

Recognition and classification of electrocardiogram signal based on convolutional neural network

Yuzhong Xia

Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710000, China

xiayuzhong23@stu.xjtu.edu.cn

Abstract. Nowadays, the prevalence of cardiovascular diseases is increasing, leading to a rise in mortality rates. The electrocardiogram (ECG) is a common way to track activity of heart, some cardiovascular diseases including arrhythmia can be diagnosed with the ECG. However, diagnosing cardiac arrhythmias still heavily relies on the subjective judgment of cardiovascular specialists based on ECG readings. This project proposes a model that utilizes machine learning algorithms for the classification and processing of ECG signals. After collecting relevant data and performing necessary data preprocessing steps, the Convolutional Neural Network (CNN) model is employed for recognizing and classifying. Experimental findings demonstrate that the proposed method performs excellently in classifying various common types of cardiac arrhythmias, yielding high accuracy of over 99%. The method exhibits fast processing speeds, making it suitable for real-time ECG analysis and monitoring. Therefore, the outcomes of this project hold significant clinical relevance in improving the diagnosis of cardiovascular diseases and enhancing patients' quality of life.

Keywords: Convolutional Neural Network, residual network, electrocardiogram signal recognition.

1. Introduction

Cardiovascular diseases (CVDs) are the top cause of mortality worldwide, with over 17 million people dying from heart disease annually [1]. Patients with heart disease exhibit abnormal heart rhythm patterns known as Arrhythmias. Common manifestations of Arrhythmias include atrial fibrillation, ventricular fibrillation, tachycardia, and premature beats, and can be grouped into life-threatening and non-life-threatening [2,3]. The diagnosis of Arrhythmias primarily relies on cardiac experts' visual judgment of ECGs [4]. Although the widely used 24-hour Holter monitoring, which records long-term ECG signals, is prevalent, it lacks the ability to process data, automatically classify signals, and identify medically significant pathological signals. It requires waiting for the end of the 24-hour monitoring period for data analysis and conclusions by medical professionals. Therefore, real-time and accurate detection of patients' Arrhythmias is paramount in preventing heart disease and sudden cardiac death.

This paper proposes an automatic classification and recognition method for ECG signals based on CNN. The aim is to classify ECG signals accurately and effectively in real-time, distinguishing between normal and four types of arrhythmic signals, including Atrial Premature Contraction (APC), Ventricular Premature Contraction (VPC), Left Bundle Branch Block (LBBB) and Right Bundle Branch Block

(RBBB). By monitoring high-risk groups, subgroups, and undetermined conditions in cardiovascular disease, intelligent analysis of ECG variations during normal daily life, work, and activities can help alert patients to determine their condition or obtain potential cardiac ECG information. The approach of this paper is as follows: Firstly, in Chapter 2, the source of the dataset and methods for data preprocessing are presented. Subsequently, in Chapter 3, the CNN model structure used in this paper is briefly described. Finally, in Chapter 4, the test results after model training are presented and analyzed.

2. ECG Signal Preprocessing

2.1. Data source

The data utilized in the paper sourced from the MIT-BIH database, which is widely recognized and extensively employed within the academic community [5]. This database, which contains a wide variety of ECG signals and a sizable amount of data, offers useful experimental information for the automatic classification of cardiac signals examined in the paper.

The MIT-BIH Arrhythmia Database consists of 48 ECG recordings, and each recording lasts 30 minutes. These recordings are obtained from 47 research subjects, including 25 males and 22 females, with ages ranging from 23 to 89 years (note that records 201 and 202 originate from the same individual). The signals are sampled at a rate of 360 Hz, with an AD resolution of 11 bits. Each record contains two channels of signals, with the first channel typically representing the MLII lead (records 102 and 104 represent the V5 lead) and the second channel representing the V1 lead. This paper selects the MLII lead ECG signals for research analysis.

2.2. Pre-processing method

2.2.1. Noise in ECG signals. ECG signals are characterized by their low amplitude, low frequency, and susceptibility to noise interference. Noise sources can originate from within the body, such as respiration, muscle tremors, and external interference due to poor electrode contact. The main types of noise interference in ECG signals are powerline interference, electromyographic (EMG) interference, and baseline drift, which need to be suppressed and removed during the filtering process.

Powerline interference is caused by electromagnetic interference from the power supply environment surrounding the ECG recording equipment. It is characterized by low amplitude and a noise frequency of approximately 50 Hz, resembling a sinusoidal waveform. This noise often masks useful ECG signals and can also affect the detection of P and T waves.

EMG interference occurs during ECG signal acquisition due to involuntary muscle tremors. This type of interference is irregular, with rapidly changing waveform morphology, a high-frequency range (0-2000 Hz), and a wide distribution. Its energy is concentrated within 30-300 Hz range, typically around 50 ms. The overlap between EMG interference and the ECG signal can make subtle changes in useful ECG signals easily overlooked.

Baseline drift is a low-frequency interference with a frequency range of 0.15-0.3 Hz. It occurs when the electrode position slides or due to respiratory movements, causing the ECG signal to deviate from its normal baseline position over time gradually. Both the amplitude and frequency of the baseline drift can vary continuously, and it particularly affects the PR and ST segments in the ECG signal, leading to distortion.

2.2.2. Removal of noise. Therefore, the first step in the classification and recognition of ECG signals is signal pre-processing. This paper employed a denoising approach based on wavelet transform to address noisy ECG signals [6]. The Wavelet Transform is an ideal tool for time-frequency analysis and allows for the analysis of signals in both the time and frequency domains. The denoising process consisted of the following three steps:

Firstly, due to the mixture of noise and signal in ECG signals, a wavelet basis function (specifically, the db5 wavelet in this paper) was selected. The ECG signal was decomposed into wavelet coefficients

at various scales using the wavelet transform. Secondly, after the ECG signal underwent wavelet transform decomposition, the wavelet coefficients with larger amplitudes represented useful signal components, while those with smaller amplitudes corresponded to noise. Based on the frequency distribution of the ECG signal and the noise interference, a thresholding process is applied to the wavelet coefficients at each scale. Coefficients below the threshold were set to zero or processed using a thresholding function. Thirdly, the low-frequency coefficients and high-frequency coefficients obtained from the wavelet scale decomposition were processed separately, followed by signal reconstruction.

Figure 1 present a comparison of selected ECG signals before and after denoising clearly demonstrating significant noise suppression. (a) shows original signals and (b) shows signals after denoising.

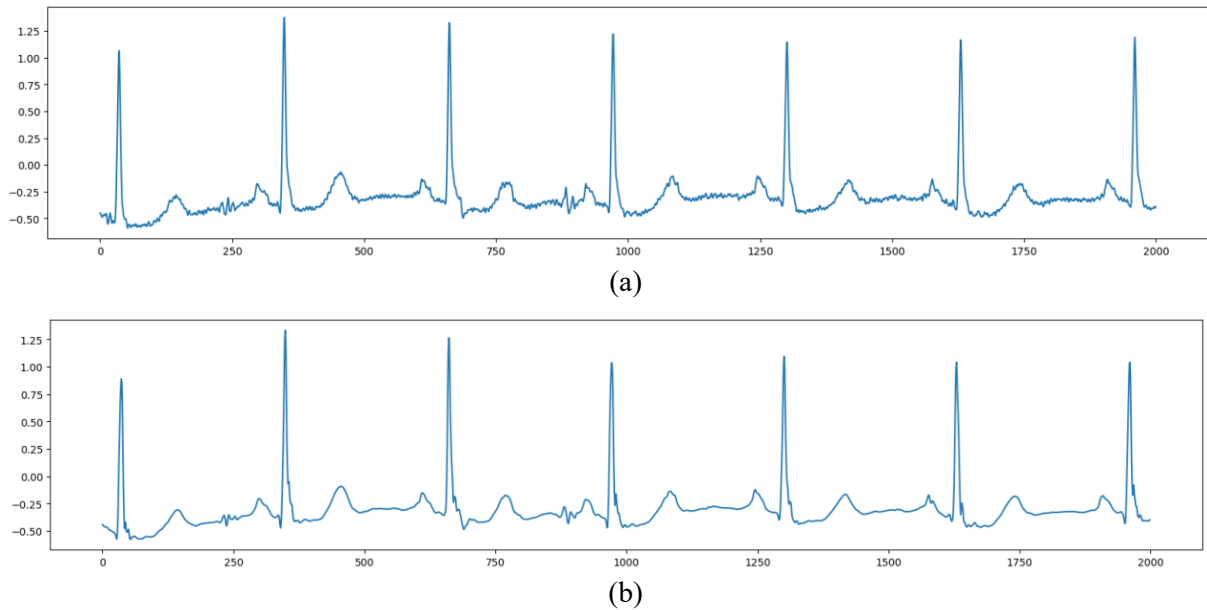


Figure 1. Comparison between ECG signals before and after denoising.

3. CNN Model Construction and Training

3.1. Data set construction

To construct the dataset, it is necessary to partition the pre-processed ECG signals from the previous steps into individual heartbeats that meet the desired criteria, which will serve as samples. The segmentation of heartbeats requires the localization of the QRS wave peak. In this paper, manual annotations provided by the MIT-BIH dataset were utilized (In practical detection, the automatic localization of QRS peak can be achieved using the Pan-Tompkins algorithm) [7]. Specifically, 99 signal points preceding the peak and 200 signal points following the peak are extracted to form a complete heartbeat.

After partitioning heartbeats, all heartbeat signals (excluding signals 102 and 104 without MLII lead) are read into the data set and the label set. Then generating a random permutation list of 92,192 (total number of heartbeats), truncate the first 30% of the values as indexes, and then extract the values subscript to those indexes from the data set and the label set to get the test data set and the test label set, the remaining data is used as a training data set and a training label set.

3.2. Model construction

This paper employed Convolutional Neural Network (CNN) for ECG signal classification, the artificial neural network that is widely used nowadays [8]. The Convolutional Neural Network mainly consists of

3 different layers, including convolutional layer, max-pooling layer, and fully-connected layer [9,10]. The key to automatic feature extraction for CNN lies in the convolutional operation. Through multiple convolution and pooling layers, lower-level features gradually form higher-level features, and the neural network performs classification based on the extracted high-level features. Compared with traditional machine learning approaches that require manual feature extraction by experts or designers based on the specific characteristics of the data, deep learning has the capability to automatically extract distinctive features from each class of data [9].

Unlike images or audio, ECG signals contain relatively fewer features and information. Therefore, deep networks may not accurately classify ECG signals. Hence, in this paper, four models with varying depths were constructed to classify ECG signals using the dataset constructed in previous steps. The goal is to identify the model with the highest accuracy in ECG signal classification and compare the distinct characteristics of different depth models in classifying ECG signals. Below are the detailed structures of the four models mentioned.

Model 1: This model consists of a total of 5 layers, including 2 convolutional layers, 1 max-pooling layer, and 2 fully-connected layers. Each convolution layer (layer 1 and 3) is convolved with the kernel size of 26 and 16, and the stride for convolution is set at 1. In layer 2, there is a max pooling which can reduce the size of the feature map. The kernel size of the max-pooling is 3, and the stride is 4. The activation function of layer 1, 3, 4, and 5 is ReLU [11]. The number of output neurons in the fully-connected layers is 128 and 5, and in the last layer, softmax function is used to separate the output into classes named N, A, V, L, and R.

Model 2: This model consists of a total of 7 layers, including 4 convolutional layers, 1 max-pooling layer, and 2 fully-connected layers. layers 1, 2, 4, and 5 are convolutional layers, and their kernel sizes are 26, 26, 16, and 16. The stride for convolution is 1. The max-pooling layer is in layer 3, and the kernel size is 3, the stride is 4. The structure of fully-connected layers is the same as in model 1.

Model 3: This model consists of a total of 10 layers, including 6 convolutional layers, 2 max-pooling layer, and 2 fully-connected layers. Layers 1, 2, 4, 5, 7, and 8 are convolutional layers whose kernel sizes are 26, 26, 16, 16, 11, and 11. The stride for convolution is 1. The max-pooling layers are in layer 3 and 6, the kernel size is 3, the strides are 3 and 2, respectively. The structure of fully-connected layers is the same as in model 1.

Model 4: A residual network structure is employed in constructing this model to investigate the classification capability of deeper networks for ECG signals [12]. Model 4 emulates the ResNet18 residual network and consists of 8 residual blocks followed by 2 fully connected layers [13]. Each residual block comprises 2 convolutional layers (with a kernel size of 3) and a shortcut connection. The stride of the first convolutional layer of the third, fifth, and seventh residual block is 2, while the stride of other convolutional layers is set at 1. And the structure of fully connected layers is the same as in model 1.

4. Results

After 30 rounds of training on 4 models, the performance of each model is shown in Table 1. The change process of accuracy and loss for each model with the training set and testing set is shown in Figure 2. Table 2-5 present the confusion matrix of each model in the testing set.

From Table 2-5, all four models achieved high recognition accuracy for ECG signals in classes N, V, L, and R, with recognition rates ranging from 98% to 99%. However, the recognition accuracy for class A ECG signals was comparatively lower, ranging from 85% to 90%. This solution can be attributed to the fact that class A, which represents APCs, exhibits ECG patterns that are more similar to normal ECG signals, making them challenging to distinguish. Additionally, the relatively small proportion of class A samples in the data set may have limited the models' ability to learn and generalize adequately. And from Table 1 and Figure 2, the differences between each model can be summarized as follows.

Table 1. The comparative analysis of 4 models.

Model	Time taken for each step(ms)	Minimum loss with testing set (%)	Maximum accuracy with testing set(%)
1	40	2.95	99.42
2	175	2.62	99.52
3	250	2.59	99.49
4	110	2.96	99.41

Table 2. The confusion matrix of model1 in the testing set.

Original	Predicted					acc(%)	ppv(%)	sen(%)	spec(%)
	N	A	V	L	R				
N	21368	45	32	2	1	99.56	99.56	99.75	99.71
A	54	554	3	0	0	89.26	87.31	91.17	90.27
V	17	0	2094	1	0	98.58	99.86	99.03	98.48
L	3	0	3	1949	0	99.85	99.85	99.85	99.85
R	3	5	0	0	1524	99.61	99.61	99.93	99.67

Table 3. The confusion matrix of model2 in the testing set.

Original	Predicted					acc(%)	ppv(%)	sen(%)	spec(%)
	N	A	V	L	R				
N	21358	43	44	2	1	99.52	99.56	99.75	99.71
A	51	556	4	0	0	89.44	88.21	91.16	90.36
V	10	2	2100	0	0	98.89	99.53	99.34	98.99
L	2	0	1	1952	0	99.65	99.95	99.95	99.69
R	3	4	2	0	1523	99.43	99.51	99.61	99.74

Table 4. The confusion matrix of model3 in the testing set.

Original	Predicted					acc(%)	ppv(%)	sen(%)	spec(%)
	N	A	V	L	R				
N	21407	20	19	2	0	99.68	99.88	99.93	99.93
A	78	529	2	0	2	86.46	86.63	86.56	85.03
V	21	0	2089	2	0	98.52	98.98	98.95	98.22
L	4	0	0	1951	0	99.77	100	99.95	100
R	7	2	0	0	1523	99.39	99.51	99.48	99.74

Table 5. The confusion matrix of model4 in the testing set.

Original	Predicted					acc(%)	ppv(%)	sen(%)	spec(%)
	N	A	V	L	R				
N	21405	18	17	6	2	99.72	99.89	99.91	99.91
A	73	536	1	0	1	85.94	86.15	86.03	85.34
V	26	1	2082	2	1	98.16	98.71	98.64	97.87
L	2	0	0	1953	0	99.87	100	99.95	100
R	3	3	0	0	1526	99.45	99.74	99.67	99.74

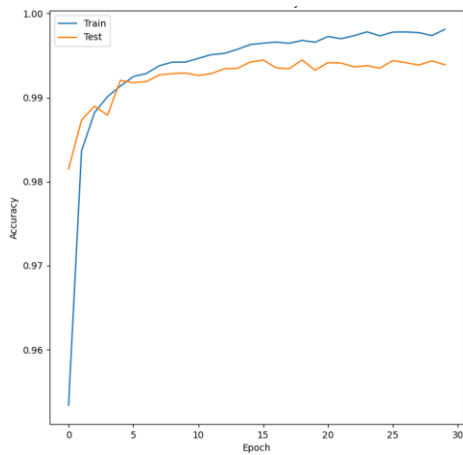
In these tables, the accuracy is acc. The positive predictive value is ppv. The sensitivity is sen and the specificity is spec.

Model 1 demonstrates the advantage of having fewer parameters, leading to reduced computational requirements and the highest recognition speed. It is capable of extracting basic features from the input data. However, this model exhibits limited ability to learn complex features and may not achieve high accuracy when dealing with complex tasks.

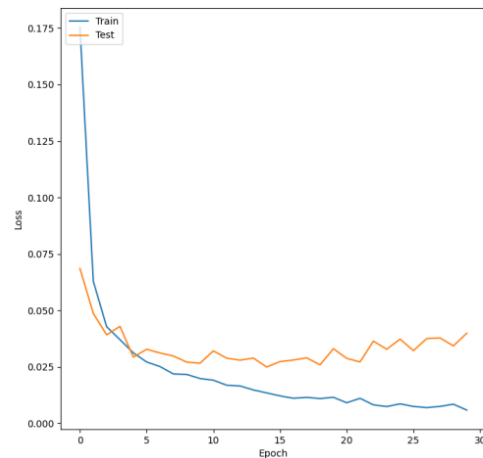
Model 2 has the ability to learn more complex and abstract features, resulting in higher accuracy for complex tasks. Nevertheless, this model requires a larger number of parameters and computational resources, resulting in a significantly lower recognition speed compared to model 1.

Model 3 has the capability to extract more complex and abstract features, leading to higher accuracy in handling complex tasks. However, due to the relatively limited features and information in ECG signals, this model is prone to overfitting, resulting in lower recognition accuracy than model 2. Additionally, model 3 exhibits the slowest recognition speed among the four models.

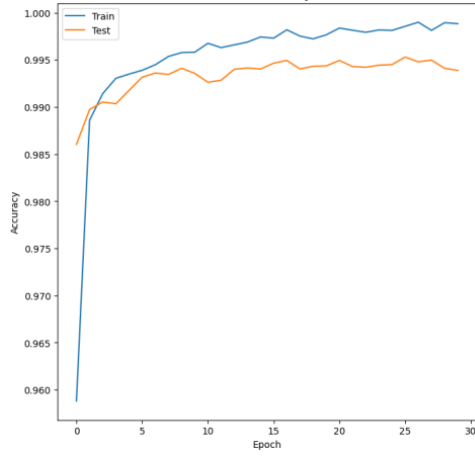
Model 4, as a deeper residual network, possesses the ability to train deep neural networks effectively and facilitates the learning of complex features and patterns. In comparison to conventional CNNs, residual networks demonstrate a relatively faster recognition speed, second only to model 1. However, in the task of ECG signal classification, model 4 faces overfitting issues on smaller datasets, adversely affecting its recognition accuracy. It may require careful parameter tuning and the incorporation of regularization techniques to extract features from ECG signals better and improve recognition accuracy.



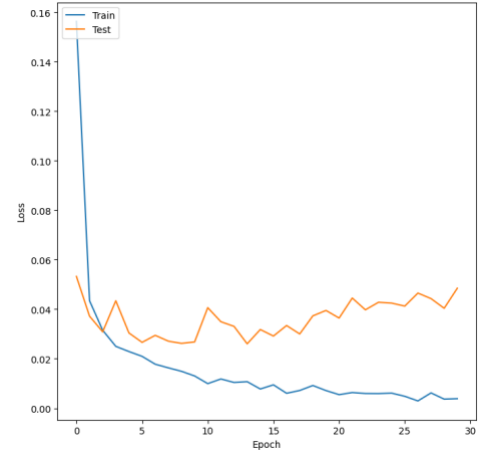
(a) Model1 accuracy



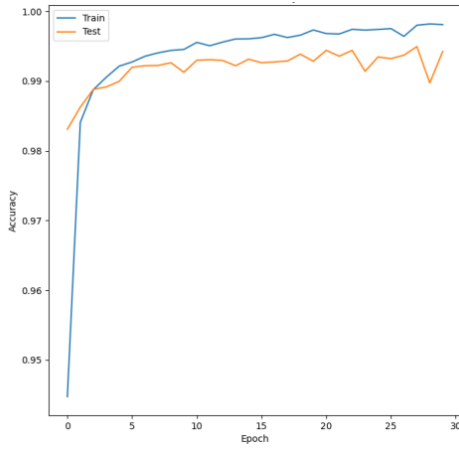
(b) Model1 loss



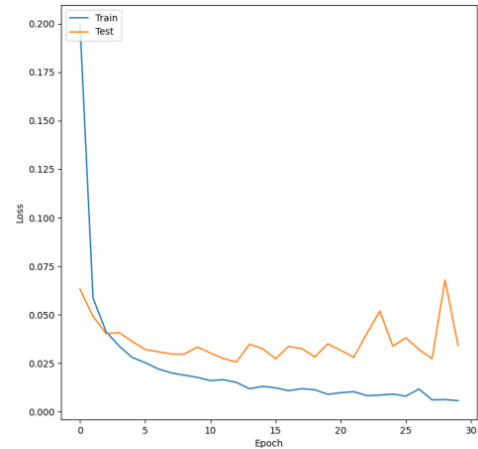
(c) Model2 accuracy



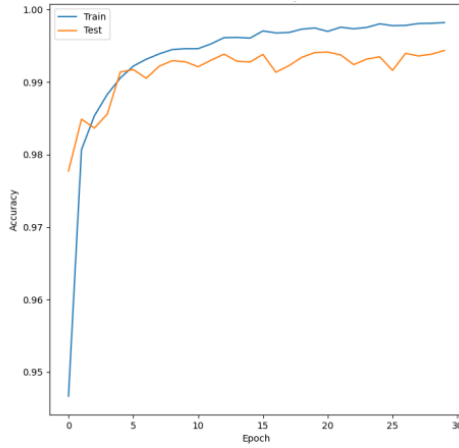
(d) Model2 loss



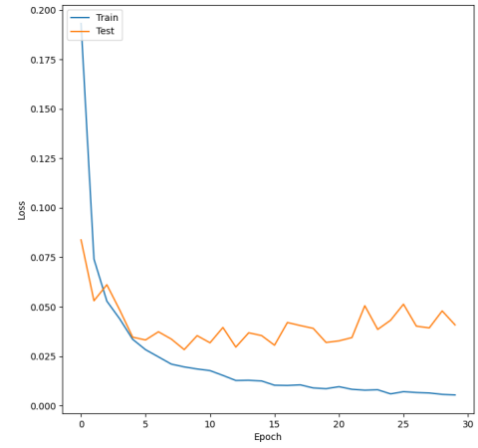
(e) Model3 accuracy



(f) Model3 loss



(g) Model4 accuracy



(h) Model4 loss

Figure 2. The change process of accuracy and loss for model1-4 with the training set and the testing set.

5. Conclusion

In the task of classifying ECG signals, the four CNN models employed in this paper achieved high recognition accuracy (>99%) for 5 types of ECG Signals: Normal beat, Left bundle Branch Block(LBBB), Right Bundle Branch Block(RBBB), and Ventricular Premature Contraction (VPC), as well as Atrial Premature Contraction (APC). Furthermore, the recognition time for all models remained below 250ms, meeting the requirements for real-time ECG signal recognition. Among these models, model 2 achieved the highest recognition accuracy of 99.52%. However, all models exhibited relatively lower recognition accuracy for APC ECG signals, representing APC, with rates ranging from 85% to 90%. Future work will involve adjusting the proportions of different classes within the dataset to enhance the models' ability to accurately recognize APC's ECG signal.

Furthermore, by comparing the training results of different models, it is observed that shallower CNN models (e.g., model 1) can extract basic ECG signal features while maintaining faster recognition speeds. As the model depth increases (e.g., model 2 and model 3), CNN models can learn more complex and abstract features, resulting in higher accuracy for handling complex tasks, albeit with slower recognition speeds. The residual network structure (e.g., model 4) demonstrated a relatively faster recognition speed compared to traditional CNNs for ECG signal classification. However, it exhibits overfitting issues, leading to lower recognition accuracy than model 1, 2, and 3. This model may require careful parameter tuning and the inclusion of regularizers to extract features from ECG signals better and improve recognition accuracy. In conclusion, the results of this paper indicate that CNN models have advantages and limitations in the task of ECG signal classification. Selecting an appropriate model depth and structure is crucial for achieving high accuracy in ECG signal classification.

References

- [1] World Health Organization. Cardiovascular diseases (CVDs). June 11, 2021. Retrieved on May 13, 2023. Retrieved from: <http://www.who.int/mediacentre/factsheets/fs317/en/>
- [2] National Heart Lung and Blood Institute. Types of arrhythmias. March 24, 2022. Retrieved on May 13, 2023. Retrieved from: <https://www.nhlbi.nih.gov/health/arrhythmias>
- [3] Acharya U, Suri J, Spaan J, and Krishnan S 2007 *Advances in cardiac signal processing*
- [4] Martis R, Acharya U, and Adeli H 2014 Current methods in electrocardiogram characterization *Comput. Biol. Med.* 48(1) p 133-149
- [5] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, Mietus J, Moody G, Peng C K and Stanley H 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals *Circulation*. 101(23) p e215-e220
- [6] Singh B and Tiwari A 2006 Optimal selection of wavelet basis function applied to ECG signal denoising *Digit. Signal Process. A Rev. J.* 16(3) p 275-287
- [7] Pan J and Tompkins W 1985 A real-time QRS detection algorithm *IEEE Trans. Biomed. Eng.* BME-32(3) p 230-236 doi: 10.1109/TBME.1985.325532.
- [8] Schmidhuber J 2015 Deep learning in neural networks: an overview *Neural Networks* 61 p 85-117
- [9] Yann L, Yoshua B and Hinton G 2015 Deep learning *Nature* 521 p 436-444
- [10] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J and Chen T 2018. Recent advances in convolutional neural networks *Pattern recognition*. 77 p 354-377
- [11] He K, Zhang X, Ren S and Sun J 2015 Delving deep into rectifiers: surpassing human-level performance on imageNet classification *IEEE Int. Conf. Comp. Vision* p 1026-1034 doi: 10.1109/ICCV.2015.123.
- [12] Wang P, Hou B, Shao S and Yan R 2019 ECG Arrhythmias Detection Using Auxiliary Classifier Generative Adversarial Network and Residual Network *IEEE Access* 7 p 100910-100922 doi:10.1109/ACCESS.2019.2930882.
- [13] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *IEEE Conf. Comp. Vision Pattern Recognit.* p 770-778 doi:10.1109/CVPR.2016.90.