
- [8] K. Tateno and L. Glass, "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and Δ RR intervals," *Med. Biol. Eng. Comput.*, vol. 39, no. 6, pp. 664–671, Nov. 2001.
- [9] S. Cerutti, L. T. Mainardi, A. Porta, and A. M. Bianchi, "Analysis of the dynamics of RR interval series for the detection of atrial fibrillation episodes," in *Computers in Cardiology 1997*, pp. 77–80.
- [10] S. Sarkar, D. Ritscher, and R. Mehra, "A Detector for a Chronic Implantable Atrial Tachyarrhythmia Monitor," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1219–1224, Mar. 2008.
- [11] D. E. Lake and J. R. Moorman, "Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices," *Am. J. Physiol. Circ. Physiol.*, vol. 300, no. 1, pp. H319–H325, Jan. 2011.
- [12] 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," *Biomed. Eng. Online*, vol. 13, no. 1, p. 18, 2014.
- [13] A. Petr nas, V. Marozas, and L. S rnmo, "Low-complexity detection of atrial fibrillation in continuous long-term monitoring," *Comput. Biol. Med.*, vol. 65, pp. 184–191, Oct. 2015.
- [14] J. Slocum, A. Sahakian, and S. Swiryn, "Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity," *J. Electrocardiol.*, vol. 25, no. 1, pp. 1–8, Jan. 1992.
- [15] S. Asgari, A. Mehrnia, and M. Moussavi, "Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine," *Comput. Biol. Med.*, vol. 60, pp. 132–142, May 2015.
- [16] P. de Chazal and R. B. Reilly, "A Patient-Adapting Heartbeat Classifier Using ECG Morphology and Heartbeat Interval Features," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2535–2543, Dec. 2006.
- [17] Y. LeCun *et al.*, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
- [18] S. C. Lee, "Using a translation-invariant neural network to diagnose heart arrhythmia," in *Images of the Twenty-First Century. Proceedings of the Annual International Engineering in Medicine and Biology Society*, pp. 2025–2026.
- [19] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu, "ECG signal enhancement based on improved denoising auto-encoder," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 194–202, Jun. 2016.
- [20] G. Clifford *et al.*, "AF Classification from a Short Single Lead ECG Recording: the Physionet Computing in Cardiology Challenge 2017," 2017.
- [21] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," 2014, pp. 818–833.
- [22] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [24] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Inf. Sci. (Nij.)*, vol. 415–416, pp. 190–198, Nov. 2017.

- [25] X. Zhou, H. Ding, W. Wu, and Y. Zhang, "A Real-Time Atrial Fibrillation Detection Algorithm Based on the Instantaneous State of Heart Rate," *PLoS One*, vol. 10, no. 9, p. e0136544, Sep. 2015.
- [26] Jinseok Lee, Yunyoung Nam, D. D. McManus, and K. H. Chon, "Time-Varying Coherence Function for Atrial Fibrillation Detection," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2783–2793, Oct. 2013.
- [27] U. R. Acharya *et al.*, "Automated detection and localization of myocardial infarction using electrocardiogram: a comparative study of different leads," *Knowledge-Based Syst.*, vol. 99, pp. 146–156, May 2016.
- [28] L. N. Sharma, R. K. Tripathy, and S. Dandapat, "Multiscale Energy and Eigenspace Approach to Detection and Localization of Myocardial Infarction," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1827–1837, Jul. 2015.
- [29] Z. Yao, Z. Zhu, and Y. Chen, "Atrial fibrillation detection by multi-scale convolutional neural networks," in *2017 20th International Conference on Information Fusion (Fusion)*, 2017, pp. 1–6.