Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 17 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i17.42499

PAPER

Empowering AI-Diagnosis: Deep Learning Abilities for Accurate Atrial Fibrillation Classification

Bambang Tutuko, Annisa Darmawahyuni, Siti Nurmaini(✉), Muhammad Naufal Rachmatullah, Firdaus Firdaus, Sutarno Sutarno, Rossi Passarella, Ade Iriani Sapitri, Anggun Islami, Muhammad Wahyu Ramansyah, Muhammad Isra Al Hadi

Intelligent System Research Group, Universitas Sriwijaya, Palembang, Indonesia

sitinurmaini@gmail.com

ABSTRACT

Artificial intelligence (AI) is a powerful technology that can enhance clinical decision-making and the efficiency of global health systems. An AI-enabled electrocardiogram (ECG) is an essential tool for diagnosing heart abnormalities such as arrhythmias. The most prevalent arrhythmia globally is atrial fibrillation (AF), which is an irregular heart rhythm that originates in the atria and can lead to other heart-related complications. A trusted AI classification of AF is explored in this study. Deep learning (DL) has been used to analyze large amounts of publicly available ECG datasets in order to classify normal sinus rhythm (NSR), AF, and other types of arrhythmias. A convolutional neural network (CNN) has been proposed to extract ECG features and classify ECG signals. Based on a 10-fold cross-validation strategy, we conducted experiments involving three scenarios for AF classification: (i) a balanced set, an imbalanced set, and an extremely imbalanced set; (ii) a comparison of ECG denoising algorithms; and (iii) the classification of AF, NSR, and other arrhythmia types (15 classes). As a result, we have achieved 100% accuracy, sensitivity, specificity, precision, and F1-score for the AF, NSR, and non-AF classifications, both for balanced and imbalanced sets. In addition, for the classification of AF, NSR, and other types of arrhythmia (15 classes), the performance results achieved an accuracy of 99.77%, sensitivity of 96.48%, specificity of 99.87%, precision of 97.03%, and F1-score of 96.68%. The results can empower AI diagnosis and assist clinicians in classifying AF on routine screening ECGs.

KEYWORDS

atrial fibrillation (AF), artificial intelligence (AI), classification, deep learning (DL)

1 INTRODUCTION

Atrial fibrillation (AF) is the most common arrhythmia worldwide, presenting an increased risk of stroke and heart failure. Within the general population, diabetes mellitus (DM), high blood pressure, and coronary artery disease are widely recognized as the primary risk factors, affecting 1–2% [1]. During this abnormal heart rate,

Tutuko, B., Darmawahyuni, A., Nurmaini, S., Rachmatullah, M.N., Firdaus, F., Sutarno, S., Passarella, R., Sapitri, A.I., Islami, A., Ramansyah, M.W., Al Hadi, M.I. (2023). Empowering AI-Diagnosis: Deep Learning Abilities for Accurate Atrial Fibrillation Classification. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(17), pp. 134–151. https://doi.org/10.3991/ijoe.v19i17.42499

Article submitted 2023-06-21. Revision uploaded 2023-07-25. Final acceptance 2023-07-25.

ventricular fibrillation often occurs, causing chaotic contractions in the ventricles. This can be seen in the screening of an electrocardiogram (ECG) [3]. A cardiologist or well-trained physician visually analyzes the recorded ECG data as a frequent non-invasive diagnostic tool for identifying AF. However, this laborious process is prone to human errors [4]. Furthermore, AF frequently goes undetected because it is usually asymptomatic and difficult to differentiate from a normal sinus rhythm (NSR) and other types of arrhythmias (non-AF) [1]. Hence, it is necessary to automatically detect AF for early diagnosis and prevention of related complications [5]. Unfortunately, the automatic detection of AF is a complex problem, and state-of-the-art performance is typically achieved through the use of machine learning (ML). Lack of ML still requires human intervention for feature representation [6–9].

Artificial intelligence (AI)-enabled ECG classification utilizing deep learning (DL) can offer a superior solution for AF classification through end-to-end learning. With automated feature engineering, DL assists in decision-making instead of replacing clinical decisions. Convolutional neural networks (CNNs) are widely used deep learning algorithms for biomedical signal processing. These networks are applied in one-dimensional (1D) AF classification and have shown promising results [10–14]. CNN models used for AF classification can perform both feature extraction and classification without the need for manual feature extraction [9]. Such a network consists of multiple back-to-back layers connected in a feed-forward manner, including convolutional, normalization, pooling, and fully connected layers [15]. The 1D-CNN technique is useful for extracting features from time sequence data for robust AF classification [9, 10]. However, the intermittent nature of AF poses a challenge when it comes to screening for this condition.

The classification of an ECG signal input into its respective class is performed at the final layers of DL models. The classification can take two forms: binary classification, which distinguishes between AF and non-AF, or multi-class classification, which involves multiple categories. Some studies have selectively included related arrhythmias, such as atrial flutter (AFL), which have morphological differences from AF but still carry a similar risk of stroke. To enhance the detection of AF using DL in AI, this study examines the classification of AF, NSR, and other arrhythmias, such as AFL. The study conducts a comprehensive experiment with diverse data proportions, aiming to simulate real clinical conditions.

2 RELATED WORKS

Artificial intelligence-enabled ECG deep learning-based AF classification has been extensively explored in state-of-the-art methods. The excellence of AI is demonstrated through its ability to achieve outstanding performance in processing ECG signals using various techniques. The implementation of single-lead ECG for AF diagnosis has achieved tremendous research potential in biomedical signal processing. Nguyen et al. [16] propose stacking a support vector machine (SVM) and CNN. They employed five-fold cross-validation, in which each ECG recording was divided into fixed-length segments of 4096 points from the PhysioNet and computing in Cardiology Challenge 2017. They implemented SVM to analyze the statistical features of the prediction sequences of ECG signals. From the experiment, they achieved an average F1 score of 84.19% for AF, NSR, noisy, and other classifications.

Chen et al. [10] combined two powerful DL algorithms, namely CNN and long short-term memory long short-term memory (LSTM), for the classification of AF, NSR, ventricular bigeminy (B), pacing rhythm (P), AFL, and sinus bradycardia (SBR).

They have segmented the signal into 10-second intervals. They have explored ECG databases from the MIT-BIH arrhythmia database, PhysioNet/Computing in Cardiology Challenge 2017, the MIT-BIH NSR database, and the MIT-BIH AF database. They proposed seven convolutional layers, seven pooling layers, and two blocks of LSTM. From the experiment with five-fold cross-validation, they achieved an accuracy of 96.62%, a sensitivity of 95.40%, and a specificity of 96.80%. Petmezas et al. [12] have also experimented with the combination of CNN and LSTM for AF, NSR, AFL, NSR, AFL, and atrioventricular junctional rhythm (AVJ). The model was trained using the MIT-BIH AF database with 10-fold cross-validation. They achieved a sensitivity of 97.87% and a specificity of 99.29%.

Serhal et al. [17] investigated the use of continuous wavelet transform and two-dimensional CNN for classifying AF and non-AF from the paroxysmal AF (PAF) prediction challenge database and PTB-XL ECG dataset. From both ECG databases, the ECG signals have been segmented into five-minute intervals before and during AF and 10 seconds, respectively. They proposed three convolutional layers, three pooling layers (MaxPooling), and a fully connected layer with two dense layers. Based on the tested data, they successfully obtained accuracy rates of 95.7%, 98.8%, and 95.8% from leads D1, D2, and V1, respectively. Ma et al. [18] proposed a CNN-LSTM model for automatic classification of AF signals using the MIT-BIH AF database for training. The proposed architecture consists of a one-dimensional convolution layer with a corresponding one-dimensional pooling layer as well as a two-dimensional convolution layer with a corresponding two-dimensional pooling layer. From the experiment, they achieved an accuracy of 97.21%, a sensitivity of 97.34%, and a specificity of 97.08%.

3 MATERIAL AND METHOD

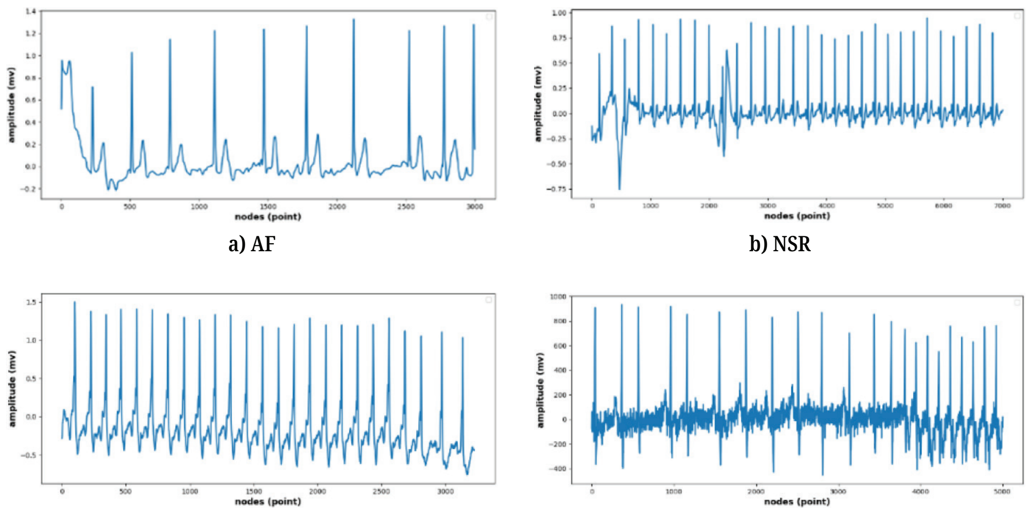
3.1 Data preparation

The experimental data for AF, NSR, and non-AF classification was obtained from various ECG databases, including PhysioNet and another repository (see Table 1). We have explored three ECG databases from the PhysioNet repository: the MIT-BIH AF [19], the Arrhythmia database from Chapman University, Shaoxing People's Hospital (Shaoxing Hospital Zhejiang University School of Medicine), and Ningbo First Hospital [20], and The PhysioNet/Computing in Cardiology Challenge 2017 [21]. MIT-BIH AF consisted of 23 long-term ECG recordings, mostly with paroxysmal AF [19]. The ECG records were sampled at a rate of 250 Hz, and each recording has a duration of ten hours. The arrhythmia database from Chapman University and Shaoxing People's Hospital consists of various ECG morphologies that belong to recorded abnormalities. The records were sampled at 500 Hz. For AF, NSR, and non-AF, the total number of records used was 1,780, 1,826, and 7,040, respectively. The PhysioNet/Computing in Cardiology Challenge 2017 [21] was sampled at a frequency of 300 Hz. The total number of records used in this study consisted of 5,154 samples of NSR and 771 samples of atrial fibrillation.

The China Physiological Signal Challenge 2018 [22] is another repository of ECG records used for this experimental study. The ECG recordings were sampled at a rate of 500 Hz, and the records consisted of 918 instances of NSR, 1,098 instances of AF, and 4,861 instances of non-AF. The main highlight of this experimental study is the utilization of different ECG morphologies with varying frequency sampling. The sample plot of AF, NSR, and non-AF recordings can be presented in Figure 1.

Table 1. The ECG recordings database

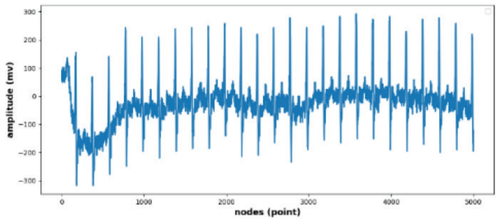
ECG Dataset	Frequency Sampling	Class		Total Records
MIT-BIH Atrial Fibrillation [19]	250 Hz	AF		23
A large-scale 12-lead electrocardiogram database for arrhythmia study (Chapman University, Shaoxing People's Hospital (Shaoxing Hospital Zhejiang University School of Medicine), and Ningbo First Hospital) [20]	500 Hz	NSR		1,826
		AF		1,780
		Non-AF	Sinus Bradycardia (SB)	3,889
			Sinus Tachycardia (ST)	1,568
			Atrial Flutter (AFL)	445
			Sinus Irregularity (SI)	399
			Supraventricular Tachycardia (SVT)	587
			Atrial Tachycardia (AT)	121
			Atrioventricular Node Reentrant Tachycardia	16
			Atrioventricular Reentrant Tachycardia	8
			Sinus Atrium to Atrial Wandering Rhythm	7
The PhysioNet/Computing in Cardiology Challenge 2017 [21]	300 Hz	NSR		5,154
		AF		771
The China Physiological Signal Challenge 2018 [22]	500 Hz	NSR		918
		AF		1,098
		Non-AF	First-degree atrioventricular block (IAVB)	704
			Left bundle branch block (LBBB)	207
			Right bundle branch block (RBBB)	1,695
			Premature atrial contraction (PAC)	556
			Premature ventricular contraction (PVC)	672
			ST-segment depression (STD)	825
			ST-segment elevated (STE)	202



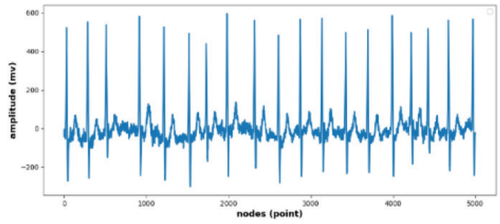
a) AF　b) NSR

c) Non-AF (Atrial Flutter)　d) Non-AF (Atrial Tachycardia)

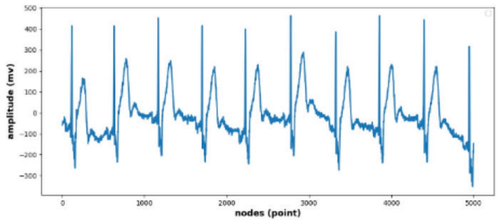
Fig. 1. *(Continued)*



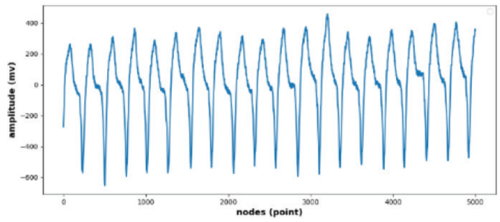
e) Non-AF (Atrioventricular Node Rentrant Tachycardia)



f) Non-AF (Sinus Bradycardia)



g) Non-AF (Sinus Tachycardia)



h) Non-AF (Supraventricular Tachycardia)

Fig. 1. The sample of AF, NSR, and non-AF recordings

3.2 ECG preprocessing

To obtain valuable information from the ECG signal, it must be processed. The challenge of ECG preprocessing is to filter and reduce ECG noise components, making it easier to identify various cardiac disorders. This study proposes the steps of ECG preprocessing to achieve effective noise removal and accurate feature extraction. The process consists of two main steps:

- **ECG denoising.** The discrete wavelet transform (DWT) is an effective approach for processing non-stationary signals. A signal can be decomposed using the wavelet transform on the time-frequency scale plane [23]. This study used various mother wavelet functions, namely *Daubechies (db), symlet (sym), biorthogonal (bior), coiflet (coif),* and *haar* wavelet. Signal-to-noise ratio (SNR) is a measure used in signal processing to compare measurement for level of a desired signal with the level of background noise. The unit of the SNR is decibels (dB). An SNR greater than 0 dB implies that the signal level exceeds the noise level. The higher the ratio, the better the quality of the signal. The results of the different wavelet families used in this study are listed in Table 2. Table 2 displays the averaged SNR for all ECG records for five wavelet families: *db, sym, bior, coif,* and *haar*. ECG records. Among the averaged SNR values, sym5 achieved the highest average SNR of 11.81 dB. Overall, sym5 outperformed all wavelet functions studied and is the best function for this experimental study.

Table 2. The SNR of wavelet families

Wavelet	Averaged SNR (dB)
db2	11.70
db4	11.58
db5	11.70
db6	11.58
db7	11.48

(Continued)

Table 2. The SNR of wavelet families *(Continued)*

Wavelet	Averaged SNR (dB)
sym5	11.81
sym6	11.60
sym7	11.69
sym8	11.58
bior1.3	11.22
bior3.5	11.18
bior6.8	11.52
coif4	11.51
coif5	11.54
haar	11.30

- **ECG Segmentation.** The aim of segmenting the ECG signal is to locate the temporal and morphological characteristics in order to detect patterns of AF, NSR, and non-AF. In our previous study [13, 14], we successfully segmented the ECG records into 2,700 nodes (each containing at least one RR-interval) for the classification of AF, NSR, and non-AF. To identify AF, the QRS complexes are irregularly irregular with varying RR intervals. The total number of ECG episodes, after being segmented into 2,700, can be listed in Table 3.

Table 3. The total of ECG episodes after segmentation preprocessing

ECG Dataset	Class	ECG Episodes (2,700 Nodes)
MIT-BIH Atrial Fibrillation	AF	262
The China Physiological Signal Challenge 2018	NSR	1,981
	AF	2,297
	Non-AF	11,443
Chapman University, Shaoxing People's Hospital	NSR	1,771
	AF	1,726
	Non-AF	6,824
The PhysioNet/Computing in Cardiology Challenge 2017	NSR	15,747
	AF	2,316
Total		44,367

3.3 Convolutional neural networks architecture

The 1D CNN is well-suited for real-time and low-cost applications in biomedical signal processing [12, 24–26]. The computational complexity of a 1D CNN is lower than that of a two-dimensional CNN. The configuration of a 1D CNN is determined by the following hyperparameters [27]: (i) the number of hidden CNN layers or neurons; (ii) the filter size in each CNN layer; (iii) the subsampling factor in each CNN layer; and (iv) the pooling and activation functions. In CNN layers, 1D forward propagation is defined as [28].

$$x_q^p = b_q^p + \sum_{i=1}^{N_{p-1}} conv1D(w_{iq}^{p-1}, s_i^{p-1}) \tag{1}$$

where x_q^p is defined as the input, b_q^p is defined as the bias of the q^{th} neuron at layer p, s_i^{p-1} is the output of the i^{th} neuron at layer $p-1$, and is the kernel from the i^{th} neuron at layer $p-1$ to the q^{th} neuron at layer p. The output y_q^p can be written from the input x_q^p as,

$$y_q^p = f(x_q^p) \tag{2}$$

and,

$$s_q^p = y_q^p \downarrow ss \tag{3}$$

where s_q^p for the neuron output and down-sampling operation with factor, ss.

In addition, for backward propagation, we assume the existence of an input layer and an output layer. Let N_P be the number of classes in the database. Then, for an input vector v, and its target and output vectors, t_i^v and $\left[y_1^P, \ldots, y_{N_P}^P \right]$, respectively.

The proposed architecture of the CNN can be found in Table 4 and Figure 2. It consists of 13 convolution layers (stride = 1) with rectified linear unit (ReLU) activation functions, five pooling layers (stride = 2), kernel sizes of 64, 128, 256, and 512, and a fully connected layer. The proposed architecture has also been implemented in our previous study for AF classification tasks [13, 14]. We present the pseudocode of the CNN architecture in Algorithm 1.

Table 4. The proposed CNN architecture

Layer	Types	Kernel Size	Stride
1	Conv + ReLU	3 × 64	1
2	Conv + ReLU Pooling	3 × 64 2	1 2
3	Conv + ReLU	3 × 128	1
4	Conv + ReLU Pooling	3 × 128 2	1 2
5	Conv + ReLU	3 × 256	1
6	Conv + ReLU	3 × 256	1
7	Conv + ReLU Pooling	3 × 256 2	1 2
8	Conv + ReLU	3 × 512	1
9	Conv + ReLU	3 × 512	1
10	Conv + ReLU Pooling	3 × 512 2	1 2
11	Conv + ReLU	3 × 512	1
12	Conv + ReLU	3 × 512	1
13	Conv + ReLU Pooling	3 × 512 2	1 2
14	Fully Connected	1000	–
15	Fully Connected	1000	–
16	Fully Connected	3	–

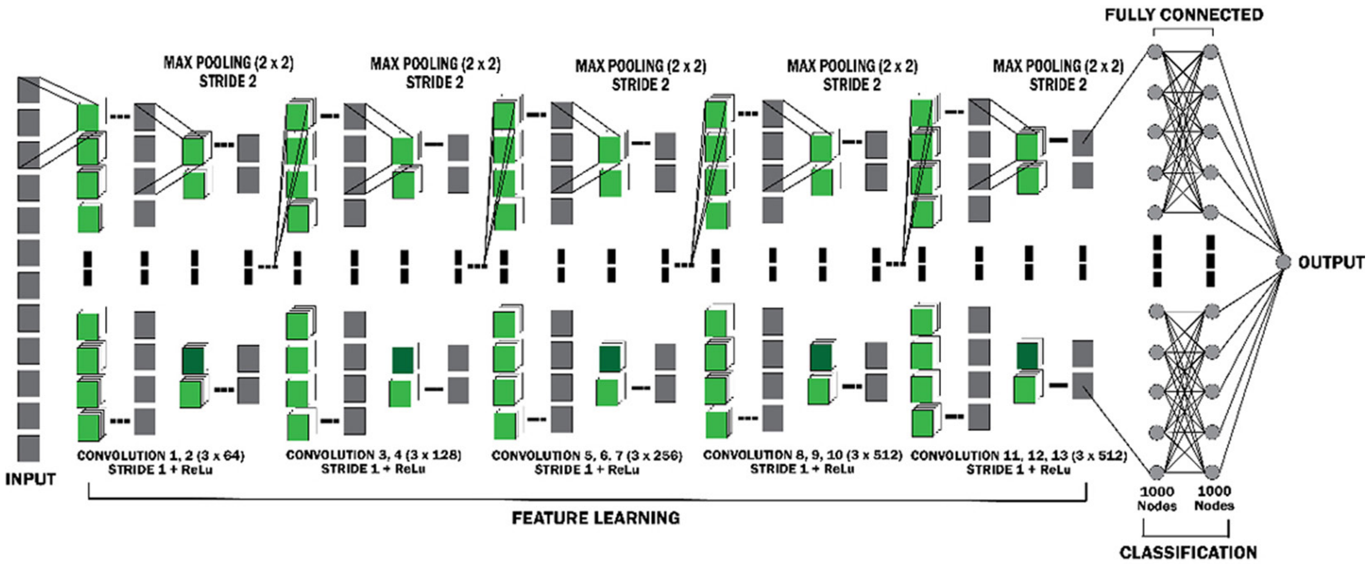


Fig. 2. The proposed CNN architecture

Algorithm 1: Pseudocode of CNN Architecture

Parameters: input x (2700, 1), output y (2700, 1)
1. For each epoch do:
    #CNN Feature Extraction
2.   For each convolution layer do:
3.   For each sample in x do:
4.   Calculate a_{ij}^{m} from x by Equation 1
5.   End for
6.   End for
    #Dimension of a is (2700, 512)
    #Max Pooling Layer
7.   For each sample in a do:
8.   Extract the maximum value
9.   End for

4 RESULTS AND DISCUSSION

We conducted an experiment using stratified K-fold cross-validation to divide the training and validation sets. It aims to reduce the variance of the performance estimate and allows you to use more data for training. In this study, we have implemented 10-fold cross-validation, which divides the dataset into ten random groups of equal size. To investigate and explore this experimental study, we aimed to empower the diagnosis of AF, NSR, and non-AF. We conducted the study based on several important cases: (i) we partitioned the dataset into balanced and imbalanced sets; (ii) we compared the effectiveness of the DWT and denoising autoencoder (DAE) for ECG denoising; and (iii) we trained the proposed model for AF, NSR, and other arrhythmia classification (15 classes).

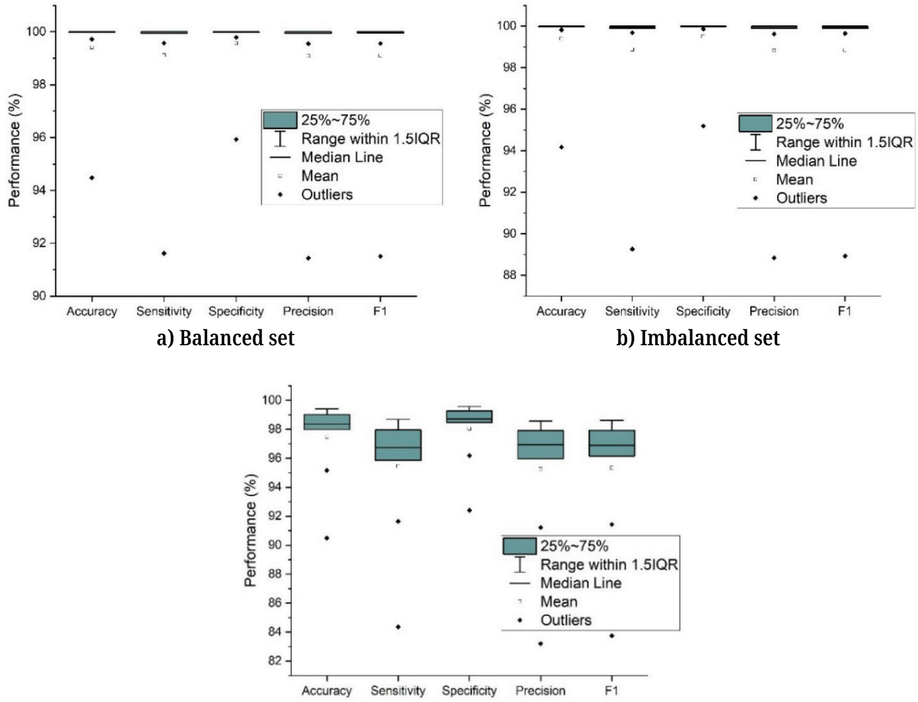
4.1 Case 1: The balanced and imbalanced dataset

This study has explored multiclass classification (AF, NSR, and non-AF) with varying numbers of total episodes in each class, which are imbalanced. In a classification task, having balanced data makes training a model easier because it helps prevent the model from becoming biased towards one class. Unlike cases with balanced data, imbalanced data is difficult to fix. The number of classes constitutes a significant portion of the dataset (majority class). The minority class constitutes a smaller proportion. Hence, it can lead to potential bias in the trained model. Table 5 presents the total number of episodes that have been categorized into balanced, imbalanced, and extremely imbalanced data. For a balanced set, we have divided the AF, NSR, and non-AF into 6,601, 7,818, and 6,824 episodes, respectively. For the imbalanced set, there are 6601 episodes of AF, 19499 episodes of NSR, and 6824 episodes of non-AF. Then, for the highly imbalanced set, there are 6601 AF, 19499 NSR, and 18267 non-AF episodes.

Table 5. The investigation of balanced, imbalanced, and extremely imbalanced set

ECG Dataset	Class	Number Episodes		
		Balanced Set	Imbalanced Set	Extreme Imbalanced Set
MIT-BIH Atrial Fibrillation	AF	262	262	262
The China Physiological Signal Challenge 2018	NSR	–	1,981	1,981
	AF	2,297	2,297	2,297
	Non-AF	–	–	11,443
Chapman University, Shaoxing People's Hospital	NSR	–	1,771	1,771
	AF	1,726	1,726	1,726
	Non-AF	6,824	6,824	6,824
The PhysioNet/Computing in Cardiology Challenge 2017	NSR	7,818	15,747	15,747
	AF	2,316	2,316	2,316
Total Episodes		21,243	32,924	44,367

The results for the balanced, imbalanced, and extremely imbalanced data can be presented in Figure 3. The boxplot is used to display the distribution of numerical data and skewness by showing the data quartiles, including the minimum, first quartile, median, third quartile, and maximum. Figures 3a–c show the performance results of the distribution for balanced, imbalanced, and extremely imbalanced sets, respectively. Figures 3a–c show that some outliers differ significantly from the performance results. Figures 3a and b for balanced and imbalanced data show that the performance can reach 100% accuracy, sensitivity, specificity, precision, and F1-score. Different from the extremely imbalanced set (refer to Figure 3c), the performance results were obtained at around 83%. The results perform poorly and are lower than those of the balanced and imbalanced sets due to the skewed or biased distribution of AF, NSR, and non-AF. The proposed algorithm is biased in favor of one class because it receives a disproportionately high number of samples from that class. It does not learn what distinguishes the other class and does not comprehend the underlying patterns that enable us to differentiate between AF, NSR, and non-AF classes.

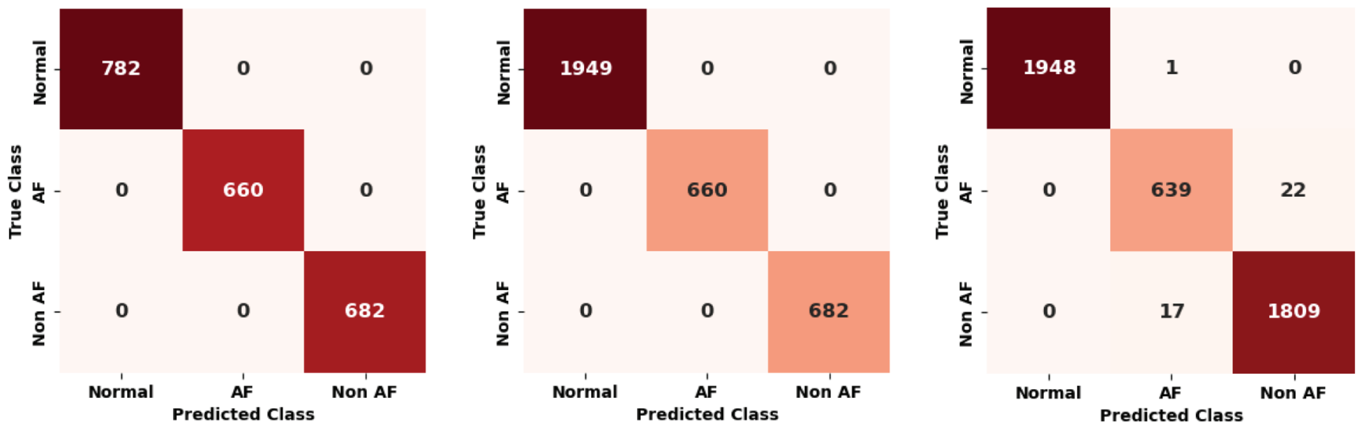


a) Balanced set b) Imbalanced set

c) Extremely imbalanced set

Fig. 3. Boxplot of the 10-fold cross-validation results for a balanced, imbalanced, and extremely imbalanced set

To visualize the performance of the proposed model on a set of validation data, a confusion matrix (CM) can be presented in Figure 4. Figure 4 shows the prediction results on the validation set for the balanced, imbalanced, and extremely imbalanced datasets. It can assess where false positives and negatives were made in the proposed model. Figures 4a and b show perfect classification, with no errors occurring. However, in the case of Figure 4c, there is one misclassified AF as NSR and 17 misclassified AF as non-AF. In addition, for the non-AF classification, there are 22 cases misclassified as AF. In our data experiment, the distribution of the non-AF class exhibits a variety of abnormality morphologies, with the majority of non-AF cases being comprised of other arrhythmia morphologies. The various abnormal morphologies of other arrhythmias belong to one class, non-AF. Hence, the misclassification mostly occurs from non-AF to AF. AF is the most common type of treated heart arrhythmia.



a) Balanced set a) Imbalanced set c) Extreme Imbalanced set

Fig. 4. Confusion matrix of the balanced, imbalanced, and extreme imbalanced set for AF classification

4.2 Case 2: Comparison of DWT and DAE for ECG denoising

To generalize the proposed experimental model, we compared the effectiveness of DWT and DAE for ECG denoising. In our previous works [29, 30], we explored the use of DAE for removing noise and artifacts from ECG signals. Denoising algorithms based on DL have been explored for performing ECG signal denoising [29–34]. DAE learns the parameters for different noisy conditions, which consist of encoding (lower-dimensional representation) and decoding layers (feature extraction). To ensure a fair comparison, we re-trained a proposed architecture using 13 convolutional layers and five max-pooling layers with the help of DAE. The selected mother function of DWT (*sym5*) is used for baseline SNR (target). The initial input x, is corrupted by a stochastic mapping $\tilde{x} \sim q(\tilde{x} \mid x)$. The DAE used corrupted data $\tilde{x}$ as input, which was initially transformed into a hidden representation using the encoder. The text was then reconstructed using a decoder. We trained a DAE to propose an architecture based on hyperparameter tuning (see Table 6). We have adjusted the number of epochs (100, 200, and 300) and the batch size (8, 16, 32, 64, 128, 256, and 512). Based on Table 6, the optimal parameters are 200 epochs, 256 batch sizes, and a learning rate of 0.00095. This conclusion is drawn from the discrepancy between the reconstructed value (34.14 dB), as well as the target value (36.11 dB) and the lowest loss value (0.02596050052903). The presentation of reconstructed (yellow color) and raw (blue color) signals can be seen in Figure 5. Figure 5 shows the morphology of the reconstructed signal near the raw signal.

Table 6. The hyperparameter tuning of the DAE model

Epoch	Batch Size	Target Value (dB)	Reconstructed Value (dB)	Loss Value
100	8	34.6125	17.6964	0.0445
	16	34.6125	27.8053	0.0291
	32	34.6125	17.6964	0.0445
	64	34.6125	17.6964	0.0445
	128	34.6125	28.2868	0.0349
	256	34.6125	17.6964	0.0439
	512	34.6125	17.6964	0.0431
200	8	34.6125	17.6964	0.0445
	16	34.6125	17.6964	0.0445
	32	34.6125	17.6964	0.0445
	64	34.6125	17.6964	0.0445
	128	34.6125	25.7426	0.0260
	256	**34.6125**	**29.0016**	**0.0260**
	512	34.6125	24.7656	0.0389
300	8	34.6125	17.6964	0.0445
	16	34.6125	17.6964	0.0445
	32	34.6125	17.6964	0.0445
	64	34.6125	17.6964	0.0445
	128	34.6125	38.9717	0.0202
	256	34.6125	17.6964	0.0439
	512	34.6125	17.6964	0.0440

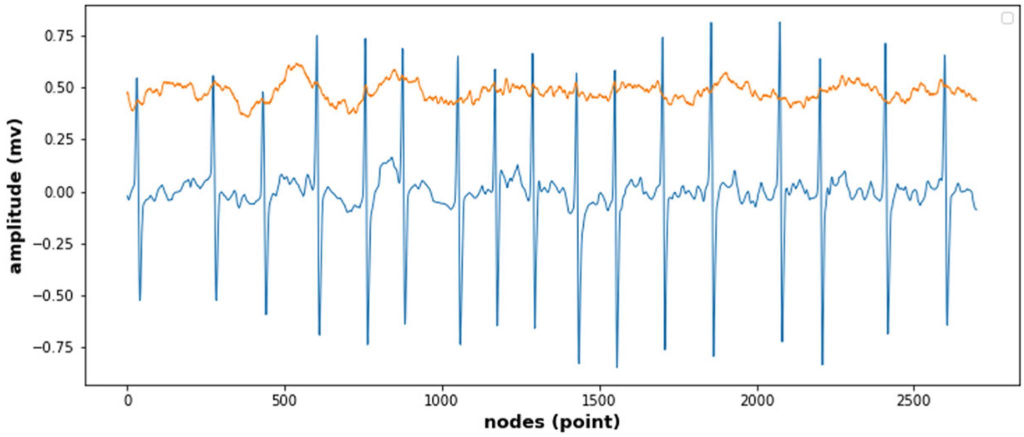


Fig. 5. The reconstructed signal using DAE

The performance results of DAE-CNN can be seen in Figure 6. Figure 6 shows the box plot of the first, second, and third quartiles for DAE-CNN. The average performance result is 94.78% for accuracy, sensitivity, specificity, precision, and F1-score. However, when comparing it to DWT-CNN, DWT-CNN outperformed DAE-CNN. The results of DAE-CNN achieve less than 50% sensitivity, precision, and F1-score. Different from DWT-CNN, the performance achieved an 83% accuracy, sensitivity, specificity, precision, and F1-score above 83%. With DAE, the noisy ECG signal is fed into the network and mapped into a lower-dimensional manifold, making noise filtering much more manageable. However, DAE is prone to a high risk of overfitting. Additionally, it is not effective in handling highly non-linear data. DWT is useful for representing the subtle variations in the signal $f(t)$ at different scales. Additionally, the function $f(t)$ can be expressed as a linear combination of functions that represent variations at different scales.

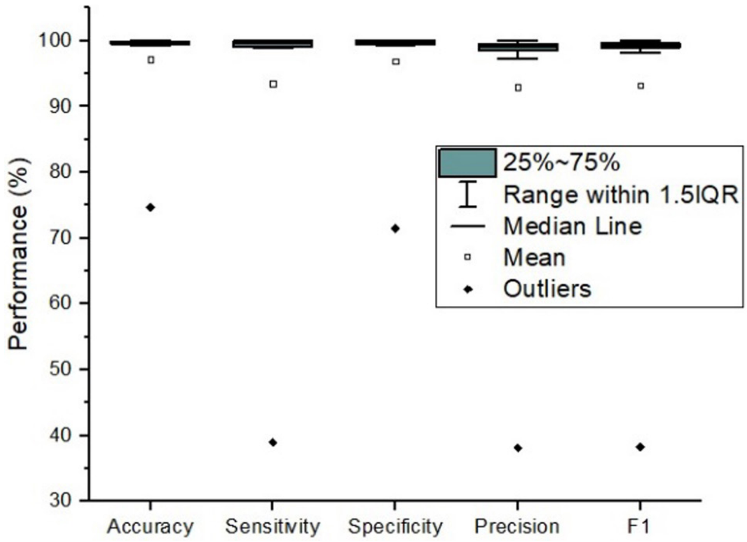


Fig. 6. Box plot of DAE-CNN performance results

4.3 Case 3: AF, NSR and other arrhythmia classification (15-classes)

To empower our AI-enabled ECG, we trained the proposed model to classify NSR, AF, SB, ST, AFL, SI, SVT, AT, IAVB, LBBB, RBBB, PAC, PVC, STD, and STE (15 classes).

We have specified the non-AF class to include specific arrhythmia types. As we mentioned earlier, AF often goes unnoticed because it typically does not cause symptoms, and it is difficult to differentiate an irregular heartbeat from an NSR or other types of arrhythmias. The performance results of each class in a 10-fold cross-validation strategy are shown in Figure 7. Figure 7 visualizes the varying results of accuracy, sensitivity, specificity, precision, and F1-score from a 10-fold classification of 15 classes, including NSR, AF, and other arrhythmia types. There are some outlier values between 68% and 85% for the performance metrics used. With the excellence of our proposed model, we achieved an average (second quartile) of 96.58% accuracy, 94.21% sensitivity, 97.04% specificity, 93.81% precision, and a 93.94% F1 score. The experiments concluded that the performance of multiclass classification reached remarkable results for AF diagnosis in clinical practice.

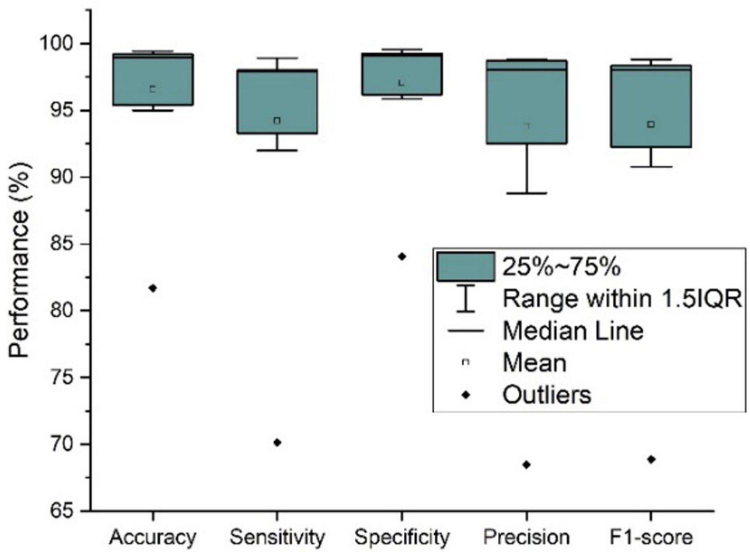


Fig. 7. Boxplot of NSR, AF and other arrhythmia classification (15-classes) performance

The visualization of CM for NSR, AF, and other arrhythmia classifications (15 classes) can be presented in Figure 8. Figure 8 plots the distribution of the number of classes that were successfully classified as actual classes (ground truth). The misclassification mostly occurs with LBBB and RBBB, which are wrongly classified as AF, PAC, PVC, STD, and STE. LBBB may be due to conduction system degeneration, while RBBB represents the abnormality of the human heart in the intraventricular electrical conduction system. In a specific case, confirming the onset of LBBB or RBBB can be challenging if no prior ECG exists. The morphologies of LBBB, RBBB, AF, PAC, PVC, STD, and STE tend to be similar due to the prevalence of LBBB, which increases ventricular myocardial infarction, and RBBB, which is associated with arterial hypertension.

Although the results look promising, there are some limitations to our study for possible avenues for future research and improvements. First, this study only focused on classifying NSR, AF, and non-AF. We did not detect episodes of AF and NSR. Second, more datasets can be generated to achieve more robust performance. Third, though the results were well-performed, we have only generalized the proposed DL model to validation data, not to testing data. Testing data can be applied to evaluate the performance and optimize it for improved results.

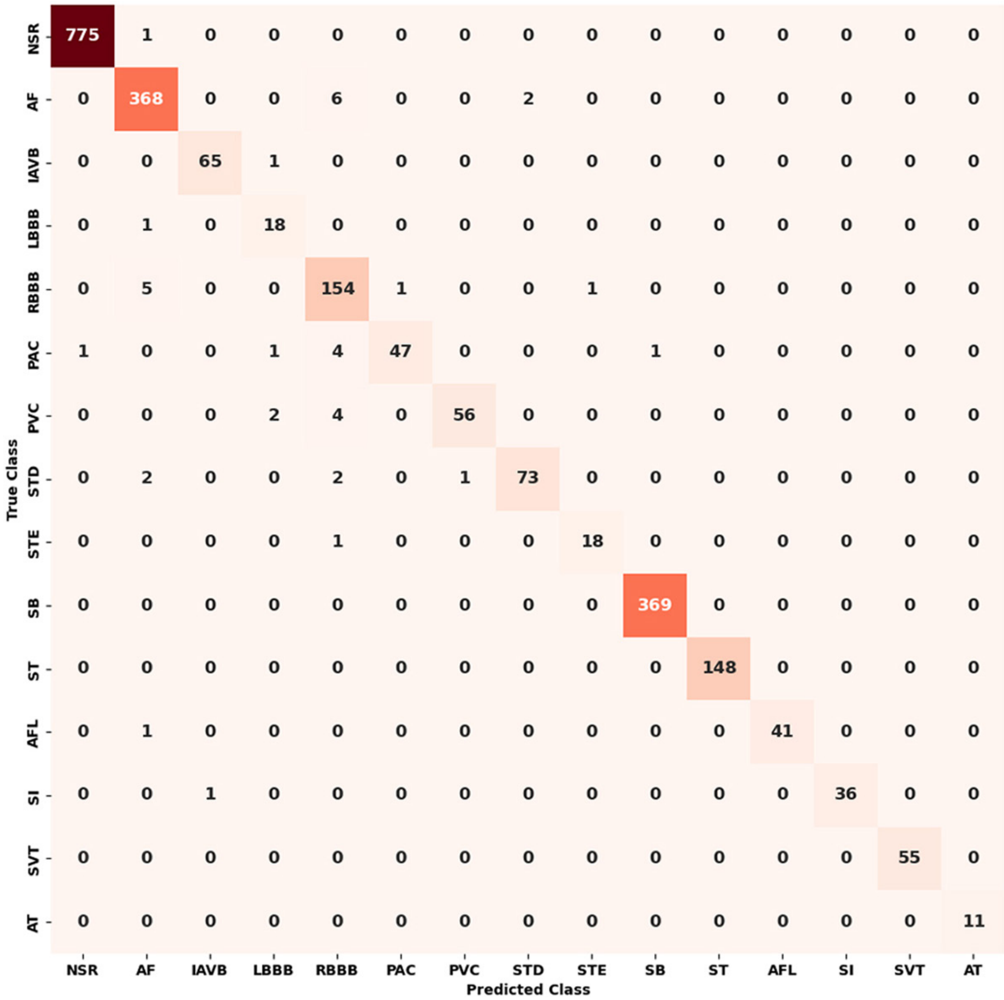


Fig. 8. CM of NSR, AF and other arrhythmia classifications (15-classes)

5 CONCLUSION

A powerful AI-enabled ECG based on the CNN algorithm achieves outstanding performance with varying large amounts of ECG datasets publicly. In this study, the proposed architecture of CNN deals with three cases: (i) a balanced, imbalanced, and extremely imbalanced set, (ii) a comparison of the ECG denoising algorithm; and (iii) the classification of AF, NSR, and other arrhythmia types (15 classes). As a result of a 10-fold cross-validation strategy, the following findings were obtained: (i) for AF, NSR, and non-AF classification, we achieved 100% accuracy, sensitivity, specificity, precision, and F1-score for both balanced and imbalanced sets; (ii) DWT outperformed DAE in this study for ECG denoising; and (iii) for the classification of AF, NSR, and other arrhythmia types (15 classes), the performance results were 99.77% accuracy, 96.48% sensitivity, 99.87% specificity, 97.03% precision, and 96.68% F1-score. We conclude that our AF classification experiment has successfully applied of AI to healthcare.

6 ACKNOWLEDGMENTS

We thank the Intelligent System Research Group (ISysRG), Faculty of Computer Science, Universitas Sriwijaya, Indonesia. This work was also supported by Universitas Sriwijaya funded by the Competitive Grant 2023.

7 REFERENCES

[1] D. Kleyko, E. Osipov, and U. Wiklund, "A comprehensive study of complexity and performance of automatic detection of atrial fibrillation: Classification of long ECG recordings based on the PhysioNet computing in cardiology challenge 2017," *Biomedical Physics & Engineering Express*, vol. 7, no. 2, p. 025010, 2020. https://doi.org/10.1088/2057-1976/ab6e1e

[2] K. Harris, D. Edwards, and J. Mant, "How can we best detect atrial fibrillation?" *The Journal of the Royal College of Physicians of Edinburgh*, vol. 42, no. 18, pp. 5–22, 2012. https://doi.org/10.4997/JRCPE.2012.S02

[3] B. M. Mathunjwa, Y. T. Lin, C. H. Lin, M. F. Abbod, and J. S. Shieh, "ECG arrhythmia classification by using a recurrence plot and convolutional neural network," *Biomedical Signal Processing and Control*, vol. 64, p. 102262, 2021. https://doi.org/10.1016/j.bspc.2020.102262

[4] S. Mousavi, F. Afghah, and U. R. Acharya, "HAN-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks," *Computers in Biology and Medicine*, vol. 127, p. 104057, 2020. https://doi.org/10.1016/j.compbiomed.2020.104057

[5] S. Asgari, A. Mehrnia, and M. Moussavi, "Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine," *Computers in Biology and Medicine*, vol. 60, pp. 132–142, 2015. https://doi.org/10.1016/j.compbiomed.2015.03.005

[6] U. Desai, R. J. Martis, C. G. Nayak, K. Sarika, and G. Seshikala, "Machine intelligent diagnosis of ECG for arrhythmia classification using DWT, ICA and SVM techniques," in *2015 Annual IEEE India Conference (INDICON)*, 2015, pp. 1–4. https://doi.org/10.1109/INDICON.2015.7443220

[7] K. Y.-K. Liao, C.-C. Chiu, and S.-J. Yeh, "A novel approach for classification of congestive heart failure using relatively short-term ECG waveforms and SVM classifier," in *Proceedings of the International Multiconference of Engineers and Computer Scientists*, IMECS March 2015, Hong Kong, 2015, pp. 47–50.

[8] M. R. Homaeinezhad, S. A. Atyabi, E. Tavakkoli, H. N. Toosi, A. Ghaffari, and R. Ebrahimpour, "ECG arrhythmia recognition via a neuro-SVM–KNN hybrid classifier with virtual QRS image-based geometrical features," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2047–2058, 2012. https://doi.org/10.1016/j.eswa.2011.08.025

[9] S. Hadiyoso, S. Aulia, I. D. Irawati, and M. Ramdhani, "Multi-Distance dispersion entropy for ECG signal classification," *International Journal of Online and Biomedical Engineering*, vol. 18, no. 7, pp. 151–160, 2022. https://doi.org/10.3991/ijoe.v18i07.30055

[10] C. Chen, Z. Hua, R. Zhang, G. Liu, and W. Wen, "Automated arrhythmia classification based on a combination network of CNN and LSTM," *Biomedical Signal Processing and Control*, vol. 57, pp. 1–10, 2020. https://doi.org/10.1016/j.bspc.2019.101819

[11] S. L. Oh, E. Y. K. Ng, R. San Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Computers in Biology and Medicine*, vol. 102, pp. 278–287, 2018. https://doi.org/10.1016/j.compbiomed.2018.06.002

[12] G. Petmezas *et al.*, "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets," *Biomedical Signal Processing and Control*, vol. 63, p. 102194, 2021. https://doi.org/10.1016/j.bspc.2020.102194

[13] B. Tutuko *et al.*, "AFibNet: An implementation of atrial fibrillation detection with convolutional neural network," *BMC Medical Informatics and Decision Making*, vol. 1, pp. 1–17, 2021. https://doi.org/10.1186/s12911-021-01571-1

[14] S. Nurmaini *et al.*, "Robust detection of atrial fibrillation from short-term electrocardiogram using convolutional neural networks," *Future Generation Computer Systems*, vol. 113, pp. 304–317, 2020. https://doi.org/10.1016/j.future.2020.07.021

[15] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification," *Expert Systems with Applications: X*, vol. 7, p. 100033, 2020. https://doi.org/10.1016/j.eswax.2020.100033

[16] Q. H. Nguyen, B. P. Nguyen, T. B. Nguyen, T. T. T. Do, J. F. Mbinta, and C. R. Simpson, "Stacking segment-based CNN with SVM for recognition of atrial fibrillation from single-lead ECG recordings," *Biomedical Signal Processing and Control*, vol. 68, p. 102672, 2021. https://doi.org/10.1016/j.bspc.2021.102672

[17] H. A. Serhal, N. K. Abdallah, J.-M. Marion, P. Chauvet, M. Oueidat, and A. Humeau-Heurtier, "An EMD-based approach for atrial fibrillation classification using wavelets and convolutional neural network," *SSRN Electronic Journal*, 2022. https://doi.org/10.2139/ssrn.4029430

[18] F. Ma, J. Zhang, W. Chen, W. Liang, and W. Yang, "An automatic system for atrial fibrillation by using a CNN-LSTM Model," *Discrete Dynamics in Nature and Society*, vol. 2020, pp. 1–9, 2020. https://doi.org/10.1155/2020/3198783

[19] G. Moody, "A new method for detecting atrial fibrillation using RR intervals," *Computers in Cardiology*, pp. 227–230, 1983.

[20] J. Zheng, H. Guo, and H. Chu, "A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0. 0)," *PhysioNet,* 2022. Available online: http://physionet.org/content/ecg-arrhythmia/1.0.0/. [Accessed Nov. 23, 2022].

[21] G. D. Clifford *et al.*, "AF classification from a short single lead ECG recording: The PhysioNet/Computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4. https://doi.org/10.22489/CinC.2017.065-469

[22] F. Liu *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, 2018. https://doi.org/10.1166/jmihi.2018.2442

[23] M. Alfaouri and K. Daqrouq, "ECG signal denoising by wavelet transform thresholding," *American Journal of Applied Sciences,* vol. 5, no. 3, pp. 276–281, 2008. https://doi.org/10.3844/ajassp.2008.276.281

[24] Y. Li, Y. Pang, J. Wang, and X. Li, "Patient-specific ECG classification by deeper CNN from generic to dedicated," *Neurocomputing*, vol. 314, pp. 336–346, 2018. https://doi.org/10.1016/j.neucom.2018.06.068

[25] D. Jia *et al.*, "An electrocardiogram delineator via deep segmentation network," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, Berlin, Germany, 2019, pp. 1913–1916. https://doi.org/10.1109/EMBC.2019.8856987

[26] A. Peimankar and S. Puthusserypady, "DENS-ECG: A deep learning approach for ECG signal delineation," *Expert Systems with Applications*, vol. 165, p. 113911, 2021. https://doi.org/10.1016/j.eswa.2020.113911

[27] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, pp. 1–21, 2021. https://doi.org/10.1016/j.ymssp.2020.107398

[28] S. Kiranyaz, T. Ince, O. Abdeljaber, O. Avci, and M. Gabbouj, "1-D convolutional neural networks for signal processing applications," in *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 8360–8364. https://doi.org/10.1109/ICASSP.2019.8682194

[29] B. Tutuko *et al.*, "DAE-ConvBiLSTM: End-to-end learning single-lead electrocardiogram signal for heart abnormalities detection," *PLoS One*, vol. 17, no. 12, p. e0277932, 2022. https://doi.org/10.1371/journal.pone.0277932

[30] S. Nurmaini *et al.*, "Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification," *Electronics*, vol. 9, no. 1, p. 135, 2020. https://doi.org/10.3390/electronics9010135

[31] K. Ochiai, S. Takahashi, and Y. Fukazawa, "Arrhythmia detection from 2-lead ECG using convolutional denoising autoencoders," 2018.

[32] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu, "ECG signal enhancement based on improved denoising auto-encoder," *Engineering Applications of Artificial Intelligence*, vol. 52, pp. 194–202, 2016. https://doi.org/10.1016/j.engappai.2016.02.015

[33] P. Liu, P. Zheng, and Z. Chen, "Deep learning with stacked denoising auto-encoder for short-term electric load forecasting," *Energies*, vol. 12, no. 12, p. 2445, 2019. https://doi.org/10.3390/en12122445

[34] P. Xiong, H. Wang, M. Liu, and X. Liu, "Denoising autoencoder for eletrocardiogram signal enhancement," *Journal of Medical Imaging and Health Informatics*, vol. 5, no. 8, pp. 1804–1810, 2015. https://doi.org/10.1166/jmihi.2015.1649

8 AUTHORS

Bambang Tutuko is working as a Lecturer and also is researcher at the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes Robotics, Deep Learning, and Machine Learning (E-mail: bambangtutuko60@gmail.com).

Annisa Darmawahyuni is a working as a Lecturer and also a researcher at the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interest includes Biomedical Signal Processing, Deep Learning, and Machine Learning (E-mail: riset.annisadarmawahyuni@gmail.com).

Siti Nurmaini is working as a Professor at the Faculty of Computer Science, Universitas Sriwijaya and IEEE Member. She was received her Master's degree in Control system, Institut Teknologi Bandung – Indonesia (ITB), in 1998, and the PhD degree in Computer Science, Universiti Teknologi Malaysia (UTM), at 2011. Her research interest includes Biomedical Engineering, Deep Learning, Machine Learning, Image Processing, Control systems, and Robotic (E-mail: sitinurmaini@gmail.com).

Muhammad Naufal Rachmatullah is working as a Lecturer and also a researcher at the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes Medical Imaging, Biomedical Signal Processing, Deep Learning, and Machine Learning (E-mail: naufalrachmatullah@gmail.com).

Firdaus Firdaus is working as a Lecturer and also a researcher at the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes Text processing, Deep Learning, and Machine Learning (E-mail: virdauz@gmail.com).

Sutarno Sutarno is working as a Lecturer and also a researcher at the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes Text processing, Deep Learning, and Machine Learning (E-mail: sutarno@unsri.ac.id).

Rossi Passarella is working as a Lecturer and also a researcher at the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes Text processing, Deep Learning, and Machine Learning (E-mail: passarella.rossi@unsri.ac.id).

Ade Iriani Sapitri is working as a Lecturer and also a researcher at the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interest includes Medical Imaging, Deep Learning, and Machine Learning (E-mail: adeirianisapitri13@gmail.com).

Anggun Islami is working as a Lecturer and also a researcher at the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interest includes Medical Imaging, Deep Learning, and Machine Learning (E-mail: anggunislami2@gmail.com).

Muhammad Wahyu Ramansyah is an undergraduate student of the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes Signal Processing, Deep Learning, and Machine Learning (E-mail: wahyu.ramansyah26@gmail.com).

Muhammad Isra Al Hadi is an undergraduate student of the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interest includes Signal Processing, Deep Learning, and Machine Learning (E-mail: israalhadi01@gmail.com).