Future Generation Computer Systems 113 (2020) 304–317

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks

Siti Nurmaini [a,*], Alexander Edo Tondas [b], Annisa Darmawahyuni [a], Muhammad Naufal Rachmatullah [a], Radiyati Umi Partan [c], Firdaus Firdaus [a], Bambang Tutuko [a], Ferlita Pratiwi [a], Andre Herviant Juliano [a], Rahmi Khoirani [a]

[a] *Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia*
[b] *Department of Cardiology & Vascular Medicine, Dr. Mohammad Hoesin Hospital, Palembang, Indonesia*
[c] *Faculty of Medicine, Universitas Sriwijaya, Indonesia*

ARTICLE INFO

Article history:
Received 12 February 2020
Received in revised form 15 June 2020
Accepted 8 July 2020
Available online 11 July 2020

Keywords:
Atrial fibrillation
Classification
Convolution neural network
Deep learning

ABSTRACT

The most prevalent arrhythmia observed in clinical practice is atrial fibrillation (AF). AF is associated with an irregular heartbeat pattern and a lack of a distinct P-waves signal. A low-cost method for identifying this condition is the use of a single-lead electrocardiogram (ECG) as the gold standard for AF diagnosis, after annotation by experts. However, manual interpretation of these signals may be subjective and susceptible to inter-observer variabilities because many non-AF rhythms exhibit irregular RR-intervals and lack P-waves similar to AF. Furthermore, the acquired surface ECG signal is always contaminated by noise. Hence, highly accurate and robust detection of AF using short-term, single-lead ECG is valuable but challenging. To improve the existing model, this paper proposes a simple algorithm of a discrete wavelet transform (DWT) coupled with one-dimensional convolutional neural networks (1D-CNNs) to classify three classes: Normal Sinus Rhythm (NSR), AF and non-AF (NAF). The experiment was conducted with a combination of three public datasets and one dataset from an Indonesian hospital. The robustness of the proposed model was evaluated based on several validation data with an unseen pattern from 4 datasets. The results indicated that 1D-CNNs outperformed other approaches and achieved satisfactory performances with high generalization ability. The accuracy, sensitivity, specificity, precision, and F1-Score for two classes were 99.98%, 99.91%, 99.91%, 99.99%, and 99.95%, respectively. For the three classes, the accuracy, sensitivity, specificity, precision, and F1-Score was 99.17%, 98.90%, 99.17%, 96.74%, and 97.48%, respectively. Potentially, our approach can aid AF diagnosis in clinics and patient self-monitoring to improve early detection and effective treatment of AF.

1. Introduction

Atrial fibrillation (AF) is the most common heart arrhythmia in clinical practice. Early and accurate detection of AF can significantly improve the prevention of complications associated with the cardioembolic event, such as stroke [1]. Nearly 25% of all individuals with AF are asymptomatic. In these individuals, screening with an electrocardiogram (ECG) may be the gold standard of AF detection. A typical waveform of ECG in sinus rhythm is composed of the P-wave, QRS-complex, and T-wave. However, in AF, chaotic fibrillatory waves can affect the morphology of P-wave, and the RR-interval is not constant among beats [2]. Some wearable or ambulatory 12-lead ECG cardiac monitoring devices are designed to ensure accurate AF detection. Unfortunately, these devices are expensive, time-consuming, complicated, and require long-term exposure for AF measurement. Currently, single-lead ECG with short-term detection is prevalent in daily application. Regardless, AF detection using short-term detection can be missed in many studies due to the existence of other non-AF rhythms with irregular RR-intervals and a lack of P-waves similar to AF [3,4]. Hence, a simple algorithm for improving short-term AF detection with satisfactory results is desirable.

AF episodes detection in several short terms ECG has the variable signal quality and lengths. The ambiguities of labels are due to multiple types of arrhythmia rhythms in the same recording, variable human physiology, and difficulty in distinguishing the features of the ECG signal. Therefore, the selected method must be robust with such conditions without declining system performance. The previous computer-aided detection

* Correspondence to: Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Palembang 30139, Indonesia.
E-mail address: siti_nurmaini@unsri.ac.id (S. Nurmaini).

https://doi.org/10.1016/j.future.2020.07.021

algorithm proposed for improving AF performance in classical machine learning (ML) indicates good results [5–7]. ML technique, however, based on hand-crafted feature extraction and feature selection, requires multiple steps to finish the classification stage [8]. With deep learning (DL), AF detection can be simply performed without hand-crafted feature engineering. The selection of the best model of the DL algorithm, however, is difficult because substantial data training is needed [9–12]. Currently, only a few public datasets about AF conditions exist with the number of healthy amounts more than AF condition or imbalance data [2,13–17]. Interest in the imbalanced condition is generally concentrated on the correct classification of the "rare" class. The classical ML classification algorithms, unfortunately, are commonly used to minimize the total error rate rather than investigate the rare class or imbalanced data [18].

With the rapid development of DL, studies in solving the problems for AF condition, such as deep neural networks (DNNs) [19], convolutional neural networks (CNNs) [20–23], recurrent neural networks (RNNs) [14,16], and autoencoders (AEs) [12] has increased. The CNNs are a type of DL that excels in processing 2D data, such as images. However, studies have shown promising results by considering ECG sequence signals as 1D data using convolutions outperformed RNNs and DNNs [14]. The 1D-CNNs were successfully classifying long-term ECG records in a fast and accurate manner, and it can examine the morphological characteristics and learn about the slit input signal variance during a short-term ECG [22,24]. The CNNs can generate local features (subsequence) of the signal sequence from the ECG to identify regional patterns in the convolution window. Since the same transformation is applied to every patch defined by the window, a pattern learned in one position may also be recognized in another position, making the translation of 1D convolution networks invariant.

Typically, the 2D-convolution window can be seen as a matrix of tunable parameters that scans from left to right and top to bottom through the image; this operation is capable of capturing visual features. Due to the matrix operation, 2D-data needs high computational power. A 1D-convolution process works on the same principle, with the only significant difference being the input and the filter layer dimensions. By compressing specific layers, the total number of weight parameters within the network is diminished, which in turn increases training efficiency and reduce the computational complexity. This finding suggests that 1D-CNNs are well suited for real-time, and low-power/low-memory devices, like mobile ECG with short term signal [24]. Unfortunately, the main drawback of 1D-CNNs is the homogeneous network, or it has the same neuron type in the whole structure [25]. In contrast, the neural network has a heterogeneous network structure like the human nervous system. Therefore it can be produce deficiency if it uses in multi-features with highly sophisticated and diversity patterns [25]. Thus, the application of 1D-CNNs in ECG signal processing must be improved with tuning hyper-parameter strategy to produce well-performance.

1D-CNNs technique is useful to extract features of time sequence data. As aforementioned, however, the acquisition of the ECG signal contains high-gain instrumentation amplifiers, which are easily contaminated by various sources of noise, with characteristic frequency spectrums depending on the source [26]. All these types of noise can interfere with the original ECG signal, which can deform ECG waveforms and cause an abnormal signal [27], thus the AF episodes hard to recognize. Consequently, to maintain as much of the AF signal as possible, the noise must be removed from the original signal to ensure an accurate diagnosis. The challenge in AF research is a robust detection to recognize the AF signal pattern. Unfortunately, some ECG signals have patterns with the same characteristics of the AF condition, such as atrial flutter, atrial extrasystole, supraventricular tachyarrhythmia, etc. [28]. Patients with AF also frequently present concomitantly with atrial flutter and/or atrial tachycardia [29]. In order to achieve accurate and robust detection with several conditions, the discrete wavelet transform (DWT) as noise removal and the 1D-CNNs as a classifier are combined in this study.

To evaluate the performance of the proposed method, three publicly well-known datasets – MIT-BIH Atrial Fibrillation [30], Physionet Atrial Fibrillation [31], MIT-BIH Malignant Ventricular Ectopy [32]– and ECG data from an Indonesian hospital are utilized to validate all results. We trained 1D-CNNs with 13 convolution layers using a 10-fold cross-validation scheme and repeated this evaluation against all the combinations of the hyperparameters. The research carried out to classify three classes: normal sinus rhythm (NSR), atrial fibrillation (AF), and non-atrial fibrillation (NAF). The data sample distribution reflects a real-world dataset, where only a small percentage of abnormal examples are available. The purpose of this research is to develop a simple and low computational classifier for AF detection with high generalization power, robustness, and high performance. Therefore, the implementation of 1D-CNNs can make AF detection more direct, simpler, and feasible. To achieve such purpose, this study has novelty and contribution as follow;

- Designing an automated AF detection with short-term ECG signal based on combination DWT and 1D-CNNs models in an efficient, high performance, and robust manner;
- Developing the best AF episodes classifier with high generalization power of the selected algorithm;
- Presenting the simple segmentation technique to solve the variable length of ECG sequence recording; and
- We are implementing the selected model for the non-AF condition with the morphology similarly with an AF condition in terms of irregular R–R-intervals and a lack of P-wave on four different dataset.

The rest of this paper is organized as follows: Section 2 explains the related work of the research, Section 3 describes the material and methods, and Section 4 presents the result and discussion. Finally, the conclusions are presented in Section 5.

2. Related works

In this section, we describe other existing researches related to this work. The hallmark of the AF condition is the absence of P-waves and an irregularly irregular ventricular rate [1,2]. Over the past decade, a variety of conventional methods based on the ECG features such as QRS complexes, R–R intervals, different R amplitude, and the heart rate variability was implemented to detect AF segments from the ECG signal automatically [30,33]. By utilizing such a method for AF detection, however, produce an unsatisfactory result, and the performance metric produces under 90%. Several analysis methods use classical ML that has been developed to improve AF detection and classification [5–7,33–36]. These algorithm structures are mainly consisting of three parts: feature extraction, feature selection, and classification. For instance, the artificial neural networks method was designed to classify AF segments and normal sinus rhythm by using a combination of four statistical features based on R–R intervals and the generalized linear classifier [37]. The coefficients of the discrete wavelet transform and dual-tree complex wavelet transform are combined with four morphological features for the analysis of AF episodes [38]. The combination coefficients of discrete cosine transform, empirical mode decomposition, and DWT with the K-nearest neighbor for AF signal classification [39]. Heart rate variability series and non-dominated sorting genetic algorithm to separate optimization AF signals [40]. An ECG hand-held device uses a random forest classifier has been developed to detect AF episodes [34]. Although these methods were demonstrated to be helpful for AF signals analysis, they have many apparent shortcomings such as, (i) due to the feature extraction and feature selection need the manual design for a long time; (ii) it needs trial and error strategy to solve the feature engineering; (iii) it fails to guarantee the reliable robustness and easily cause the over-fitting problems, especially validated against different types of datasets; (iv) it is unreliable to ensure the generalization ability of the method; and finally (v) all the previous studies show that the ML algorithm produces an unsatisfactory result for AF detection with accuracy values below 99% for multiclass classification.

In order to overcome the above problems, the DL approach has been proposed in the detection and classification of AF episodes. Unlike classical learning algorithms, DL approach has the feature extraction and feature selection embedded into the network. Furthermore, by using the deep structure of learning, the performance of the DL model can be continuously improved. Their typical architecture utilizes in ECG signal processing mainly includes CNNs, SAE, DBNs, RNNs, and a combination of them [13,14,17,41–43]. AF classification and analysis have reported in some literature that DL models are more effective and accurate than classical machine learning algorithms [2,13,15,17,27,43–46]. The previous studies by using a computer-aided detection system proposed for AF detection with DL algorithm including, Yuan et al. [42] proposed a DL framework based on sparse autoencoder that using R-R interval time series with a window size of 10 s. For the results, the accuracy is 75.15% and 98.30%, before and after fine-tuning, respectively.

In the following year, Ghiasi et al. [2] proposed CNNs to improve the accuracy of AF by understanding the occurrence of P-wave, and the overall accuracy achieved 71%. Zihlmann et al. [17] combined the long short-term memory (LSTM) in CNNs architecture and obtained an F1 score of 82.1% on the hidden challenge testing set. Yao et al. [27] proposed multi-scale CNNs (MCNNs) based on AF detectors using public and private datasets. They achieved the overall accuracy is 98.18% on the public dataset. Rubin et al. [45] classified the normal sinus rhythm, AF, others, and noisy signals with the CNNs. The best result achieved at the official phase on the blind test set was 80%. On the other hand, Teijeiro et al. [13] was enhanced by Rubin et al. work and obtained a final score of 83%. Xiong et al. [14] proposed a 16-layer one-dimensional CNNs to classify the ECGs, including AF segments. They compared RNNs and spectrogram learning. The final results showed CNNs outperformed RNNs and spectrogram learning with an accuracy of 82% for AF. Limam et al. [16] were also used RNNs as classifiers that combined with CNNs for classified normal sinus, AF, an alternative rhythm, and noisy signal. They only achieved 77% using the validation dataset. Acharya et al. [47] used an 11-layer deep CNNs for representing the normal sinus, AF, atrial flutter, and ventricular fibrillation class. They have used ECG signals of two seconds and five seconds duration without QRS detection. The results showed the accuracy was 92.50% and 94.90%, for two seconds and five seconds period, respectively. All the studies by using their proposed method, unfortunately, produce an unsatisfactory result with all performances under 99%.

In the previous two years of publication, Xia et al. [48] proposed deep CNNs using input generated by the short-term Fourier transform (STFT) and stationary wavelet transform (SWT), and presented accuracy of 98.29% and 98.63%, for AF detection, respectively. Warrick et al. [43] proposed an ensembling convolutional and LSTM networks that predict an AF classification at every 18th input sample. The newly proposed method improved the test score of 82% in the follow-up phase of the challenge. Erdenebayar et al. [49] designed a CNNs model and optimized it by dropout and normalization for automatically predicting AF using a short-term normal EG signal. However, all the results produce average accuracy is still under 99%. Therefore, improving the performance of AF detection is needed to investigate deeply. Cai et al. [15] built a novel one-dimensional deep densely connected neural network (DDNN) to detect AF with a length of 10 s in 12-lead ECG recordings. The DDNN obtained high performance with an accuracy of 99.35%. Nevertheless, for daily application, the short-term mobile ECG more often used compare to 12-lead ECG recording, due to simple and low cost. Thereby, simple AF detection with short-term signals in an accurate and robust manner is important for an in-depth investigation.

3. Materials and methods

3.1. Convolutional neural network

CNNs are composed of three common layers, i.e., convolutional layer, pooling layer, and fully connected layer. The convolutional layer aims to represent the features from the input, and the pooling layer aims to reduce the resolution of the feature maps [50]. In the fully connected layer, the classification function is performed, the final class vector is generated [48]. All of the activation neurons from the previous layer are connected to the neurons in the fully connected layer to produce a decision. By using the activation function, the network determines the classification results. CNNs have a hierarchical architecture; starting from the input x, each subsequent layer x_j is computed as [51],

$$x_j = \rho W_j x_{j-1} \tag{1}$$

where W_j and ρ represent a linear operator and a nonlinear operator, respectively. For CNNs, W_j and ρ represent a stack of convolution filters and activation functions, respectively, ReLU or sigmoid. W_j is learned by a stochastic gradient descent. For computer gradients, a backpropagation algorithm is employed. A sum of convolutions of the previous layers can be written as [51],

$$x_j(u, k_j) = \rho(\sum_k (a_{j-1}(., k)^* W_{j,k_j}(., k))(u)) \tag{2}$$

where $(^*)$ is the discrete convolution operator [51],

$$(f^*g)(x) = \sum_{u=-\infty}^{\infty} f(u)g(x-u) \tag{3}$$

A sequence of 1D convolutions comprises a linear weighted sum of two 1D arrays, in which forward and backward operations can be effectively executed in parallel. The forward process is expressed as [24],

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} conv1D(w_{ik}^{l-1}, s_i^{l-1}) \tag{4}$$

where x_k^l is the input, b_k^l is the bias of the kth neuron at layer l, s_i^{l-1} is the output of the ith neuron at layer $l-1$, and w_{ik}^{l-1} is the kernel from the ith neuron at layer $l-1$ to the kth neuron at layer l. The output y_k^l can be expressed by passing the input x_k^l through the activation function [24]:

$$y_k^l = f(x_k^l) \text{ and } s_k^l = y_k^l \downarrow ss \tag{5}$$

where s_k^l is the output of the kth neuron of layer l, and $\downarrow ss$ is defined as the downsampling operation with the scalar factor ss. For backward operations, assume $l = 1$ and $l = L$ for the input layer and output layer, respectively. In the output layer for the input p, the mean squared error (MSE) E_p can be defined as [24],

$$E_p = MSE(t^p, [y_1^L, \ldots, y_{N_L}^L]') = \sum_{i=1}^{N_L} (y_i^L - t_i^p)^2 \tag{6}$$

where t^p represents the target vectors.

The delta error $\Delta_k^l = \frac{\partial E}{\partial x_k^l}$ should be computed to obtain the derivative of E_p. The chain rule is used to update the bias and neurons, and all weights [24]:

$$\frac{\partial E}{\partial w_{ik}^{l-1}} = \Delta_k^l y_i^{l-1} \tag{7}$$

and

$$\frac{\partial E}{\partial b_k^l} = \Delta_k^l \tag{8}$$

When the weight and bias are calculated with the learning factor ε, the biases and weights are updated as follows [24]:

$$w_{ik}^{l-1}(t+1) = w_{ik}^{l-1}(t) - \varepsilon \frac{\partial E}{\partial w_{ik}^{l-1}} \tag{9}$$

$$b_k^l(t+1) = b_k^l(t+1) = b_k^l(t) - \varepsilon \frac{\partial E}{\partial b_k^l} \tag{10}$$

3.2. Experimental results

In this section, the proposed framework in Fig. 1 is divided by five processes: (i) normalization based on normalizing bound; (ii) noise removal using a discrete wavelet transform (DWT); (iii) segmentation based on a window size of 9 s; (iv) classification with the proposed CNNs architecture; and (v) evaluation of the proposed model with five performance metrics, i.e., accuracy, sensitivity, specificity, precision, and F1-Score.

3.2.1. Dataset

In this paper, ECG raw data and NSR and AF data are obtained from MIT-BIH Atrial Fibrillation [30], Physionet Atrial Fibrillation [31], MIT-BIH Malignant Ventricular Ectopy [32], and from an Indonesian hospital. All data are listed in Table 1. The data validation is performed using MIT-BIH Atrial Fibrillation, Physionet Atrial Fibrillation, and the Indonesian hospital. All data are collected from a short single-lead ECG recording with different frequencies. Regarding the data distribution, 5076 NSR samples, 758 AF samples, and 280 NAF samples creates a class imbalanced condition. The samples of ECG signals and the NSR, AF, and NAF samples are shown in Fig. 2.

3.2.2. Pre-processing

In this phase, three main processes – normalization, noise removal, and segmentation – exist. Raw ECG signal records require normalization because the features have different ranges (refer to Fig. 3(a)). Moreover, this process can reduce the time computation. The process of normalization implements a normalized bound, which is constrained to the range between 0 (lower limit, *lb*) and 1 (upper limit, *ub*). The normalized bound only changes the amplitude scale of the signal in a certain range and does not change the morphology of the signal. The mathematical function of the normalization with the Normalized Bound is expressed as [52],

$$f(x) = x^{*}coef - (x^{*}_{mid}coef) + mid \tag{11}$$

where

$$coef = \frac{ub - lb}{x_{\max} - x_{\min}} \tag{12}$$

$$mid = \frac{ub(ub - lb)}{2} \tag{13}$$

The normalization results are shown in Fig. 3(b). An ECG signal is always corrupted due to the interference of different types of artifacts and power lines [53]. Hence, raw ECG signals must be improved by removing numerous kinds of noise and artifacts. This paper reconstructs an ECG signal from a noisy signal by using a DWT. Such an approach performs a correlation analysis, and therefore, the output is expected to be maximal when the input signal mostly resembles the mother wavelet. Unlike the old denoising method (i.e., Fourier transform), the DWT provides an analysis of the signal, which is localized in both time and frequency.

Conversely, the Fourier transform is localized only in frequency. The algorithm of the wavelet denoising contains three steps: wavelet decomposition, coefficient processing, and wavelet reconstruction [54]. Soft Thresholding is utilized for ECG noise removal as [55],

$$thresholding(t) = \sigma\sqrt{2\log N} \tag{14}$$

where $\sigma = (median|cD_j|/0.6457)$, N is the length of the ECG signal, and σ is the standard noise deviation [55],

$$c\hat{D} = \begin{cases} sign(cD_j)(|cD_j| - t), |cD_j| \geq t \\ 0, |cD_j| \leq t \end{cases} \tag{15}$$

To remove the baseline wander and eliminate noise, the ECG signal will undergo denoising after normalization. The sample result of the ECG signal after denoising is shown in Fig. 3(c). The baseline wander of the signal is corrected, and the ECG signal nears zero points. In a previous study [9], the biorthogonal mother function (bior6.8) has applied for noise removal in ECG heartbeat. When such a function implemented in ECG rhythm for this study, however, the reconstruction results are poor. Hence, based on some experiments, Sym5 is selected as the mother wavelet because it offers better denoising of the noisy ECG signal. The DWT was designed with eight levels of decomposition for a low-pass filter and eight levels of decomposition for a high-pass filter (refer to Fig. 4), in which the largest frequency sequence ranges from level 1 to level 8 [55].

Before the classification process with the 1D-CNNs, the clean ECG signal is segmented. In the Physionet Atrial Fibrillation dataset, the length of minimum signal recording about 2700 nodes in 9 s and the maximum value of signal recording about 13 800 nodes (approximately 60 s). In this study, minimum length is utilized, which is enough to analyze the ECG rhythm regarding the possibility of AF conditions and to simplify the computation. Therefore, the length of the ECG signals with more than 2700 nodes will be adjusted to 2700 nodes (window size). The result of the ECG segmented sample, as depicted in Fig. 3(d).

3.2.3. 1D-CNNs classifier

After the preprocessing phase, the ECG signal is classified to achieve a satisfactory interpretation. As illustrated in Fig. 5, two distinct layer types are proposed in 1D CNNs: (1) the CNNs-layers as feature learning, where both 1D convolutions and sub-sampling (pooling) occur, and, (2) Fully-connected (FC) layers that are identical to the layers of a typical Multi-layer Perceptron (MLP) as the classifier. In the feature learning, data that are used as a feature will pass through the stages of convolution, non-linearity, and pooling. The proposed CNNs comprises several hyper-parameters that were cautiously selected after widespread experimentation. Each hyper-parameter was tuned by keeping all other network parameters constant and evaluating the effect of incremental adjustment of the value to the training set with 10-fold cross-validation. For each hyper-parameter, the values that yielded the best cross-validation accuracy, sensitivity, specificity, precision, and F1-score were selected for the final model. The proposed 1D-CNNs architecture is shown in Fig. 5. The ECG feature map is constructed by using 13 convolutional layers, which reduces 2700 nodes to 78 nodes for the two classes NSR and AF. To describe the convolution and max-pooling process to produce a feature map, we present all processes in Fig. 6. In the first stage,

Table 1
ECG raw data resume.

Dataset	Classes	Training	Validation	Testing (unseen)
Physionet Atrial Fibrillation	NSR	4556	520	148
	AF	694	64	47
MIT-BIH Atrial Fibrillation	NSR	–	–	285
	AF	–	–	291
	AFL	234	28	–
	J	7	–	–
MIT-BIH Malignant Ventricular Ectopy	SVTA	5	–	–
	VFL	8	–	–
Indonesian hospital	NSR	–	–	50
	AF	–	–	3

*Abbreviations: NSR (normal sinus rhythm), AF (atrial fibrillation), AFL (atrial flutter), J (AV junctional rhythm), SVTA (supraventricular tachyarrhythmia), VFL (ventricular flutter).

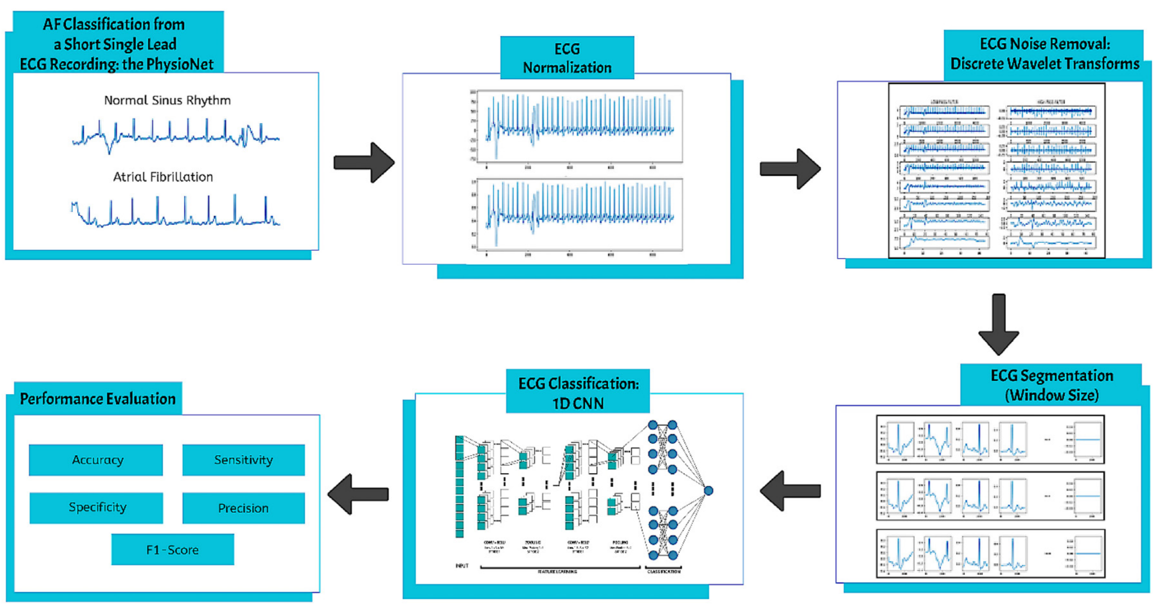


Fig. 1. Research framework.

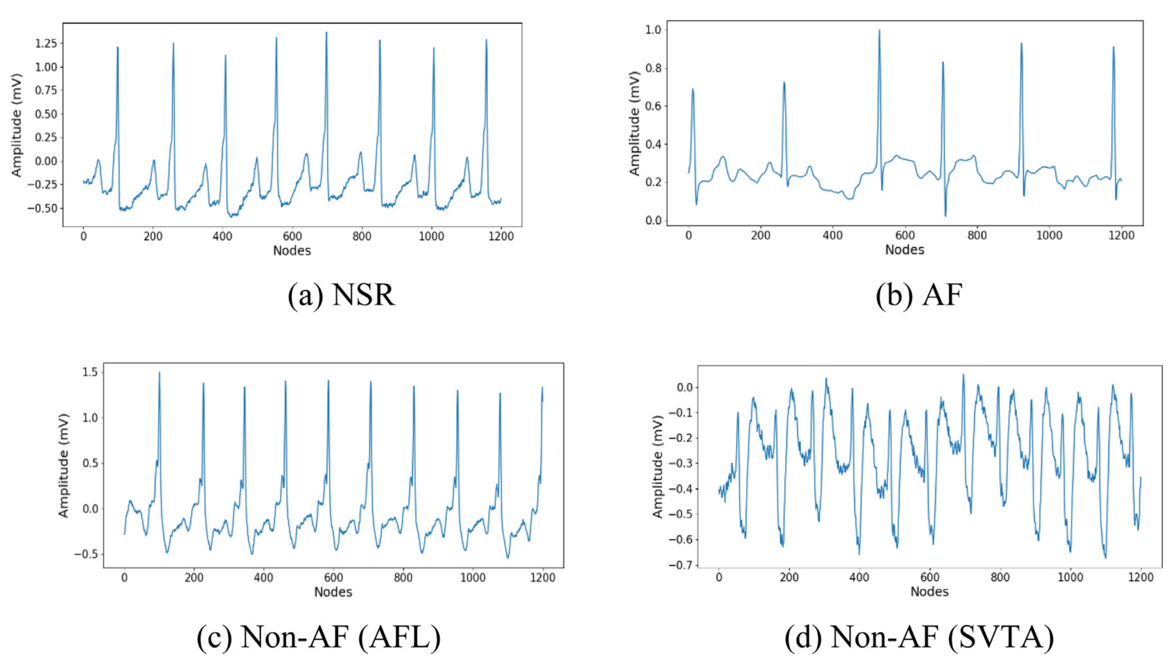
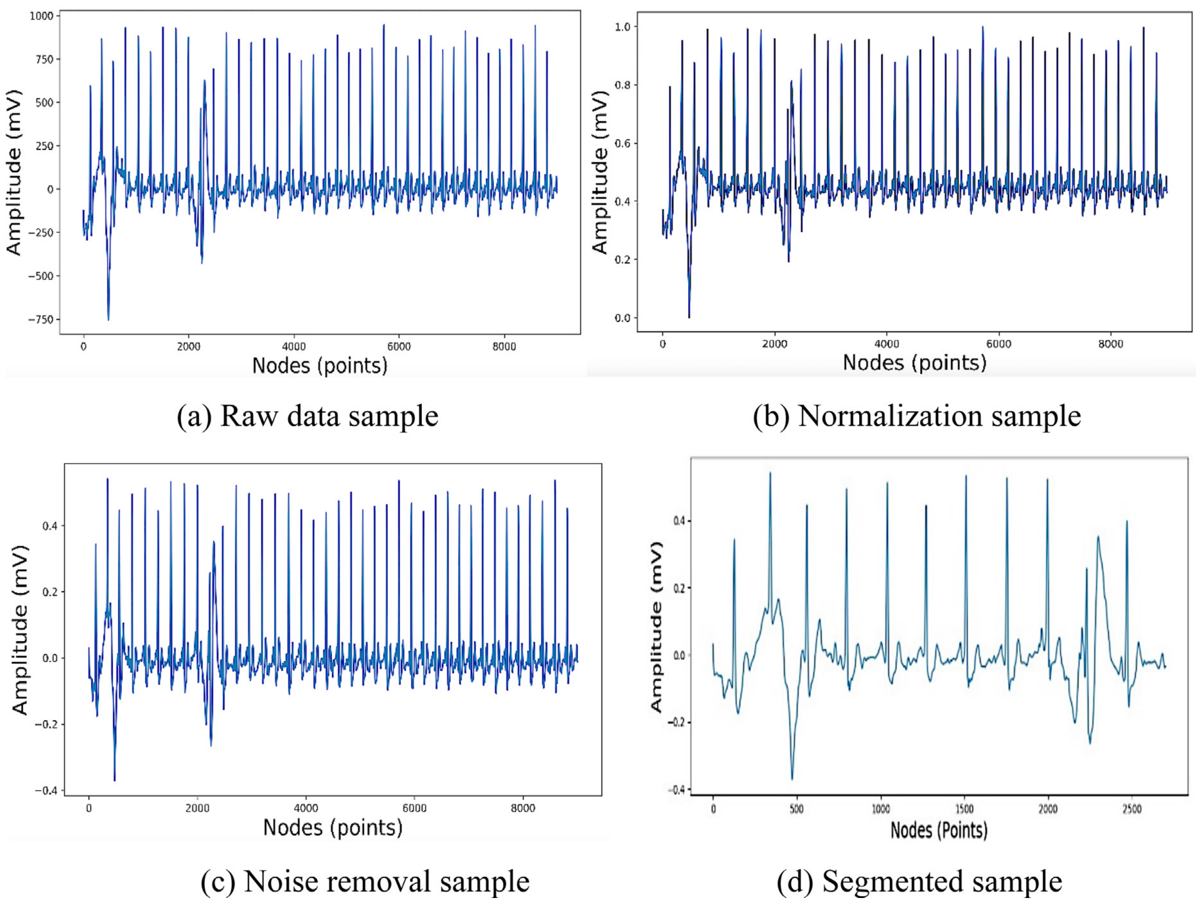


Fig. 2. Sample of ECG recording to describe NSR, AF, and NAF.



(a) Raw data sample (b) Normalization sample

(c) Noise removal sample (d) Segmented sample

Fig. 3. All phase of data preprocessing for NSR sample in terms of raw data, normalization, noise removal, and segmentation.

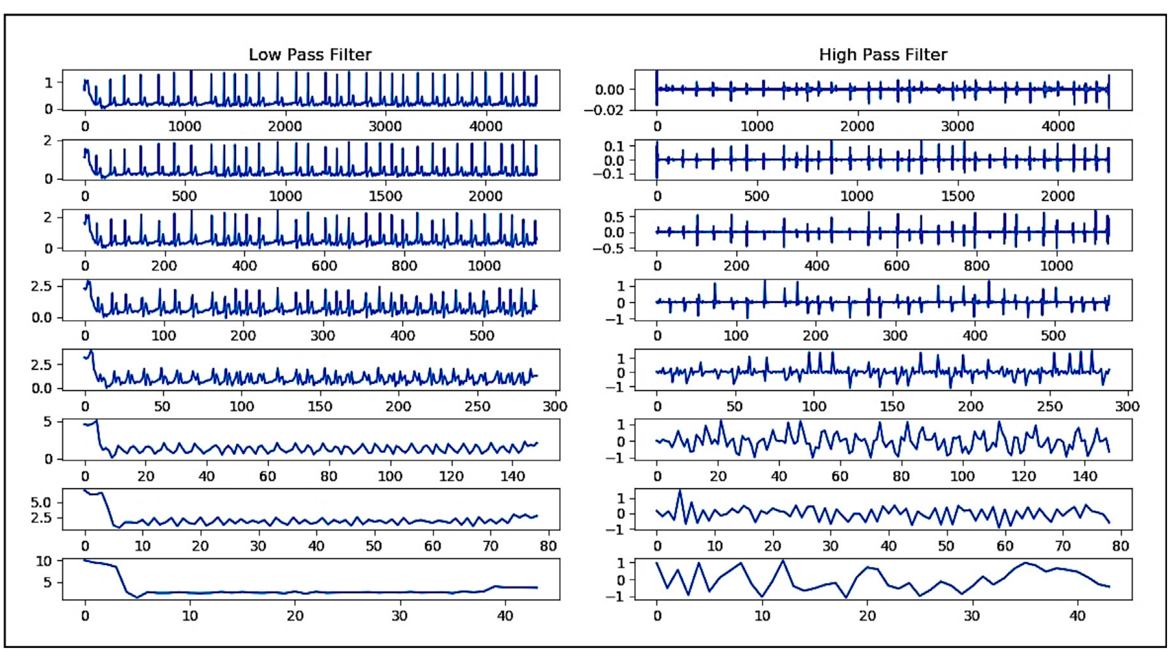


Fig. 4. Low-pass and high-pass filters for noise removal in ECG signal with Sym5.

a signal with a length of 2700 nodes will be convoluted with 64 filters or 3 × 1 kernels with stride 1.

The convolution process will produce a feature map that will be utilized as new input data for the next process. All features will be generated after the non-linearity process and only during each convolution process; this step never occurs in the pooling process. This allowed the model to generate 64 unique features on the first layer of the network. We use ReLU as a non-linearity function because of its ability to prevent vanishing gradient in the training phase. The result of the convolution layer called the feature map.

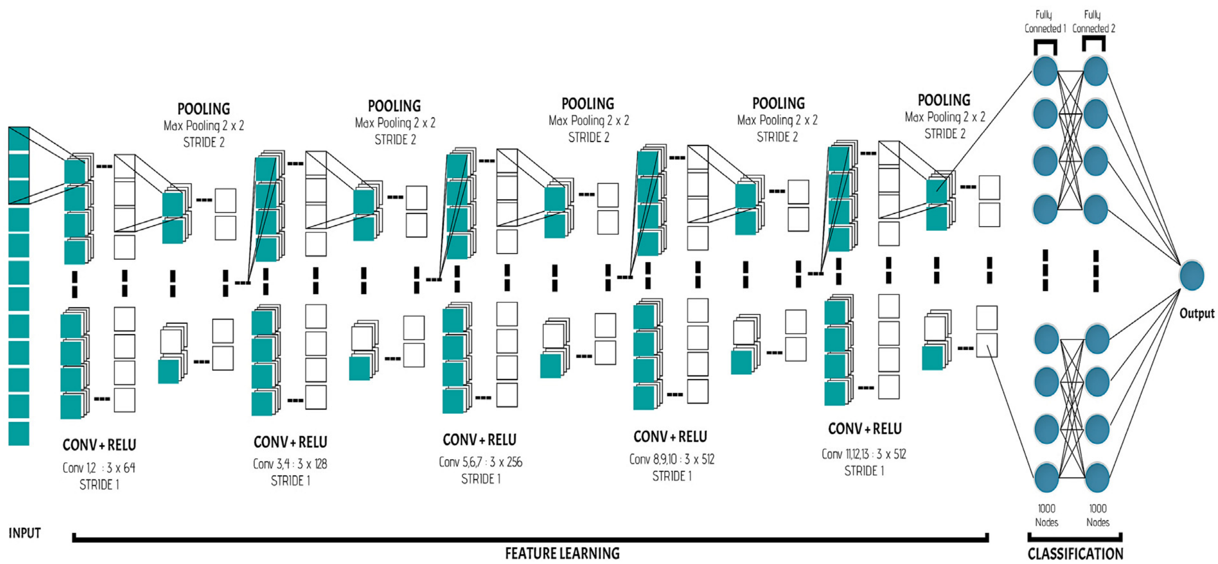


Fig. 5. Proposed 1D-CNNs architecture.

After the second convolution layer, we add the pooling layer. This layer aims to reduce the complexity of the feature map. Max pooling layer is used due to it can extract the most important features from the feature map. In the fully connected part, we define two hidden layers, each with 1000 nodes, while the output layer we define a sigmoid function to classify the data.

The 1D-CNNs model with 13 layers comprises two fully connected layers, with 1000 nodes for each layer and one node for the output layer. This model requires 3D input, which consists of *n* samples, *n* features, and time-steps, while a 2D ECG signal only consists of *n* samples and *n* features. Therefore, 2D data is reshaped into 3D data. To achieve a good generalization capability, the CNNs training process requires large numbers of datasets. To overcome this problem, we segment and combine several datasets. The results of the segmented signal become an input for the CNNs process.

All parameters, including the number of samples for the NSR, AF, and NAF condition, which is approximately 6116 samples; the number of features, which is approximately 2700 nodes; and the time steps, which is approximately 9 s. In a fully connected layer, classification consists of training, validation, and testing processes. All data are validated with 10-fold cross-validation to select the best model. In the fully connected layer, an automatic prediction of the classes is provided by learning these feature maps. To implement the proposed CNNs model, Python 3.6 software on the Keras library with a Tensor-Flow background was employed. The model was trained and evaluated by using the Nvidia graphics processing unit GeForce RTX 2080 TI with the Windows 10 operating system environment.

4. Results and discussion

The 1D-CNNs classifier workflow consists of several stages: determining the number and size of the convolution layer, max pooling, and the number of nodes in the fully connected layer. All data are splitting and resampling for training and validation with 10-fold cross-validation. Nine models of CNNs hyperparameters were compared to produce the best performance of the classifier. The hyperparameters are tuned with different nodes, such as 2700 nodes or 9 seconds, and 18300 nodes or 60 seconds, and 18300 nodes with a time step, as shown in Tables 2 and 3. All models are conducted with three models of CNNs with

2700 → ?

Table 2
Nine models of CNNs structure for tuning hyper-parameters.

Model	Segmentation	Convolution layers	Fully connected and output layers
1		7	(1000,1000,1)
2	9 seconds	10	(1000,1000,1)
3		13	(1000,1000,1)
4		7	(1000,1000,1)
5	60 seconds	10	(1000,1000,1)
6		13	(1000,1000,1)
7	60 seconds	7	(100,100,1)
8	with time step	10	(100,100,1)
9		13	(100,100,1)

Table 3
Results of model benchmarking of 1D-CNNs.

Model	Performance results (%)				
	Accuracy	Sensitivity	Specificity	Precision	F1-Score
1	87.83	71.72	71.72	57.82	60.08
2	89.03	77.66	77.66	62.32	65.92
3	**92.97**	**87.46**	**87.46**	**71.78**	**81.63**
4	83.63	51.52	51.52	50.73	50.24
5	84.66	58.59	58.59	54.84	55.62
6	87.48	0	0	50.00	0
7	83.29	49.89	49.89	49.94	49.19
8	81.66	52.27	52.27	51.65	51.72
9	82.34	51.40	51.40	50.87	50.70

7, 10, and 13 convolutional layers. All models of CNN's hyper-parameters utilize activation function ReLUs in the hidden layers, tanh-sigmoid in the output layers, a learning rate of 0.0001, a batch size of 16, 100 epochs, and adaptive moment estimation (Adam) as optimizers the weight of parameters. The objective function in the output layer is binary cross-entropy. From Table 3, the bold values at model 3 indicate the highest accuracy with corresponding to sensitivity, specificity, precision, and F1-Score for model benchmarking.

From the previously mentioned models, all performances of model 3 outperformed other models in terms of accuracy, sensitivity, specificity, precision, and F1-score. Model 3 produces satisfactory results; therefore, this model becomes a baseline to show the performance in three types of convolutional layers. Adjusting the number of nodes of the CNNs layer produces a significant change in the classifier performance. Based on the

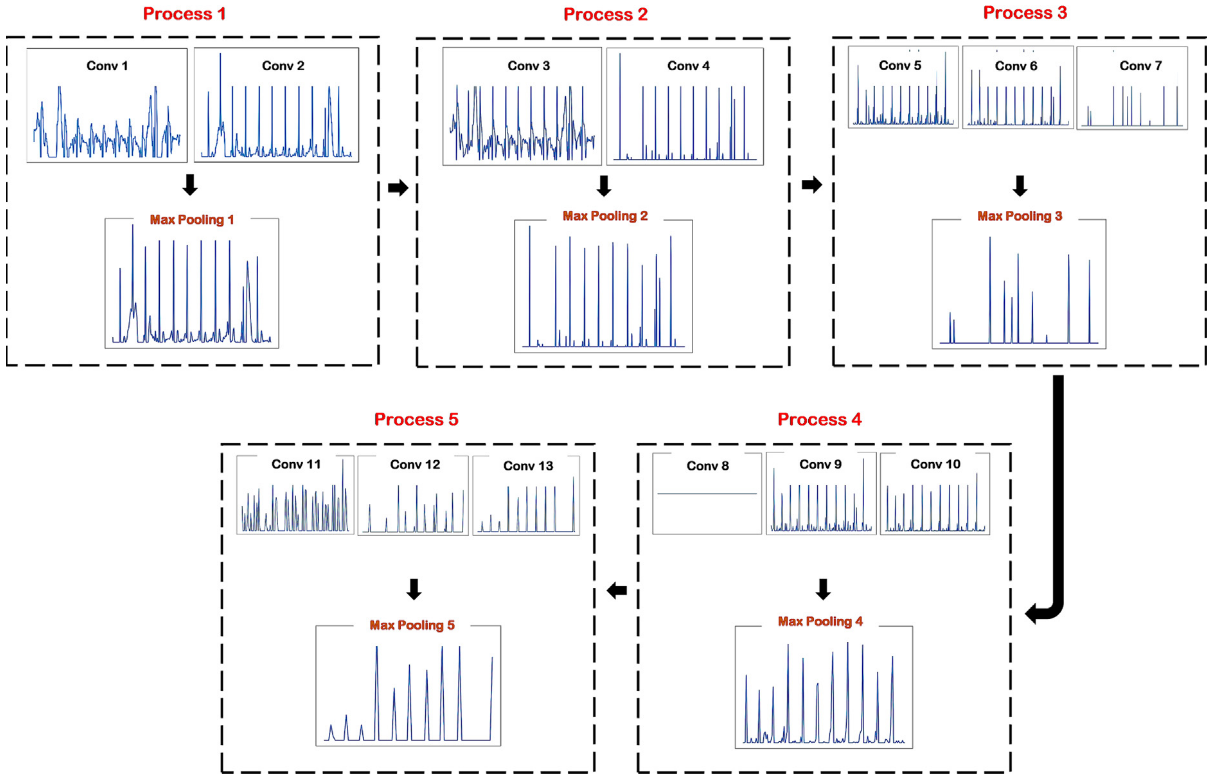


Fig. 6. Features map from convolution and max-pooling process.

result, however, classifier performance with three models (7, 10, and 13 layers) produces an unsatisfactory result—all values achieved under 90% because of the class imbalance. Therefore, the training model is enhanced by using k-fold to improve the results. Such a procedure involves splitting the training dataset into k-folds. Three schemes of splitting data were created, including stratified, shuffle, and fixed sampling to evaluate the proposed model. Based on such a process, the proposed 1D-CNNs structure is selected with k = 10-fold. It produces high performance in accuracy, sensitivity, specificity, precision, and F1-Score, which reveals the ability of the classifier to predict the increase in the minority class (refer to Fig. 7).

In this study, to support our proposed 1D-CNNs architecture, we benchmark the algorithm with other DL techniques, including DNNs and RNNs. This section also discusses the use of the data ratio between imbalanced and balanced class. Although slight skewness in the balance levels is acceptable, an increase in this gap causes improper classifier training and inappropriate predictions. In an imbalanced class, a classifier tends to predict the majority of classes effectively. However, the minority class prediction levels are substantially reduced, which reduces the reliability levels of the model. An algorithm analysis in terms of the level of imbalance, data size, and their impact on classifier metrics is presented. Three types of RNNs are implemented with a data ratio of 1:8 for the balanced and imbalanced classes. A variance sampling technique was utilized: Synthetic Minority Oversampling Technique (SMOTE) and Random Oversampling (ROS). In Table 4, with the RNNs imbalanced data ratio between classes, all performance metrics exhibit a downward decline. When the data are balanced by SMOTE, the results increased significantly and outperformed the CNNs, DNNs, and RNNs. RNNs with SMOTE has accuracy, precision, sensitivity, specificity, and F1-Score of 94.83%, 94.94%, 94.95%, 94.78%, and 94.78%, respectively. This result was also obtained for the RNNs with a maximum ROS, the accuracy, sensitivity, specificity, precision, and F1-score of 88.93%.

Therefore, the partition of the data between classes is affected by the performance of the classifier. The RNNs with balanced data, achieve effective performance. However, this method can increase the weight of the minority class by replicating the minority class examples. Although it does not increase the amount of information, it increases the overfitting, which causes the model to be too specific. The accuracy of the training set is high, but the performance in new datasets is worse. When a 10-fold cross-validation scheme is implemented, the performance of the CNNs increases significantly (refer to Table 5). The imbalanced class can be overcome with a k-fold cross-validation strategy because it tunes the class weight with the resampling procedure for the total data and reduces the bias. The CNNs approach is proven to be sensitive to the ECG signal quality to present a promising classifier.

In our study, the model is constructed by using two classes condition (NSR and AF), and three classes (NSR, AF, and NAF). An in-depth investigation is carried out to ensure the robustness of the selected model. The proposed model was tested with other datasets in the testing stage as the unseen patterns (refer to Tables 5 and 6). Table 5 indicates the satisfactory performance of the validation data of Physionet Atrial Fibrillation. For the NSR class, a perfect sensitivity of 100% is achieved, which is similar to the AF class for the specificity and precision. The performance of all testing data produces consistent results for the NSR and AF conditions, even though the imbalance ratio for the data between NSR and AF is 1:8.

Our proposed model was implemented for three classes (NSR, AF, and NAF). The characteristics of AF is the absence of P-waves and irregular P-waves among beats. However, the baseline fibrillatory waves (f-waves) can be characterized by other waves, such as flutter waves, and can be misinterpreted. To ensure the robustness of the proposed model, NAF only uses the ECG signal with irregular RR-intervals and lacks P-waves, such as atrial flutter, AV junctional, supraventricular tachyarrhythmia, and VFL

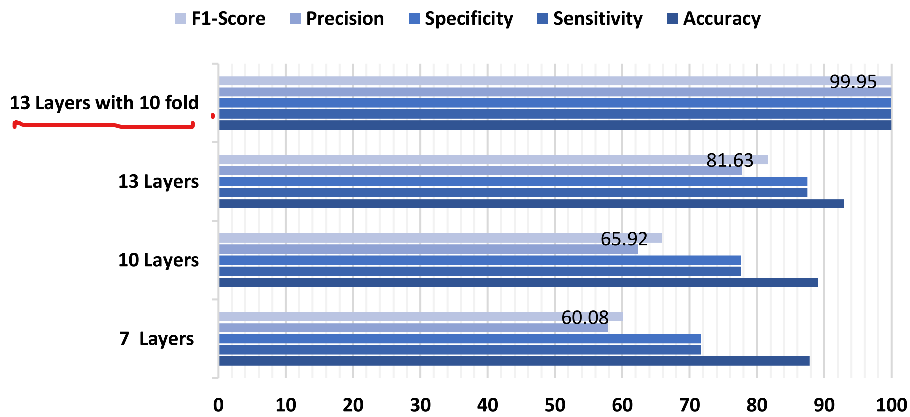


Fig. 7. Sample validation results based on model 3 of CNNs structure.

Table 4
The average performance of the classifier with several DL models for AF and NSR classes based on the model.

Classifier	Pre-processing	Performances (%)				
		Accuracy	Precision	Sensitivity	Specificity	F1-Score
DNNs	DWT	83.54	51.56	52.95	52.95	51.42
RNNs	DWT	84.23	62.21	65.72	65.72	63.55
RNNs	DWT+ROS	92.65	89.16	88.99	94.55	88.93
RNNs	DWT+SMOTE	94.83	94.94	94.95	94.78	94.78
1D-CNNs	DWT	92.27	77.78	87.46	87.46	81.63
1D-CNNs	DWT+10-fold	99.98	99.91	99.91	99.99	99.95

Table 5
Performance of each class of ID-CNNs for two classes.

Metrics	Validation		Testing 1		Testing 2		Testing 3	
	NSR (%)	AF (%)	NSR (%)	AF (%)	NSR (%)	AF (%)	NSR (%)	AF (%)
Accuracy	99.98	99.98	99.50	99.50	86.18	86.18	98.33	98.33
Sensitivity	100.0	99.83	100.0	98.0	80.60	94.16	97.50	100.0
Specificity	99.83	100.0	98.0	100.0	94.16	80.60	100.0	97.5
Precision	99.98	100.0	99.38	100.0	95.32	77.90	100.0	96.70
F1-Score	99.99	99.91	99.68	98.89	87.10	87.10	98.57	98.00

*The validation process for two classes use Physionet Atrial Fibrillation dataset, and testing 1, 2, 3 process use unseen data from Physionet Atrial, MIT-BIH Atrial Fibrillation, and Indonesian hospital data respectively.

Table 6
Performance for each class of ID-CNNs for three classes.

Metrics	Class			
	NSR (%)	AF (%)	NAF (%)	Average (%)
Accuracy	98.90	98.94	99.87	99.17
Sensitivity	98.88	98.02	99.57	98.90
Specificity	99.02	98.99	99.88	99.17
Precision	99.88	92.96	97.39	96.74
F1-Score	99.36	94.68	98.41	97.48

*The Validation process NSR and AF using Physionet Atrial Fibrillation data, and validation for NAF using MIT-BIH Atrial Fibrillation, and MIT-BIH Malignant Ventricular Ectopy dataset (see Table 1).

ventricular flutter (refer to Table 1). As indicated by the results in Table 6, our proposed 1D-CNNs model can be developed by three classes. The average results in the three classes are 96.16% for all performance metrics. Besides, the average performance for two or three classes (refer to Table 7) with the training, validation, and testing data is 86%. Our proposed 1D-CNNs remains robust in other datasets. The model can be generalized and developed for binary or multiclass classification.

The proposed model produces a receiver operating characteristic (ROC) with an area under the curve (AUC) of 99% for the validation/test set and is robust to motion artifacts that are inherent to ECG signals (refer to Fig. 8(a) and (b)). Moreover, the 1D-CNNs with hyperparameter tuning that employed 10-fold cross-validation can overcome the imbalance classes, with high accuracy for two and three classes (refer to Fig. 8(c) and (d)). The performance of the 1D-CNNs for the two classes model does not degrade when predicting other datasets as unseen patterns, including MIT-BIH Atrial Fibrillation data, Physionet Atrial Fibrillation, and patient data from the Indonesian hospital. Similarly, when the proposed model is employed again in three classes, to validate the Non-AF rhythms with irregular RR-intervals and a lack of P-waves similar to AF, it produces robust detection. The f-waves amplitudes may vary from small to large. Large f-waves should not be mistaken for flapping waves (f- waves) observed in atrial flutter. Our proposed method successfully predicts conditions, such as a Non-AF rhythm.

1D-CNNs operation complexities are proportional to the total number of connections between two consecutive layers (convolutional layer and pooling layer), which are the multiplication of the number of neurons at each layer. In this study, we define a model with seven convolution layers with three polling layers, ten convolution layers with four polling layers, and 13 convolution layers with five polling layers. To reduce the complexity of the proposed model, we add a more pooling layer. Therefore, the number of feature maps can reduce before enter the fully-connected layer. Based on such a combination, structure in consecutive layers

Table 7
Average performances of ID-CNNs for two classes and three classes.

Classes	Learning	Performances (%)				
		Accuracy	Sensitivity	Specificity	Precision	F1-Score
NSR and AF	Training	99.97	99.88	99.88	99.98	99.93
	Validation	99.98	99.91	99.91	99.99	99.95
	Testing 1	86.18	87.38	87.38	86.59	86.00
	Testing 2	98.33	98.75	98.75	98.33	98.29
	Testing 3	99.50	99.00	99.00	99.68	99.28
AF and NAF	Training	100.0	100.0	100.0	100.0	100.0
	Validation	99.71	99.68	99.68	99.58	99.63
NSR, AF, NAF	Training	100.0	100.0	100.0	100.0	100.0
	Validation	99.17	98.90	99.17	96.74	97.48

*Testing 1 use unseen data from MIT-BIH datasets, testing 2 use unseen data from Indonesian Hospital data, testing 3 use unseen data from Physionet Atrial Fibrillation dataset.

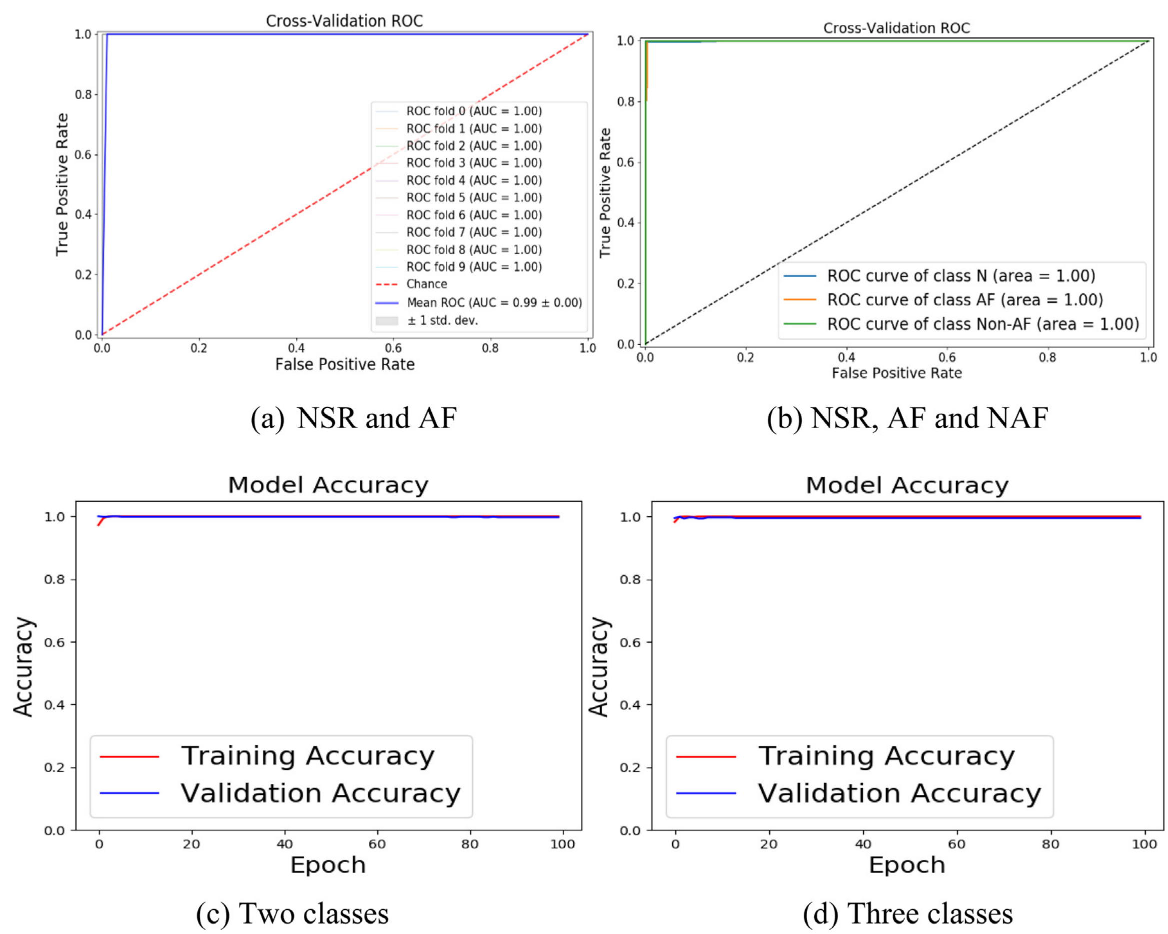


Fig. 8. Classifier performances based on ROC and accuracy curve.

produce the number of weight parameters, about 86 million, 87 million, and 45 million, respectively. However, in the 1D-CNNs classification process, only a scalar (weight) multiplication and addition are performed for each connection. Even though the number of weight parameters is high, but it is quite negligible in the implementation. That is, due to the scalar operation, so that a minimal computational complexity is achieved against the competing (conventional) methods. In this study, we evaluated low computational complexity associated with processing time utilize the ordinary computer to prove it can process with a satisfactory result. There are five computers specification used to implement the 1D-CNNs, including one core CPU, four-core CPU with and without GPU, an eight-core CPU with and without GPU. Amount of data about 5250, with input shape 2700 x 1, and 100 epochs.

Table 8 presents the processing times of both the training and testing process. The result found with one core CPU with two threads, 2.2 GHz, memory about 13 GB, and disk about 33 GB, take 33 h for training, and 0.515 s for testing a single instance. But with an eight-core CPU without GPU takes 7.7 h for training and 0.1389 s to test a single instance which only 3.7 times faster, the value does not significantly. Nevertheless, it can still process all the AF classification phase with good performance. It proves that the 1D-CNNs produce low computational complexities that suitable for low cost and low power hardware.

In recent years DL has become increasingly popular, as it eliminates the need for extraction of features. An increasing amount

Table 8
Processing time of 1D-CNNs results by using 13 layers.

Computer specification	Processing time of training	Processing time of testing
CPU: 1 Core, 2 thread, 2.2 GHz Memory: 13 Gb Disk: 33 Gb GPU: -	33 hours (Around 1137 seconds per epoch)	Single instance: 0.515 seconds All test data (584 instances): 57.458 seconds
CPU: 4 Core, 8 thread, @2.8 GHz Memory: 16 Gb Disk: 1000 Gb GPU: -	73600 seconds	Single instance: 0.3026 seconds All test data (584 instances): 29.0081 seconds
CPU: 4 Core, 8 thread, @2.8 GHz Memory: 16 Gb Disk: 1000 Gb GPU: GTX 1050 Ti, 4 Gb	6345 seconds	Single instance: 0.1826 seconds All test data (584 instances): 1.6125 seconds
CPU: 8 Core, 16 thread, @3.6 GHz Memory: 32 Gb Disk: 1000 Gb GPU: -	27400 seconds	Single instance: 0.1389 seconds All test data (584 instances): 3.5426 seconds
CPU: 8 Core, 16 thread, @3.6 GHz Memory: 32 Gb Disk: 1000 Gb GPU: RTX 2080 Ti, 11 Gb	1310 seconds	Single instance: 0.021 seconds All test data (584 instances): 0.296 seconds

of research has applied the DL approach to AF classification and other types of arrhythmia given their superior performance. However, high sensitivity and specificity are necessary for AF detection to avoid causing many false negatives that can generate needless apprehension in patients and additional costs of follow-up inspections. The applications of 1D-CNNs on ECG signals and a comparison between our model and other DL approaches were summarized in Table 9. Some DL algorithms are an effective way of defining discriminatory characteristics from a list of sorted RR-intervals and analysis of atrial activity by understanding if the P-wave is present in the ECG signal. This is done by investigating the morphology of P-waves.

Ghiasi et al. [2] used the segments with 600 samples as the input of a one-dimensional CNNs. Whitaker et al. [7] learned a 32-element sparse coding dictionary on the sorted RR-intervals of the ECG signals. Yuan et al. [42] are also proposed the stack sparse autoencoder with high accuracy of 98.30%. However, the autoencoder model cannot develop a mapping that stores the training data since our input and target output have a different shape. This is similar to sparse autoencoder, which requires that the individual nodes of a trained model that activate are data-dependent, and that different inputs will result in activations of different nodes through the network. ECG signal is categorized as sequence data, which is the order of the data is matters.

CNNs were implemented to AF classification and outperformed the other DL techniques [2,14,27,45–49] In Acharya et al. [47] used CNNs architecture with separated two different durations of the window size of a fixed length instead of analyzing one beat of ECG signal, i.e., two seconds and five seconds. They used ReLU as an activation function with 11 layers, and it validated using 10-fold cross-validation without the pre-processing of QRS detection. Luo et al. [46] proposed a 52-layer CNNs as the best performance and obtained 83.8% accuracy. They use 5-fold cross-validation to evaluate the proposed CNNs architecture. Cao et al. [27] trained the fast down-sampling residual CNNs (FDResNets) with a large difference in length is pre-segmented into short samples of 9 s. The [0, 9.375 Hz] reconstruction dataset trained by FDResNet using 6-fold cross-validation and reached test accuracy and F1 score, 92.1%, and 89.9% of the multi-scale residual neural network, respectively. Unfortunately, their researches have validated in limited datasets, i.e., 2017 Computing in Cardiology (CinC) and MIT-BIH arrhythmia database. The possibility of a proportion of occurrence and feature space distribution is diverse. Therefore, the algorithms for AF detection are not stable and robust. All the studies have an accuracy below 90%. In [2,13–17,27,43–46], all reviews studies use imbalance data and only produce F1-Score to show the comparison between false positives and false negatives. The F1-score is perfect if close to 100%. However, all F1-score values reach under 90%, due to the minority and majority class affect the overall performance.

Cai et al. [15] improved the algorithm for AF detection using a private dataset from an intra-inter patient from 3 different sources, i.e., the Chinese PLA General Hospital (301 Hospital); CardioCloud Medical Technology (Beijing) Co. Ltd; and 11 hospitals (The China Physiological Signal Challenge, 2018). They achieved excellent performance based on Deep Densely Connected Neural Network (DDNN); nevertheless, the data samples are balanced. Our 1D-CNNs model produces values that are similar to those in their work, but our data was imbalanced. The F1-score shows that the 1D-CNNs architecture achieved 97.48% for an imbalanced class, while Cai et al. performed 90.70% for a balanced class. An imbalanced class does not affect our proposed model due to its adequate performance with other datasets. All results of this research indicate that our model is a promising screening tool for AF. The accuracy of our model could be further improved by adding further datasets of arrhythmia during training and testing since our approach is completely data-driven and can only improve with more data to learn from. This is especially the case for improving AF detection, as our study only used 5076 NSR samples, 758 AF samples, and 280 NAF samples, which is a small number of standards of today.

In our study, we have shown the superiority of the 1D-CNNs model for AF detection produces high performance with short-term signal ECG, great ability of the automatic feature extraction, but low computational complexity. 1D-convolution works on vector instead of matrix operation, thus during the process, the number connection of the 1D-CNNs layer, which is a linear weighted sum [25]. A 1D-CNNs are exceptionally viable to extract features from shorter (fixed-length) sections of the general

Table 9
Comparison of previous studies of AF detection using 1D ECG signal.

Authors	Method	Features type	Acc. (%)	Sens. (%)	Spec.(%)	F1-Score
Oh et al. [41]	CNNs+ LSTM	Beats	98.10	97.50	98.70	–
Erdenebayar et al. [49] 50	CNNs	Rhythm 30s	98.70	98.60	98.70	–
Archarya et al. [47]	CNNs	Rhythm 5 seconds	94.90	99.13	81.44	–
Yao et al. [56]	Multi-scale NN	Beats	98.11	98.22	98.18	–
Xia et al. [48]	SWT+ CNNs	Rhythm 5 seconds	98.63	98.79	97.87	–
Yuan et al. [42]	AE	Rhythm 10 seconds	98.31	95.56	98.04	–
Xiong et al. [14]	CNNs	Deep features	–	–	–	81.80
Zihlmann et al. [17]	FE+LSTM	Deep features	–	–	–	82.10
Warrick et al. [43]	Ensembled CNNs+ LSTM	Beats	–	–	–	84.50
Parvaneh et al. [44]	Dense net	Beats	–	–	–	82.00
Rubin et al. [45]	CNNs	Rhythm 15 seconds	–	–	–	80.0
Ghiasi et al. [2]	CNNs	Rhythm 2 seconds	–	–	–	71.0
Limam et al. [16]	CRNNs	0.5 ms before and after R peak	–	–	–	77.0
Teijeiro et al. [13]	RNNs	Rhythm	–	–	–	83.0
Luo et al. [46]	CNNs	Rhythm	–	–	–	83.8
Cao et al. [27]	CNNs	Rhythm	–	–	–	89.9
Cai et al. [15]	DDNN	Rhythm 10 seconds	99.35	99.19	99.44	90.70
Proposed model	CNNs (3 classes) (2 classes)	Rhythm 9 seconds	99.17 99.98	98.90 99.91	99.17 99.91	97.48 99.95

*two classes are AF and NSR, three classes are AF, NAF, and NSR, Acc is accuracy, Sens. is sensitivity, Spec. is specificity.

informational collection and where the area of the component inside the portion is not of high pertinence. CNNs work a similar way whether they have 1, 2, or 3 measurements. The thing that matters is the structure of the information and how the channel, otherwise called a convolution piece or highlight identifier, moves over the information. To define the architecture of CNNs model, the number of filters and the depth of the model must be considered.

These parameters determine the feature map as well as the complexity of the model. If the model is to shallow, then it cannot extract unique feature. On the other hand, if the model is too deep, then the complexity of the model will increase, which slow the training process. The proposed CNNs models that compressed the input and the filter layer dimensions, based on two (forward and backward) 1D arrays operation that can be effectively executed in parallel. Therefore, the total number of weight parameters within the network is decreased, which in turn increases training efficiency and reduce the computational complexity. The proposed model contains several hyper-parameters that were carefully selected after extensive experimentation. All parameters of 1D-CNNs architecture selected to deliver a decent balance between learning capacity and avoiding over-fitting to increase the robustness and generalization capability.

From all performances of our proposed model, we can conclude that the advantage of our proposed model as follows;

- The proposed 1D-CNNs method for detecting atrial fibrillation can distinguish AF signal with another AF-like signal (atrial flutter, ventricular flutter, supraventricular tachyarrhythmia, and AV junctional rhythm) and normal sinus rhythm with over 99% classification accuracy;
- The 1D-CNNs produce low computational complexity with small dimension the input and the filter layer, but difficult tasks involving AF signals can be learned;
- The proposed model has the generalization ability, due to an imbalanced class does not affect the classification performance; and
- The robustness is achieved in our proposed model caused its adequate performance with other datasets, and the result shows our model gives a satisfactory result on unseen data.

Although the results look promising, there are some limitations of our study such as,

- We cannot detect the starting point of AF, and we can predict AF episodes before the 9 s;
- To expand this study to other arrhythmia conditions might add a great contribution to this line of research;
- In order to achieve a high generalization capability for other arrhythmias condition aside of AF, larger and more varied data are still needed for learning and training; and
- A recent analysis indicates that the standardization of ECG diagnostic criteria is expected to increase the consent of clinical experts and the efficiency of computer algorithms concerning the ECG interpretation system.

5. Conclusion

A computer-aided AF detection based 1D-CNNs approach with 13-layers and 10-fold cross-validation is proposed in this study. To improve the raw ECG signal from numerous kinds of noise and artifacts, the eight levels of DWT decomposition used to reconstruct the ECG signal from the noise with Symlet5. The 1D-CNNs structure evaluates four datasets, such as MIT-BIH Atrial Fibrillation, Physionet Atrial Fibrillation, MIT-BIH Malignant Ventricular Ectopy, and from an Indonesian hospital. The experiment shows that 9 s of signal segmentation produces a good result. To prove that 1D-CNNs produce low computational complexities, the experiment conducted with five CPUs. In the tuning process, the size of filters was not evaluated, but the computational complexities due to three consecutive layers (7, 10, and 13 layers) are validated associated with processing time. The result found that the combination between DWT and 1D-CNNs for AF detection/classification not only allows the model to automatically extract meaningful features from the raw ECG signals but also helps to improve the performance of the final classification. The results revealed that two classes (NSR and AF) have an accuracy, sensitivity, specificity, precision, and F1-Score of 99.98%, 99.91%, 99.91%, 99.99%, and 99.95%, respectively, and 99.17%, 98.90%, 99.17%, 96.74%, and 97.48%, respectively, for three classes (NSR, AF, and NAF). Moreover, the proposed model also produces robust detection with an unseen pattern in three testing processes and high generalization ability with several datasets. We realize that these results need more improvement in further research, especially multiclass classification, with an imbalanced dataset. Eventually, this study will encourage more precise diagnostic assistance in places with scarce access to cardiologists or other medical resources. Hence, the proposed model has great potential to expand to a hospital computer software platform to reduce mortality and save lives. In the future, we will continue a more depth investigation to detection atrial premature complex, atrial flutter, and other arrhythmia conditions with similar ECG signal characteristic to AF by using deep learning-based 1D-CNNs frameworks to improve the model robustness.

CRediT authorship contribution statement

Siti Nurmaini: Supervision, Writing - original draft, Formal analysis, Funding acquisition. **Alexander Edo Tondas:** Validation. **Annisa Darmawahyuni:** Writing - review & editing, Methodology, Data curation. **Muhammad Naufal Rachmatullah:** Software, Writing - review & editing, Data curation, Methodology. **Radiyati Umi Partan:** Validation. **Firdaus Firdaus:** Writing - review & editing, Data curation. **Bambang Tutuko:** Resources, Data curation. **Ferlita Pratiwi:** Data curation. **Andre Herviant Juliano:** Software, Data curation. **Rahmi Khoirani:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Basic Research Grants (096/SP2H/LT/DRPM/2019) from the Ministry of Research, Technology, and Higher Education, Indonesia and Unggulan Profesi Grants 2020 from Universitas Sriwijaya Indonesia.

References

[1] P. Schwab, G.C. Scebba, J. Zhang, M. Delai, W. Karlen, Beat by beat: Classifying cardiac arrhythmias with recurrent neural networks, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[2] S. Ghiasi, M. Abdollahpur, N. Madani, K. Kiani, A. Ghaffari, Atrial fibrillation detection using feature based algorithm and deep convolutional neural network, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[3] H. Julian, V. Charle, C. Ashley, Irregular, narrow-complex tachycardia, Cardiovasc. J. Afr. 29 (2018) 195.
[4] R. Sankaranarayanan, G. Kirkwood, R. Visweswariah, D.J. Fox, How does chronic atrial fibrillation influence mortality in the modern treatment era? Curr. Cardiol. Rev. 11 (2015) 190–198.
[5] J.-M. Gregoire, N. Subramanian, D. Papazian, H. Bersini, P6568 Forecasting atrial fibrillation using machine learning techniques, Eur. Heart J. 40 (2019) ehz746–1157.
[6] R. Firoozabadi, R.E. Gregg, S. Babaeizadeh, P-wave analysis in atrial fibrillation detection using a neural network clustering algorithm, in: 2018 Comput. Cardiol. Conf., 2018, pp. 1–4.
[7] B.M. Whitaker, M. Rizwan, V.B. Aydemir, J.M. Rehg, D.V. Anderson, AF classification from ECG recording using feature ensemble and sparse coding, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[8] A. Dey, Machine learning algorithms: A review, 7, 2016, pp. 1174–1179.
[9] S. Nurmaini, R. Umi Partan, W. Caesarendra, T. Dewi, M. Naufal Rahmatullah, A. Darmawahyuni, V. Bhayyu, F. Firdaus, An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique, Appl. Sci. 9 (2019) 2921.
[10] A. Darmawahyuni, S. Nurmaini, W. Caesarendra, V. Bhayyu, M.N. Rachmatullah, et al., Deep learning with a recurrent network structure in the sequence modeling of imbalanced data for ECG-rhythm classifier, Algorithms 12 (2019) 118.
[11] S. Nurmaini, A. Darmawahyuni, S. Mukti, A. Noviar, M.N. Rachmatullah, F. Firdaus, B. Tutuko, Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification, Electronics 9 (2020) 135.
[12] A. Darmawahyuni, S. Nurmaini, et al., Deep Learning with Long Short-Term Memory for Enhancement Myocardial Infarction Classification, in: 2019 6th Int. Conf. Instrumentation, Control. Autom., 2019, pp. 19–23.
[13] T. Teijeiro, D. Castro C.A. García, P. Félix, Arrhythmia classification from the abductive interpretation of short single-lead ECG records, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[14] Z. Xiong, M.K. Stiles, J. Zhao, Robust ECG signal classification for detection of atrial fibrillation using a novel neural network, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[15] W. Cai, Y. Chen, J. Guo, B. Han, Y. Shi, L. Ji, J. Wang, G. Zhang, J. Luo, Accurate detection of atrial fibrillation from 12-lead ECG using deep neural network, Comput. Biol. Med. 116 (2020) 103378.
[16] M. Limam, F. Precioso, Atrial fibrillation detection and ECG classification based on convolutional recurrent neural network, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[17] M. Zihlmann, D. Perekrestenko, M. Tschannen, Convolutional recurrent neural networks for electrocardiogram classification, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[18] M. Colak, E. Karaaslan, C. Colak, A.K. Arslan, N. Erdil, et al., Handling imbalanced class problem for the prediction of atrial fibrillation in obese patient, Biomed. Res. 28 (2017).
[19] S. Mousavi, F. Afghah, A. Razi, U.R. Acharya, ECGNET: Learning where to attend for detection of atrial fibrillation with deep visual attention, in: 2019 IEEE EMBS Int. Conf. Biomed. Heal. Informatics, 2019, pp. 1–4.
[20] B.S. Chandra, C.S. Sastry, S. Jana, S. Patidar, Atrial fibrillation detection using convolutional neural networks, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[21] B. Pourbabaee, M.J. Roshtkhari, K. Khorasani, Feature leaning with deep convolutional neural networks for screening patients with paroxysmal atrial fibrillation, in: 2016 Int. Jt. Conf. Neural Networks, 2016, pp. 5057–5064.
[22] B. Pourbabaee, M.J. Roshtkhari, K. Khorasani, Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients, IEEE Trans. Syst. Man Cybern. Syst. (2017) 1–10.
[23] X. Zhou, X. Zhu, K. Nakamura, M. Noro, Atrial fibrillation detection using convolutional neural networks, in: 2018 9th Int. Conf. Aware. Sci. Technol., 2018, pp. 84–89.
[24] S. Kiranyaz, T. Ince, R. Hamila, M. Gabbouj, Convolutional neural networks for patient-specific ECG classification, in: 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2015, pp. 2608–2611.
[25] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1D convolutional neural networks and applications: A survey, 2019, ArXiv Prepr. ArXiv:1905.03554.
[26] E. Castillo, D.P. Morales, A. Garc'ia, F. Mart'inez-Mart'i, L. Parrilla, A.J. Palma, Noise suppression in ECG signals through efficient one-step wavelet processing techniques, J. Appl. Math. 2013 (2013).
[27] X.-C. Cao, B. Yao, B.-Q. Chen, Atrial fibrillation detection using an improved multi-Scale decomposition enhanced residual convolutional neural network, IEEE Access 7 (2019) 89152–89161.
[28] C. Buttà, A. Tuttolomondo, D. Di Raimondo, G. Milio, S. Miceli, M.T. Attanzio, L. Giarrusso, G. Licata, A. Pinto, The supraventricular tachycardias: Proposal of a diagnostic algorithm for the narrow complex tachycardias, J. Cardiol. 61 (2013) 247–255.
[29] S. Sareen, M. Nayyar, B. Wheeler, M. Skelton, R.N. Khouzam, Electrocardiographic artifact potentially misleading to the wrong management, Ann. Transl. Med. 6 (2018).
[30] G. Moody, A new method for detecting atrial fibrillation using RR intervals, Comput. Cardiol. (1983) 227–230.
[31] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and Physionet: components of a new research resource for complex physiologic signals, Circulation 101 (2000) e215–e220.
[32] S.D. Greenwald, The Development and Analysis of a Ventricular Fibrillation Detector, Massachusetts Institute of Technology, 1986.
[33] K. Horoba, R. Czabanski, J. Wrobel, A. Matonia, R. Martinek, T. Kupka, R. Kahankova, J.M. Leski, S. Graczyk, Recognition of Atrial Fibrilation Episodes in Heart Rate Variability Signals Using a Machine Learning Approach, in: 2019 Mix. Int. Conf. Mix. Des. Integr. Circuits Syst., 2019, pp. 419–424.
[34] M. Zabihi, A.B. Rad, A.K. Katsaggelos, S. Kiranyaz, S. Narkilahti, M. Gabbouj, Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[35] J. Grégoire, C. Gilon, N. Subramanian, H. Bersini, Forecasting episodes of atrial fibrillation using a new machine learning algorithm, Arch. Cardiovasc. Dis. Suppl. 12 (2020) 103.
[36] N.R. Hill, D. Ayoubkhani, P. McEwan, D.M. Sugrue, U. Farooqui, S. Lister, M. Lumley, A. Bakhai, A.T. Cohen, M. O'Neill, et al., Predicting atrial fibrillation in primary care using machine learning, PLoS One 14 (2019).
[37] N. Henzel, J. Wróbel, K. Horoba, Atrial fibrillation episodes detection based on classification of heart rate derived features, in: 2017 Mix. Int. Conf. Mix. Des. Integr. Circuits Syst., 2017, pp. 571–576.
[38] M. Thomas, M.K. Das, S. Ari, Automatic ECG arrhythmia classification using dual tree complex wavelet based features, AEU-Int. J. Electron. Commun. 69 (2015) 715–721.
[39] H. Khorrami, M. Moavenian, A comparative study of DWT CWT and DCT transformations in ECG arrhythmias classification, Expert Syst. Appl. 37 (2010) 5751–5757.

[40] K.H. Boon, M. Khalil-Hani, M.B. Malarvili, Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III, Comput. Methods Programs Biomed. 153 (2018) 171–184.
[41] S.L. Oh, E.Y.K. Ng, R. San Tan, U.R. Acharya, Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats, Comput. Biol. Med. 102 (2018) 278–287.
[42] C. Yuan, Y. Yan, L. Zhou, J. Bai, L. Wang, Automated atrial fibrillation detection based on deep learning network, in: 2016 IEEE Int. Conf. Inf. Autom., 2016, pp. 1159–1164.
[43] P.A. Warrick, M.N. Homsi, Ensembling convolutional and long short-term memory networks for electrocardiogram arrhythmia detection, Physiol. Meas. 39 (2018) 114002.
[44] S. Parvaneh, J. Rubin, A. Rahman, B. Conroy, S. Babaeizadeh, Analyzing single-lead short ECG recordings using dense convolutional neural networks and feature-based post-processing to detect atrial fibrillation, Physiol. Meas. 39 (2018) 84003.
[45] J. Rubin, S. Parvaneh, A. Rahman, B. Conroy, S. Babaeizadeh, Densely connected convolutional networks and signal quality analysis to detect atrial fibrillation using short single-lead ECG recordings, in: 2017 Comput. Cardiol., 2017, pp. 1–4.
[46] J. Luo, C. Fu, M. Bai, Y. Zhao, Atrial Fibrillation detection with convolutional neural network, in: Proc. 2018 2nd Int. Conf. Comput. Sci. Artif. Intell., 2018, pp. 94–98.
[47] U.R. Acharya, H. Fujita, O.S. Lih, Y. Hagiwara, J.H. Tan, M. Adam, Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network, Inf. Sci. (Ny). 405 (2017) 81–90.
[48] Y. Xia, N. Wulan, K. Wang, H. Zhang, Detecting atrial fibrillation by deep convolutional neural networks, Comput. Biol. Med. 93 (2018) 84–92.
[49] U. Erdenebayar, H. Kim, J.-U. Park, D. Kang, K.-J. Lee, Automatic prediction of atrial fibrillation based on convolutional neural network using a short-term normal electrocardiogram signal, J. Korean Med. Sci. 34 (2019).
[50] Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, B. Yu, Recent advances in convolutional neural network acceleration, Neurocomputing 323 (2019) 37–51.
[51] J. Koushik, Understanding convolutional neural networks, 2016, ArXiv Prepr. ArXiv:1605.09081.
[52] G.B. Moody, WFDB Applications guide, Harvard-Mit Div. Heal. Sci. Technol. 10 (2003) 30–31.
[53] J. Wang, Y. Ye, X. Pan, X. Gao, Parallel-type fractional zero-phase filtering for ECG signal denoising, Biomed. Signal Process. Control 18 (2015) 36–41.
[54] Z. Wang, J. Zhu, T. Yan, L. Yang, A new modified wavelet-based ECG denoising, Comput. Assist. Surg. (2019) 1–10.
[55] H.-Y. Lin, S.-Y. Liang, Y.-L. Ho, Y.-H. Lin, H.-P. Ma, Discrete-wavelet-transform-based noise removal and feature extraction for ECG signals, IRBM 35 (2014) 351–361.
[56] Z. Yao, Z. Zhu, Y. Chen, Atrial fibrillation detection by multi-scale convolutional neural networks in: 2017 20th Int. Conf. Inf. Fusion, 2017, pp. 1–6.

Prof. Siti Nurmaini, is currently a professor with the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. She received her Master's degree in electro technique/full systems, from the Institut Teknologi Bandung — Indonesia (ITB), in 1998, and her Ph.D. degree in computer science from the Universiti Teknologi Malaysia (UTM), in 2011. Her research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotics.