A 2D CNN APPROACH FOR ECG ARRHYTHMIA CLASSIFICATION

Cao Kieu Van Manh, Tran Dinh Minh Khoa, Luong Thien and Nguyen Quoc Cuong

The University of Da Nang, University of Science and Technology, Da Nang, Viet Nam

Abstract. Currently, cardiovascular diseases are the leading cause of death in the world, especially in underdeveloped countries. Many people do not have the opportunity to receive regular check-ups and care for their cardiovascular health because of the high cost and insufficient time. Nowadays, smartphones have become commonplace, and most families own at least one device. Therefore, in this study, we propose a model to warn of abnormal signals in the human heart rhythm as well as proceed to build a practical, convenient, low-cost system for monitoring the heart rate and alerting the user to abnormal signals. This system can measure and give diagnostic results in real time. The model selected in this study is CNN, trained on the MIT-BIH Arrhythmia Database medical dataset with high accuracy for the abnormal heart rhythm classification task.

Keywords: Electrocardiogram; Convolutional Neural Network; Image classification.

1 Introduction

According to a report by the World Health Organization (WHO), cardiovascular diseases are the number one cause of death today. It is estimated that 17.9 million people died from cardiovascular diseases in 2019, accounting for 32% of total global deaths, of which more than 75% came from low- and middle-income countries.

Currently, many products identify and diagnose cardiovascular diseases. However, those systems are often extensive, consume a lot of data, and are associated with high costs. Our solution was to provide a simple and effective method for diagnosing cardiovascular disease without using complex equipment. Instead, users can use an affordable ECG sensor to collect ECG data, a lightweight deep-learning model to classify abnormal signals, and a mobile application to track real-time results.

In this study, we will experiment with classifying abnormal ECG signals using two-dimensional images, evaluating the effectiveness of the CNN model presented in [1] and a self-developed model on collected and real-life data.

2 Related research

Methods used to classify heart rate have been presented in many different articles and studies. Research [7] has introduced a method using feed-forward neural networks (FFNN). Support Vector Machine (SVM) is also a widely applied classification method for detecting arrhythmias based on electrocardiograms.

Research [8] proposed using SVM with Particle Swarm Optimization (PSO) to classify six different heart rate types. Other machine learning techniques besides FFNN and SVM are also used, as the study [9] proposed recurrent neural networks (RNN) as a classification model for four different types of heart rhythms with feature extraction methods based on eigenvectors. Research [10] introduced the random forest method (RFT) as a classification tool, and research [11] proposed a KNN classifier to classify 17 types of heart rhythms.

In particular, in the reference study [1], a method for classifying ECG arrhythmias was proposed using a CNN model with ECG images on grayscale (black and white images). By converting a 1D ECG signal into a 2D ECG image, noise filtering, and feature extraction are no longer necessary. This is important because some ECG rhythms are missed during noise filtering and feature extraction. Additionally, training data can be enhanced by enhancing ECG images for higher classification accuracy. Data augmentation has been difficult to apply in previous studies because distortion of the 1D ECG signal can reduce the performance of the classifier. Moreover, when the electrocardiogram signal is converted into a 2D image, the CNN model can automatically ignore noisy data while extracting the relevant features in the convolution and pooling layers. Therefore, the CNN model can be applied to ECG signals from different ECG devices with different sampling rates and amplitudes, while previous studies required a different model for different devices.

The results of the method in [1] achieved quite high accuracy when compared with previous 1D ECG signal processing methods or other methods mentioned above. Research [1] has also brought a new approach to classifying abnormal ECG signals using 2D images.

In this study, we will experiment with the approach of classifying abnormal ECG signals using 2D images as well as test the CNN model structure in the study [1]. At the same time, we also built a new CNN model structure as well as configured a device to apply the research model in real life.

3 Proposed methods

3.1 DFRobot Heart Rate Monitor Sensor

The DFRobot Heart Rate Monitor Sensor [3] is used to measure the electrical activity of the heart. This electrical activity can be charted as an ECG and output as an analog reading. It this research, we use this sensor to apply our model in real life.

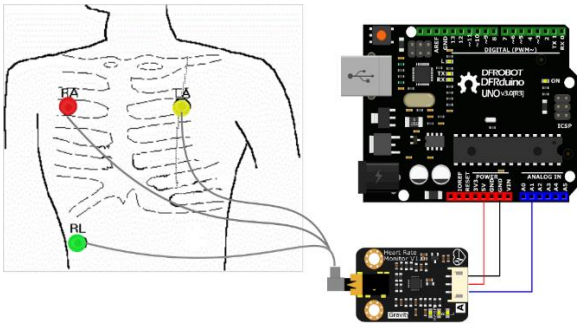

Fig. 1. DFRobot Heart Rate Monitor Sensor connection diagram

3.2 ECG signals

An electrocardiogram (ECG) is a test that records the natural electrical impulses that coordinate the heart's contractions to keep blood circulating. The QRS complex is an important component in the ECG, representing the correspondence between potentials on the body's surface.

Any changes in electrical impulses detected through an electrocardiogram and QRS complex can also indicate cardiovascular problems such as arrhythmia, heart failure, valvular stenosis, heart attack, and other conditions.

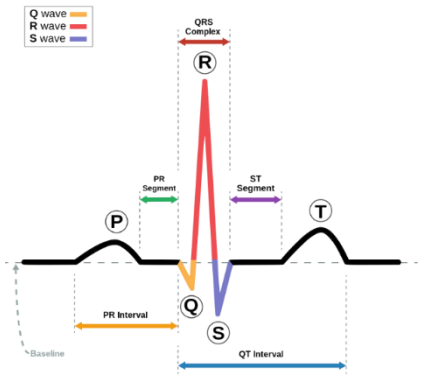

Fig. 2. QRS complex in ECG signal

3.3 Data acquisition

The ECG signal data source is taken from the MIT-BIH Arrhythmia Database [2], measured from 45 patients, including 19 women (23–89 years old) and 26 men (32–89 years old) at 360 Hz frequency. The data includes 48 records; each record is about 30 minutes long. The dataset has a total of 19 classes. The group proceeds to select six classes to use, as shown in Table 1.

Table 1. ECG arrhythmia description

ECG arrhythmia	Description	Dataset records
Normal (NOR)	Normal sinus rhythm.	100, 101, 103, 105, 108, 112, 113, 114, 115, 117, 121, 122, 123, 202, 205, 219, 230, 234
Premature Ventricular Contraction (PVC)	Abnormal ventricular rhythm originates from the ventricles.	106, 116, 119, 200, 201, 203, 208, 210, 213, 215, 221, 228, 233
Paced Beat (PAB)	Heart rate is generated by an electrical pacemaker plugged into the body.	102, 104, 107, 217
Right Bundle Branch Block (RBB)	Conduction disorders due to the right bundle branch block are prevented.	118, 124, 212, 231
Left Bundle Branch Block (LBB)	Conduction disorders caused by left bundle branch block.	109, 111, 207, 213
Atrial Premature Beat (APB)	Abnormal sinus rhythm originating from the atria.	209, 220, 222, 223, 232

In the dataset, retrieve records '100', '101', '103', '106', '107', '118', '109', and '209' to use as model testing data.

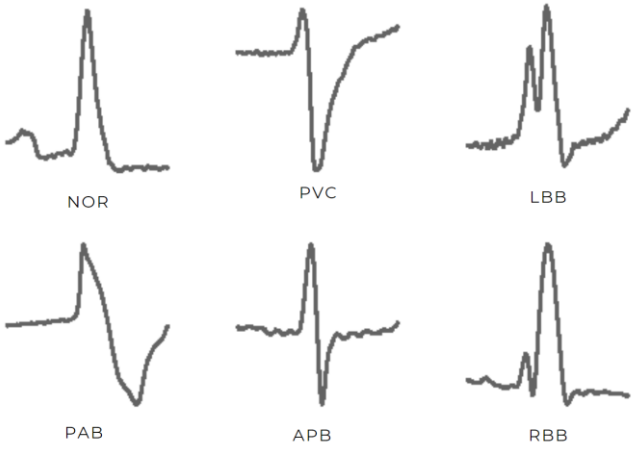

Fig. 3. ECG arrhythmia as an 2D image

3.4 Algorithm to find peaks in ECG signal

We find the R peak of the ECG signal using the neurokit2 library based on the article [5]. The algorithm to find the R peak of an ECG signal in the support library can be described as follows:

- Create a window with a length of 800 ms, and slide that window over each sample of the signal in turn. With each slide we obtain a signal segment with a length of 800 ms.
- Find the largest value on that signal segment. If the position of that largest sample is in the middle of the signal segment, it is an R peak.
- After processing all signal segments output from the sliding window, we obtain a list of R peaks for that signal.

3.5 Preprocess ECG signal

Process data obtained from sensors
We test the sensor on frequencies of 20 Hz, 33 Hz, and 50 Hz. With a frequency of 20 Hz, the sensor data loses many important samples due to the low frequency. With 33 Hz, the sensor produces an ECG signal with a larger number of heartbeats and quite good signal quality. With 50 Hz, the sensor produces a lot of noise and a smaller number of heartbeats because the sampling rate is too fast. So we chose 33 Hz as the frequency level on the sensor.

Downsample signals in the original dataset
Actual measurement data from our sensors only produced the best results at 33 Hz, while the signal in the MIT-BIH data set had a frequency of 360 Hz. So we decided to downsample the collected data frequency from 360 Hz to 33 Hz. Here, we calculate the coefficient M = 10, which is the integer part of the division between the original frequency and the resulting frequency. For each M sample of the original signal, we select one sample to include in the resulting signal. After performing comparisons, we found that downsampling the signal did not cause much change in the ECG image signal.

About the number of samples in an image
We cropped images with the number of samples in an image being 12, 16, 18, 24, and 30 samples, respectively. With measured data from the sensor, increasing the number of samples helps to have more information; however, when increasing to 30 samples, some cropped images will contain information from other peaks, leading to possible distortion of the results.

Here, we convert each heartbeat in the signal into a grayscale image with dimensions of 128 x 128, with each heartbeat having a length of 24 samples in the heartbeat signal, equivalent to 0.72 seconds, with the R peak located in the center.

Data augmentation

For each original image, we crop the corners of the image according to nine cropping methods: left-top, middle-top, right-top, left-middle, middle-right, left-bottom, middle-bottom, right-bottom. Each crop method produces an image that is 112 x 112. The images are then resized to 128 x 128.

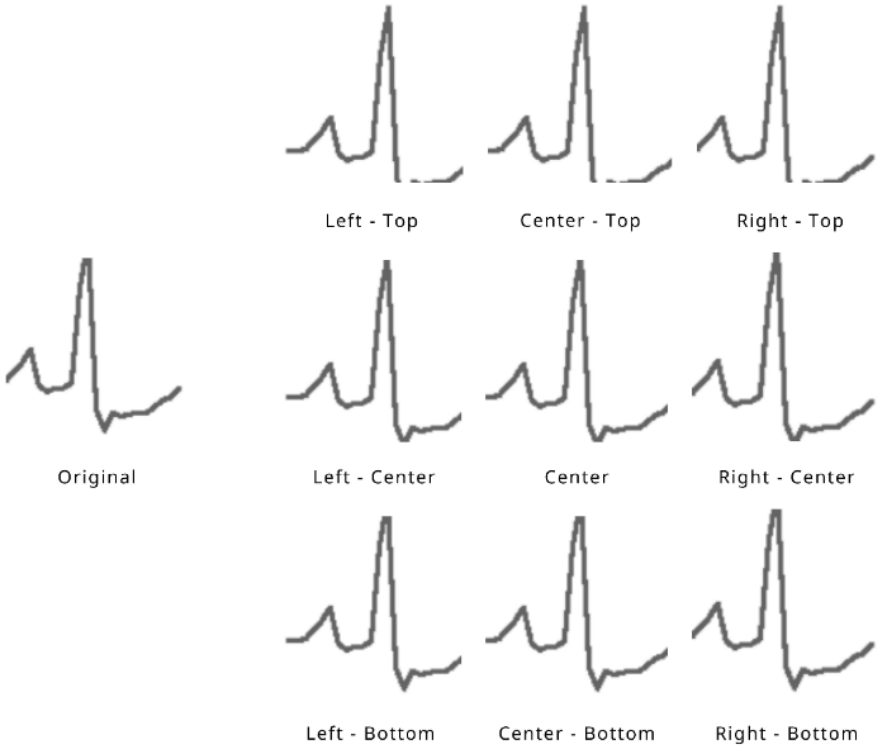

Fig. 4. Data augmentation for 2D ECG

Training and testing data

After processing the data from the original dataset, we synthesized it into 3 datasets and 2 test sets, as shown in Figure 5.

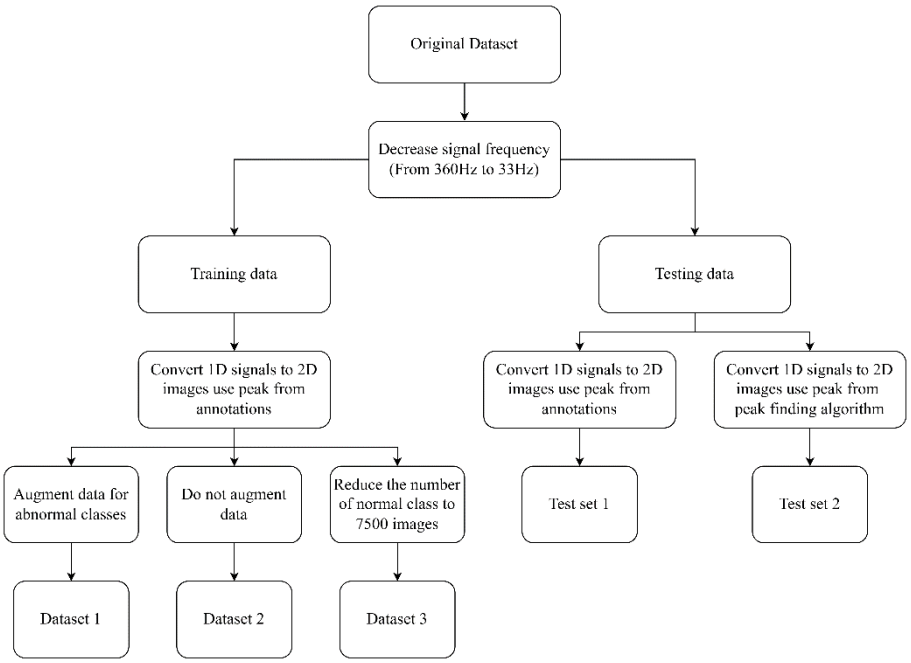

Fig. 5. Data preprocessing process

3.6 ECG arrhythmia classifier

Train the model to recognize abnormal ECG signals with the CNN model
The reference CNN structure from the article [1] includes layers as shown in Table 2, with the number of parameters being 135,392,200.

Table 2. CNN model structure that we referenced.

	Type	Kernel size	Stride	Kernel	Input size
Layer 1	Conv2D	3 x 3	1	64	128 x 128 x 1
Layer 2	Conv2D	3 x 3	1	64	128 x 128 x 64
Layer 3	Pool	2 x 2	2	-	128 x 128 x 64
Layer 4	Conv2D	3 x 3	1	128	64 x 64 x 64
Layer 5	Conv2D	3 x 3	1	128	64 x 64 x 128
Layer 6	Pool	2 x 2	2	-	64 x 64 x 128
Layer 7	Conv2D	3 x 3	1	256	32 x 32 x 256
Layer 8	Conv2D	3 x 3	1	256	32 x 32 x 256
Layer 9	Pool	2 x 2	2	-	32 x 32 x 256
Layer 10	Full	-	-	2048	16 x 16 x 256
Layer 11	Out	-	-	8	2048

In addition, we build a simpler and less complex CNN model with the layers described in Table 3, with the number of parameters reduced to 2,190,278.

Table 3. The architecture of our CNN model.

	Type	Kernel size	Kernel	Activation	Input size
Layer 1	Conv2D	3 x 3	32	ReLU	126 x 126 x 32
Layer 2	MaxPooling2D	2 x 2	-	-	63 x 63 x 32
Layer 3	Conv2D	3 x 3	64	ReLU	61 x 61 x 64
Layer 4	MaxPooling2D	2 x 2	-	-	30 x 30 x 64
Layer 5	Conv2D	3 x 3	128	ReLU	28 x 28 x 128
Layer 6	MaxPooling2D	2 x 2	-	-	14 x 14 x 128
Layer 7	Flatten	-	-	-	25,088
Layer 8	Dense	-	64	ReLU	64
Layer 9	Dense (Output)	-	6	Softmax	6

In this study, we developed six models. Models 1.1, 1.2, and 1.3 use the architecture from the article [1] and are trained on dataset 1, dataset 2, and dataset 3, respectively. Models 2.1, 2.2, and 2.3 use our architecture and are trained on dataset 1, dataset 2, and dataset 3, respectively.

4 Experiments and results

4.1 Data preprocessing results

The following table describes 3 dataset and 2 test set that we use in this study.

Table 4. Dataset 1

Class	NOR	PVC	PAB	RBB	LBB	APB
Train set	51537	52380	28930	40710	44630	16200
Validation set	12884	13090	7230	10180	11160	4050

Table 5. Dataset 2

Class	NOR	PVC	PAB	RBB	LBB	APB
Train set	51537	5238	2893	4071	4463	1620
Validation set	12884	1309	723	1018	1116	405

Table 6. Dataset 3

Class	NOR	PVC	PAB	RBB	LBB	APB
Train set	6000	5238	2893	4071	4463	1620
Validation set	1500	1309	723	1018	1116	405

Table 7. Test dataset

Class	NOR	PVC	PAB	RBB	LBB	APB
Test 1	10300	634	2076	2165	2490	516
Test 2	2194	52	405	499	551	61

4.2 Training results

Training results on two models 1.2 and 2.2

The following are graphs of the accuracy and loss values during the training process of the two models that gave the best results: 1.2 and 2.2. In these graphs, model 2.2 uses the early stopping technique and has stopped training at the 8th epoch.

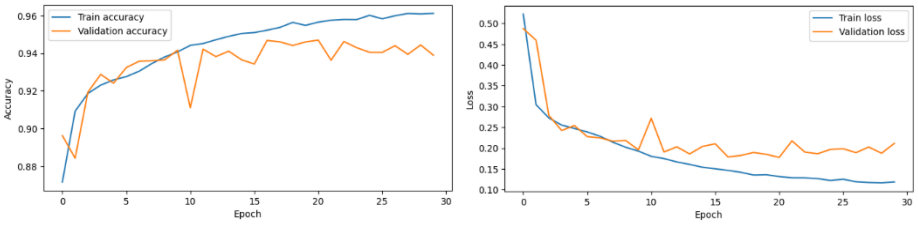

Fig. 6. Training results on model 1.2

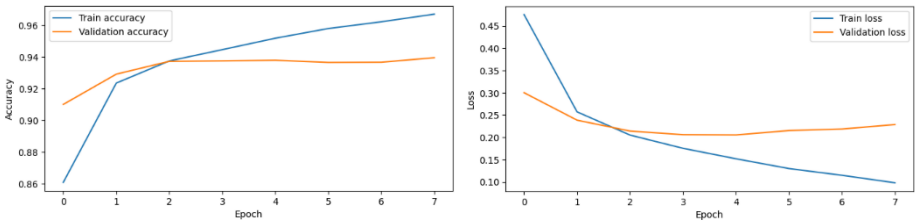

Fig. 7. Training results on model 2.2

4.3 Evaluation results

Evaluation results

Perform testing on three models with their respective test datasets. Here, we only present the results of the two metrics that are of most interest, which are the accuracy and precision of the 'Normal' class. The results are shown in Table 8.

Table 8. Test dataset

Model	Test 1		Test 2	
	Accuracy	Precision	Accuracy	Precision
1.1	0.96	0.95	0.63	0.89
1.2	**0.96**	**0.95**	**0.71**	**0.86**
1.3	0.91	0.99	0.51	0.88
2.1	0.96	0.96	0.67	0.81
2.2	**0.96**	**0.96**	**0.71**	**0.85**
2.3	0.89	0.98	0.55	0.94

Review

- On the first test set, it is processed based on the peak that was marked in the original data set. The model testing metrics all showed good results (over 90%). On the second test set, which is processed based on the peak-finding algorithm, Test results dropped quite a bit.
- In the 6 models trained above, models 1.2 and 2.2 (trained on dataset 2) give the highest performance on both test sets 1 and 2.

Evaluation

- The reason for the reduced performance may be because some diseases also depend on other peaks such as Q and S, not just the R peak. Therefore, searching only by the R peak can lead to inaccuracies when checking.
- We can also see the importance of having a good enough peak-finding algorithm to ensure the collected data contains accurate and complete information about a heartbeat.
- Models trained with augmented datasets have test results that have not improved compared to non-augmented datasets, possibly due to the following two reasons:
  - Low-quality augmented data: because ECG is a continuous signal, image cropping can produce inaccurate or inappropriate data. This can distort the model and affect test performance.
  - Augmented data is not diverse; data augmentation, as above, is mainly a repetition of the original data and does not help the model learn new samples, and performance on test data is not improved.
- From there, we can see that data enhancement in problems with ECG data is quite difficult, even with image data.

4.4 Experiments with real data

We used the two best models, 1.2 and 2.2, to test with actual data measured from the sensor. Below is a table of summary results when measuring on our team (normal people). Each person was measured five times consecutively. This results table contains the Normal class ratio for each measurement.

Table 9. Actual usage results (%)

User	Model	1st	2nd	3rd	4th	5th
Cao Kieu Van Manh	l .2	86.67	75.41	78.57	85.94	83.93
	2.2	90.23	86.56	87.33	87.69	88.72
Tran Dinh Minh Khoa	l .2	76.59	78.83	75.51	88.04	87.65
	2.2	88.23	95.83	89.79	100.0	94.00
Luong Thien	l .2	83.45	85.11	82.33	83.70	84.70
	2.2	90.00	93.54	90.16	88.52	93.65
Nguyen Quoc Cuong	l .2	73.33	78.26	100.0	96.15	100.0
	2.2	84.44	85.66	88.37	87.66	89.96

Results of testing the model's execution speed

We test the execution time of predicting results on 20 real data files using two models 1.2 and 2.2. We have a result table as follows:

Table 10. The model's execution speed.

Model	Size (MB)	Time to predict 20 files	Average time per file (s)
1.2	516	320	16
2.2	25	144	7,2

Review

- The above data was measured on four normal people. Therefore, the results of the two models above are quite close to reality. However, we have not yet concluded whether the above results are completely accurate or not because they have not been verified by a specific medical test.
- The self-developed group model is only 1/20 the size of the model in the article [1]. With the same hardware configuration and network connection, the model developed by us has a prediction time nearly twice as fast as the model in the article [1].

5 Conclusion

Through this study, we have experimented with the results in the study [1] and compared them with our proposed models in terms of their ability to identify and diagnose some abnormal signals based on electrocardiograms. At the same time, we also tested the hardware and applied the above models in practice.

Our research results show that using two-dimensional ECG signals and classifying using the CNN architecture gives high results. This result shows that scalability and practical application are completely feasible and can be further developed. However, there

are still some difficulties in implementing abnormal signal classification models into practice that we have concluded during the research process, specifically as follows:

- About data: data on good-quality ECG signals are quite scarce, and data enhancement on two-dimensional ECG has not shown an improvement in this study.
- About hardware devices: an ECG sensor is needed to ensure accuracy. However, these devices currently have high costs, so building a cheap device that still ensures accuracy is still difficult. Another development direction is to use this research to build a system to help doctors diagnose cardiovascular diseases in hospitals with more modern and accurate equipment.
- About medical: support from hospitals and specialized doctors is needed to provide necessary medical evidence and authentication to ensure the accuracy of the model when applied in practice.

References

1. Tae Joon Jun, Hoang Minh Nguyen, Daeyoun Kang, Dohyeun Kim, Daeyoung Kim, Young-Hak Kim. ECG arrhythmia classification using a 2-D convolutional neural network. Computer Vision and Pattern Recognition (2018), https://arxiv.org/abs/1804.06812
2. George Moody, Roger Mark, MIT-BIH Arrhythmia Database (2023), https://www.physionet.org/content/mitdb/1.0.0/
3. DFRobot, Product Wiki - DFRobot Heart Rate MonitorSensor (2023), https://wiki.dfrobot.com/Heart_Rate_Monitor_Sensor_SKU__SEN0213
4. Khiem H. Le, Hieu H. Pham, Thao BT. Nguyen, Tu A. Nguyen, Tien N. Thanh, Cuong D. Do, LightX3ECG: A Lightweight and eXplainable Deep Learning System for 3-lead Electrocardiogram Classification (2022), https://arxiv.org/abs/2207.12381v1https://arxiv.org/abs/2207.12381v1
5. Mohsen Nabian, Yu Yi, Jolie Wormwood, Karen S Quigley, Lisa F Barrett, Sarah Ostadabbas, An Open-Source Feature Extraction Tool for the Analysis of Peripheral Physiological Data (2018), https://pubmed.ncbi.nlm.nih.gov/30443441/
6. James Beckerman, MD, FACC, What to Know About Your Heart Rate (2022), https://www.webmd.com/heart/ss/slideshow-heart-rate
7. Gler, beyl ED. ECG beat classifier designed by combined neural network model. Pattern recognition 38(2):199-208 (2005), https://ui.adsabs.harvard.edu/abs/2005PatRe..38..199G/abstract
8. Melgani F, Bazi Y. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. IEEE transactions on information technology in biomedicine 12(5):667-677 (2008), https://ieeexplore.ieee.org/document/4487050
9. Beyli ED. Combining recurrent neural networks with eigenvector methods for classification of ECG beats. Digital Signal Processing 19(2):320-329 (2009), https://www.sciencedirect.com/science/article/abs/pii/S1051200408001516
10. Kumar RG, Kumaraswamy YS. Investigating cardiac arrhythmia in ECG using random forest classification. International Journal of Computer Applications 37(4):31-34 (2012), https://www.mdpi.com/1999-4893/13/4/75
11. Park J, Lee K, Kang K. Arrhythmia detection from heartbeat using k-nearest neighbor classifier. IEEE International Conference on Bioinformatics and Biomedicine pp 15-22 (2013), https://ieeexplore.ieee.org/document/6732594