

# ECG arrhythmia classification using recurrence plot and convolutional neural network

Bhekumuzi M. Mathunjwa <sup>1</sup>, Yin-Tsong Lin <sup>2</sup>, Chien-Hung Lin <sup>2</sup>, Maysam F. Abbod <sup>3</sup>, Jiann-Shing Shieh <sup>1,\*</sup>

<sup>1</sup> Department of Mechanical Engineering and Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan, Chung-Li 32003 Taiwan; mathunjwabhekie@gmail.com

<sup>2</sup> AI R&D Department, New Era AI Robotic Inc., Taipei City 105, Taiwan; lotusytlin@neweraai.com, lance\_lin@kinpo.com.tw

<sup>3</sup> Department of Electronics and Computer Engineering, Brunel University London, UB8 3PH UK; maysam.abbod@brunel.ac.uk

\* Correspondence: jsshieh@saturn.yzu.edu.tw

Received: date; Accepted: date; Published: date

**Abstract:** Diseases that affect the cardiovascular system affect about 50 million people in the world, thus heart disease prevention is one of the most important tasks of any health care system. Despite being popular, better automatic ECG signal analysis methods are still needed. This research is dedicated to design a new deep learning based method that effectively classify the arrhythmia in two stages using two second segments of 2D recurrence plot images from the ECG signal. The first stage discriminates between Noise and Ventricular fibrillation. The second stage discriminate between Atrial fibrillation, Normal, Premature atrial fibrillation and Premature ventricular fibrillation. Models are trained and tested using ECG databases that are publicly available in PhysioNet. The databases include the MIT-BIH Arrhythmia Database, Creighton University Ventricular Tachyarrhythmia Database, MIT-BIH Atrial Fibrillation Database, and MIT-BIH Malignant Ventricular Ectopy Database, comparing six types of arrhythmia, to achieve testing accuracy of up to  $95.3 \pm 1.27\%$  and  $98.41 \pm 0.11\%$  arrhythmia detection for the first and second stage respectively, from fivefold cross-validation. In conclusion, this contribution substantially advances the current methodology for discriminating between different types of arrhythmia and provide clinicians with an effective tool to assist in arrhythmia detection.

**Keywords:** Arrhythmia, Recurrence Plot, Convolutional neural network, Electrocardiography

---

## 1. Introduction

Globally, disease that affects the cardiovascular system is the leading cause of death, accounting for about one-third of deaths in 2008 [1, 2]. A majority of the deaths are due to Ventricular Fibrillation, a life-threatening arrhythmia, and sudden cardiac deaths, which is common among adults. Arrhythmia is a class of conditions that falls under cardiovascular diseases (CVDs), characterized by irregular changes from the normal rhythm of the heart [3]. There are a number of arrhythmias, which include premature heartbeats, supraventricular (atrial fibrillation) tachycardia, bradyarrhythmias, and ventricular arrhythmias (ventricular fibrillation). Effective treatment of a number of arrhythmias relies on the early stage of detection [3]. However, atrial fibrillation is an irregular heart rate that has a very serious health implication, like an increased risk of stroke [10, 11]. The normal heart rate of a healthy person is between 60 to 100 beats per minute; however, AF causes a hike in heart rate of over 100 beats per minute [12, 13]. Most often, AF

results from an underlying condition like hypertension. It can affect people of any age but more common in the elderly population and rare in the younger generation [14]. Premature atrial contractions (PAC) is another common form of heart arrhythmia that is characterized by premature heartbeat from the atria. This kind of arrhythmia is common with healthy young and elderly people, and it is unclear as to what is the exact cause [15]. Premature ventricular contractions (PVCs) are extra abnormal beats that come from the ventricles. They usually occur when the ventricular contractions occur sooner than the next regular beat and followed by an even stronger heartbeat, causing disorder in pumping blood. PVCs that occur frequently can be a risk factor for arrhythmia-induced cardiomyopathy, a condition whereby the heart muscle becomes less effective and may develop heart failure [16]. Ventricular fibrillation is a type of arrhythmia that is often caused by a rapid heartbeat known as ventricular tachycardia (VT), which results from an abnormal electrical impulse in the ventricles. During this abnormal heartbeat, there is often a chaotic contraction in the ventricles. This can be seen in the study of the ECG, which shows the variation of the ECG signal without the QRS complex [17]. Therefore, the instinctive detection of ECG arrhythmia becomes an essential tool for risk assessment.

Over the years, there has been an increased interest in time series methods and is justified by the number of applications used to generate data with time. There is also a large number of signal processing methods that advance change in the characterization of features in the power spectrum [4-6]. Despite the fact that frequency might be the most preferred representation domain, other possible alternatives in this subject are autocorrelation, wavelets, shapelet based time series, and principle component analysis [7-9]. However, recurrence plots are an advanced technique that distinguishes how resemblances between particular orders vary over time. Recurrence plots are a generally used procedure for qualitative evaluation of time series in dynamic systems.

The ECG signal is essential in uncovering and examining the nature of cardiac health [18]. Therefore, the morphology of ECG signals incorporates crucial information about the state of the heart. However, it is very difficult to spot any changes in the ECG signals as its amplitude is measured in millivolts thus highly nonlinear [18, 19].

Nowadays, wearable devices dominate arrhythmia classification tools and they are recently giving outstanding results with the help of deep learning. Convolutional neural networks (CNN) are one of the widely used ECG classification methods. They are applied in a wide range of applications and recently used in 1D ECG arrhythmia classification where they produced promising results in the future of arrhythmia classification [20, 21]. Even though this technique gives promising results in a specific type of data, other methods are needed to enhance performance improvement. This may be because the data used is affected by the limitations of using one-dimensional data. Tae [22] proposed an ECG arrhythmia classification method using deep two-dimensional CNN with grayscale ECG images. Transforming 1D ECG signals to 2D images has a number of advantages. Classification accuracy can be improved by using features like data augmentation to enlarge the training data. Feature extraction and noise filtering are no longer necessary which reduces the chances of ignoring some ECG beats. CNN's are commonly used in deep learning tasks that are related to object recognition using a large number of images where it produces the highest level of performances [23-25].

In this paper, CNN is used to classify 2D ECG arrhythmia in two stages. The work benefited from the advances of using 2-dimensional data to eliminate the limitations brought up by using 1-dimensional data that requires feature extraction, [26-28], and filtering for noise [29-31]. Two-dimensional image ECG classification greatly improves the classification accuracy in the sense that it utilizes every beat of the data that may be eliminated during feature extraction and noise filtering. Another advantage of using 2-D images in classifying ECG is that accuracy can be enhanced by features like data augmentation [32-34] that can be applied in case of data deficiency.

The approach taken in the paper is evaluated based on its potential to discriminate between six types of arrhythmia (AF, Normal, Noise, PAC, PVC and VF) using ECG data acquired from the MIT-BIH

Arrhythmia Database, Creighton University Ventricular Tachyarrhythmia Database, MIT-BIH Atrial Fibrillation Database, and MIT-BIH Malignant Ventricular Ectopy Database in PhysioNet. RP is applied to convert the ECG data to 2D images, which allows the application of the above-mentioned features, thus improve the classification capabilities of the method. Two-stage classification is proposed in this research to improve the classification capability of the method in all the labels. Processing starts at the first stage, which discriminate between data with labels that needs immediate attention (Noise and VF). VF can be life threatening, and noise can hinder the classification process, thus needs to be detected first. Furthermore, these labels cannot be discriminated in the second stage due to the absence of the R-peak from their beats [44, 46]. The second stage classify the rest of the labels into respective types of arrhythmia by applying the R-peak detection algorithm as the labels are annotated beat by beat [41]. The results show that CNN using images generated using RP can be suitably applied for arrhythmia discrimination.

The remaining part of the paper is arranged as follows: Sections 2 and 3 describe the analysis of the RP to produce the 2D segment representation of the ECG signal and the CNN architecture. In sections 4 and 5, the methodology is analyzed, which includes the steps for the proposed method, training environment, evaluation procedures and results of the proposed approach. Finally, sections 6 and 7 provide the discussion and conclusions of the work respectively.

## 2. Recurrence Plot

Past researchers studied a way to quantify the recurrent behavior of a trajectory through a phase space [35]. Since the phase space does not have low dimensions to be pictured on, it is usually hard to visualize. Eckmann et al. [36] introduced a representation called recurrence plot (RP), which is able to reveal when these trajectories visit the same area. The RP representation allows us to investigate the m-dimensional trajectories in a bi-dimensional representation. A two-dimensional square matrix R, with both axes as time axes provide insight to picture the recurrence of a state at time  $i$  and  $j$ .

The numerical expression of the RP is shown in Equation (1).

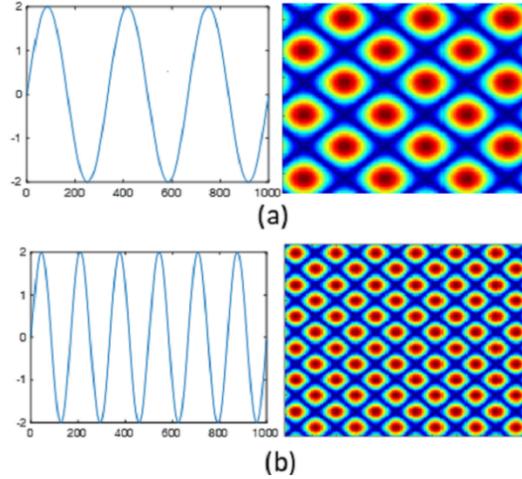
$$R_{i,j}^{m,\varepsilon_i} = \Theta\left(\varepsilon_i - \left\|\bar{x}_i - \bar{x}_j\right\|\right), \quad \bar{x}_i \in \mathbb{R}^m, \quad i, j = 1 \dots N. \quad (1)$$

where  $\Theta$  is the Heaviside function defined in equation 2,  $\varepsilon$  is a threshold for closeness,  $\bar{x}_i$  and  $\bar{x}_j$  are the observed subsequences at the positions  $i$  and  $j$ .  $\|\cdot\|$  is the norm (e.g. Euclidean norm) and  $N$  is the number of states.

$$\Theta(z) = \begin{cases} 0 & \text{if } z > 0 \\ 1 & \text{otherwise,} \end{cases} \quad (2)$$

Equation 2 describes that the recurrence plot portrays the pairs of times the trajectory is at the same place. Since the RP is identified by  $R_{i,j}=1$  ( $i = 1 \dots N$ ), it has a black main diagonal line, which is the line of identity, that has an angle of  $\pi/4$ . Phase space trajectories are reconstructed from the totality of all recurrence points [56-58]. However, they cannot be reconstructed from a single recurrence point( $i, j$ ). It puts a black dot at coordinates ( $i, j$ ) if the m-dimensional trajectory of the time series  $j$  is close to  $i$  and otherwise a white dot. Figure 1 shows typical RP examples for different types of signals.

From the Heaviside function  $\varepsilon_i$  point of view, a RP is a state  $x_j$  that is sufficiently close to  $x_i$ , which means all the states  $x_j$  that are within an m-dimensional neighborhood of the size  $\varepsilon_i$  centered at  $x_i$  are recurrence points. Originally, the neighborhood of the RP uses L2-norm and its radius contains a fixed number of closest states  $x_j$  [59]. In this case, the radius  $\varepsilon_i$  changes for every  $x_i$ , ( $i = 1 \dots N$ ) and  $R_{i,j} = R_{j,i}$  because the neighborhood of  $x_i$  does not have to be the same as that of  $x_j$  which leads to asymmetric RP. However, using a matrix and a fixed radius ensures that  $R_{i,j} = R_{j,i}$  which is the case in a symmetric RP. Choosing a neighborhood type to use depends on the application [56, 60].



**Figure 1.** Typical examples of recurrence plots for a sinusoid with two frequencies, figure (a) shows the plot of the sinusoid data and the corresponding RP with a frequency of 3 Hz, and figure (b) shows the plot of the sinusoid data and the corresponding RP with a frequency of 6 Hz.

A pivotal parameter in the RP analysis is the recurrence threshold  $\epsilon$  [56, 61-62]. Therefore, recurrence threshold selection remains an important task. The selected threshold needs to be as small as possible and at the same time an adequate number of recurrences and recurrence structures. The optimal choice of  $\epsilon$  depends on the application and the experimental conditions like noise. To select a threshold that will be a few percent of the maximum phase space diameter, a rule of thumbs is used, which should be in a region of 10% of the maximum phase space diameter. However, the presence of noise may require a larger threshold, because the influence of noise would distort the structure in the RP [56].

### 3. CNN architecture

In this section, the proposed CNN architecture used to develop the classifier is described. Since we aimed at classifying arrhythmia, there is a need to find a genetic model that will recognize the type of arrhythmia in the RP segments. However, selecting the right model for the classifier is an important task. The ImageNet large scale visual Recognition Competition (ILSVRC) recently demonstrated the fact that performance is closely related to the complexity of the CNNs [63]. Three different models re compared which took part and won the ILSVRC- 2012 [37] and runner-up 2014 [64] completion respectively to find the suitable classifier for the discrimination of arrhythmia. Besides being famous for winning the ILSVRC competition, these models have a simple enough architecture for the classification with relatively small number of layers.

Figure 2 summarizes the structure of the three architectures that are developed in this study. The number of layers for the three models ranges from eight layers deep (a), 16 layers deep (b) and 19 layers deep (c). AlexNet [37] architecture consists of five convolutional layers, some of which precedes maximum pooling layers, and followed by three fully connected layers, and lastly a three-way softmax classifier for first stage or a four-way classifier for second stage. The VGG 16 and VGG19 architectures follow a similar setup as AlexNet except that they have more layers. The unique thing about the two architectures is that they follow the same arrangement of applying convolutional layers of  $3 \times 3$  filters, same padding and maximum pooling layer of  $2 \times 2$  filters consistently throughout the architecture. In the end, they apply two fully connected layers followed by the softmax for the output.

Table 1, summaries in details some of the main features of the three architectures. It accounts for the depth of each model, kernels and Units, and other parameters including activation function, sizes of the pooling layer and strides, and dropout rate. The purpose of applying these three CNN models is to contrast the adaptability of the architectures on the arrhythmia classification using RP segments from ECG signals.

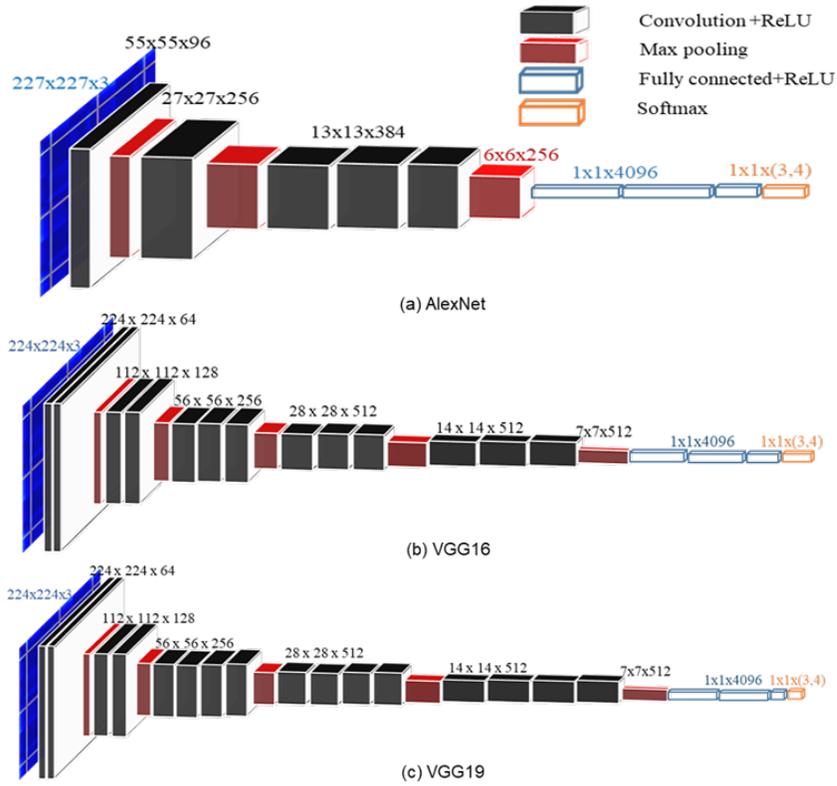


Figure 2. The architecture for the CNNs.

Table 1. Detailed parameters used for all the layers of 2D CNN models.

Layer			Kernel Size x Units		Other Layers	
AlexNet	VGG16	VGG19	AlexNet	VGG16, 19	AlexNet	VGG{16, 19}
Conv2D	Conv2D Conv2D	Conv2D Conv2D	11 × 96	3 × 64 3 × 64	Activation=ReLU, Stride = 4	ReLU, Stride = 2 ReLU, Stride = 2
MaxPool2D	MaxPool2D	MaxPool2D	-	-	Pool Size =3, stride = 2	Pool Size =2, stride = 2
Conv2D	Conv2D Conv2D	Conv2D Conv2D	5 × 256	3 × 128 3 × 128	Activation = ReLu	Activation = ReLu Activation = ReLu
MaxPool2D	MaxPool2D	MaxPool2D	-	-	Pool Size = 3, stride = 2	Pool Size = 2, stride = 2
Conv2D	Conv2D	Conv2D	3 × 384	3 × 256	Activation = ReLu	Activation = ReLu
Conv2D	Conv2D	Conv2D	3 × 384	3 × 512	Activation = ReLu	Activation = ReLu
Conv2D	Conv2D	Conv2D	3 × 256	3 × 512	Activation = ReLu	Activation = ReLu
-	-	Conv2D	-	3 × 512*	-	Activation = ReLu*
MaxPool2D	MaxPool2D	MaxPool2D	-	-	Pool Size = 3, stride =2	Pool Size = 2, stride =2
-	Conv2D Conv2D Conv2D	Conv2D Conv2D Conv2D	-	3 × 512 3 × 512 3 × 512	-	Activation = ReLu Activation = ReLu Activation = ReLu
-	-	Conv2D	-	3 × 512*	-	Activation = ReLu*
-	MaxPool2D	MaxPool2D	-	-	-	Pool Size = 2, stride = 2
-	Conv2D Conv2D Conv2D	Conv2D Conv2D Conv2D	-	3 × 512 3 × 512 3 × 512	-	Activation = ReLu Activation = ReLu Activation = ReLu
-	-	Conv2D	-	3 × 512*	-	Activation = ReLu*
-	MaxPool2D	MaxPool2D	-	-	-	Pool Size = 2, stride =2
Flatten	Flatten	Flatten	-	-	-	-
Dense	Dense	Dense	1 × 4096	1 × 4096	Tanh, Dropout Rate = 0.5	ReLU, Dropout=0.5
Dense	Dense	Dense	1 × 4096	1 × 4096	Tanh, Dropout Rate = 0.5	ReLU, Dropout=0.5
Dense	Dense	Dense	1 × {3,4}**	1 × {3,4}**	Softmax	Softmax

\* indicate that the parameters do not include VGG16. \*\* indicate the output for first and second classification stage.

### 3.1. ReLU layer

Real-world data learned by ConvNet are usually non-negative linear values, so there is a need to introduce nonlinearity in the ConvNet. The Rectified Linear Unit (Relu) is by far the most commonly used activation function in deep learning  $f(x) = \max(0, x)$  [38]. The Relu function returns the input directly if it receives any positive value  $x$ , but returns a zero for any negative value it receives [39]. Other functions like the saturating hyperbolic tangent ( $f(x) = \tanh(x)$ ) and sigmoid function  $f(x) = (1 + e^{-x})^{-1}$  can be applied to increase the nonlinearity. ReLU is often preferred over them because of its ability to perform better than the other two [40] [64]. Having a value of 0 on the negative axis means that Relu networks will run faster.

### 3.2. Pooling layer

Pooling is a form of nonlinear downsampling applied in CNNs, and among the number of nonlinear functions available, max pooling is the most commonly applied. Generally, max-pooling aggressively down sample feature maps by taking the largest element from the rectified feature map. The idea behind the use of pooling is to reduce the size of the features, minimize the number of parameters and reduce the amount of memory applied in running the calculations. Pooling also reduces the amount of computation in the network, while contributing to controlling overfitting.

## 4. Method

Since the aim of this work is to classify arrhythmia, ECG data is obtained from Physio-Bank in Physio-Net [32], which offers a huge collection of physiological signals. ECG data with AF, Normal, PAC, and PVC was obtained from the MIT-BIH Arrhythmia Database (MITDB) [41, 42] with a sampling frequency of 360 Hz. Out of the 48 half-hours, ECG recordings from subjects in the MITDB, data from four subjects did not take part in the research because they used a pacemaker. The first channel is used, which is the lead II recording. Among the 44 subjects from MITDB that took part in the research, eight of them were involved in AF classification, 39 took part in Normal, 27 took part in PAC classification, and 37 took part in PVC classification. However, the MITDB database lack data type to cover all the arrhythmia understudy, additional databases with AF, Noise, and VF data was also used from MIT-BIH Atrial Fibrillation Database (AFDB) [43], Creighton University Ventricular Tachyarrhythmia Database (CUDB) [44, 45], and MIT-BIH Malignant Ventricular Ectopy Database (VFDB) [46, 47] respectively. The sampling frequency used in AFDB, CUDB, and VFDB was 250 Hz. The MIT-BIH Atrial Fibrillation consists of 23 accessible long-term ECG recordings of subjects with atrial fibrillation. The CUDB incorporates 35 ECG recordings of subjects with ventricular tachycardia, ventricular flutter, and ventricular fibrillation. The VFDB incorporates 22 ECG recordings of subjects who experienced episodes of ventricular tachycardia, ventricular flutter, and ventricular fibrillation. More information about the databases is provided in Table 2. The table includes the databases used, the sampling frequency for each database, the total number of subjects that makes up the data in the database, and the number of subjects per database that took part in each type of arrhythmia.

Initially, the window size is selected to convert the ECG to images. The normal ECG consists of a normal series of the P, Q, R, S, T waves as shown in Figure 3. The normal sinus rhythm results from the occurrence of this pattern about 60-100 times per minute. Two-seconds segments are used to convert the 1D ECG signals to 2D images using recurrence plot, which helps provide a way to visualize the nature of a trajectory through phase space. The reason behind the selection is due to the fact that the number of R peaks to discriminate among the classes understudy is adequate. Figure 4 shows samples 2-second 1D ECG

segments for the six classes and their corresponding recurrence plot under study. They include data with (a) AF rhythm, (b) Normal beat, (c) PAC beat, (d) PVC beat, (e) VF, and (f) Noise respectively.

Table 2. Summary of databases used for each label. The databases including MIT-BIH Atrial Fibrillation Database (AFDB), Creighton University Ventricular Tachyarrhythmia Database (CUDB), MIT-BIH Arrhythmia Database (MITDB), and MIT-BIH Malignant Ventricular Ectopy Database (VFDB).

Database	Sample Hz	Total Subject	Number of Subjects / database					
			AF	N	PAC	PVC	VF	Noise
AFDB	250	23	8	-	-	-	-	-
CUDB	250	35	-	-	-	-	-	17
MITDB	360	44	8	39	27	37	-	-
VFDB	250	22	-	-	-	-	11	-

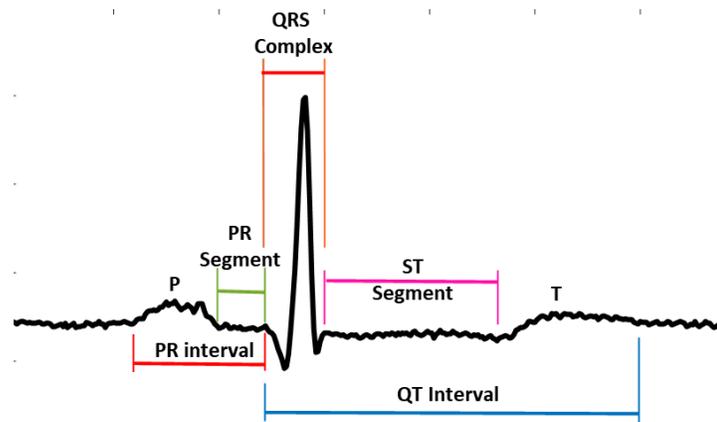


Figure 3. The normal ECG beat showing the normal series of the P, Q, R, S, T waves.

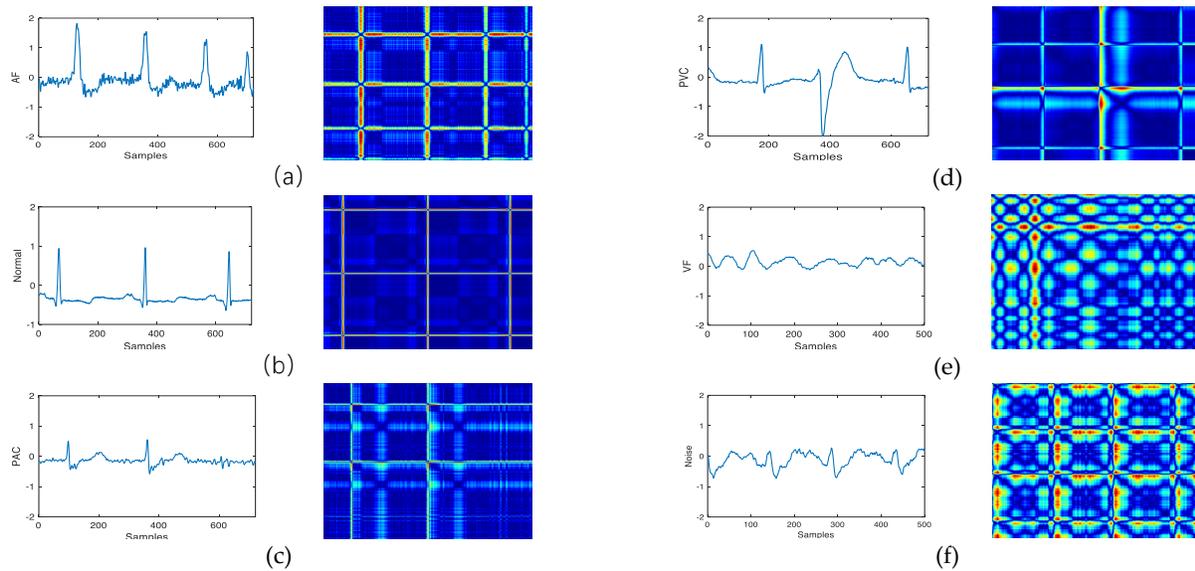


Figure 4. Typical ECG segments for the six classes that make up the 2-second segments and the corresponding recurrence plot. (a) 2-sec segment for AF, (b) 2-sec segment for Normal, (c) 2-sec segment for PAC, (d) 2-sec segment for PVC, (e) 2-sec segment for VF, and (f) 2-sec segment for Noise. The difference in sample size between ECG segments is caused by the difference in sampling rate between databases.

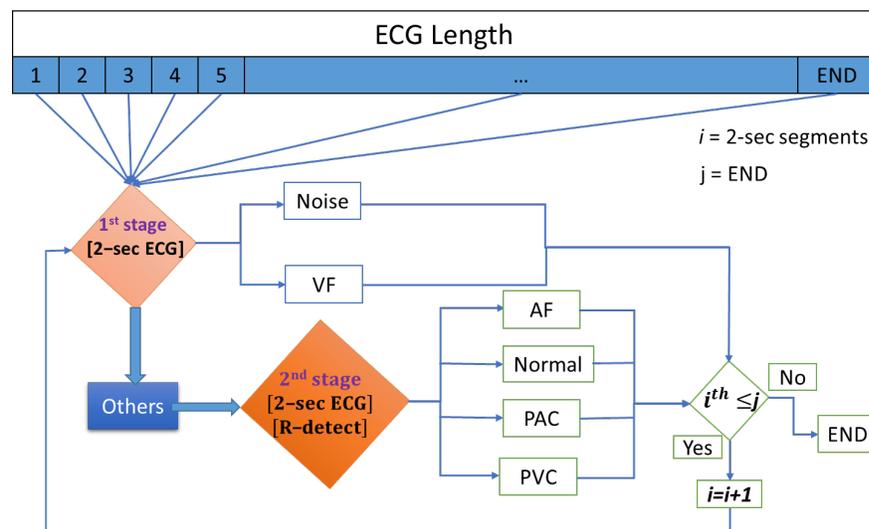
#### 4.1. Flowchart

The real-time application of arrhythmia discrimination should be able to segregate between different types of arrhythmia without prior knowledge of the subject's medical history. The chief purpose of the study is to develop an algorithm to be applied to devices and be able to detect and classify different types of arrhythmia in users. As shown in Figure 5, the designed flowchart illustrates the process of identifying the different types of arrhythmia. The classification procedure discriminates the data in two stages. Regardless of the type of arrhythmia, the two-second data segment of the ECG signal is converted to 2D using RP and input to the first stage. The first stage classifies the data between three classes, Noise, VF and Other segments. Noisy data can be difficult to deal with and the VF arrhythmia can be deadly, thus require immediate attention and discrimination in the first stage. The label Others in the first stage contains four different types of arrhythmia, which includes AF, Normal, PAC and PVC. If the algorithm detects either VF or Noise in the first stage, and more ECG data is available, it takes as an input the next 2-second segment of the ECG data to the first stage or otherwise ends the process. If classified in the first stage, the arrhythmia labeled Others goes to the second stage for further processing.

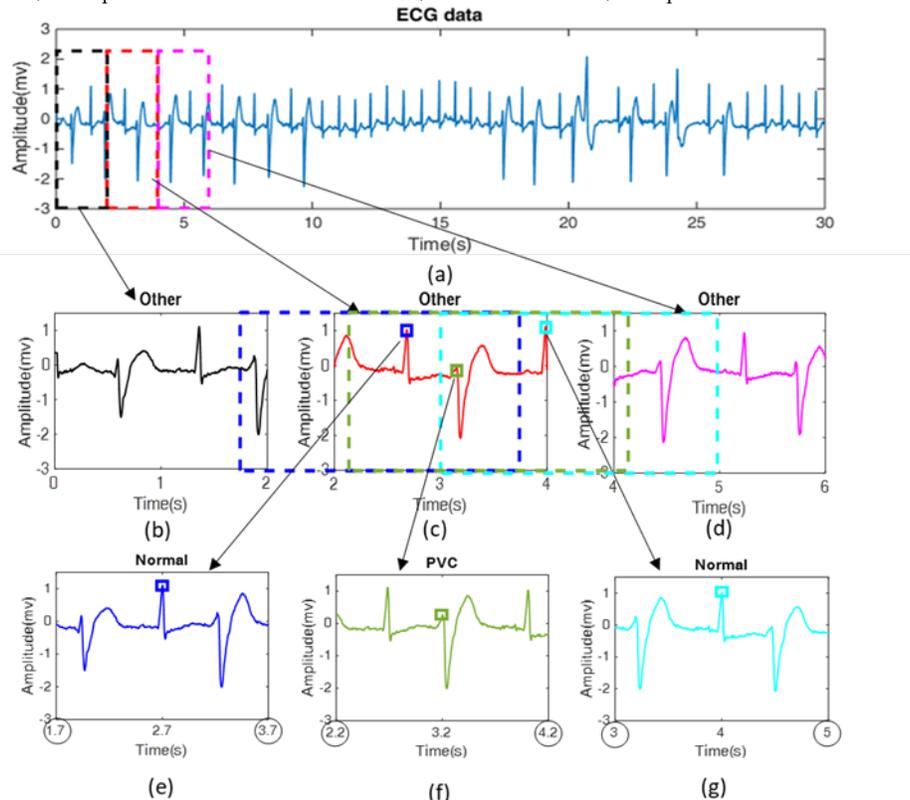
The second stage discriminates the data labeled Others in the first stage, into AF Normal, PAC, and PVC. In the second stage, the same 2-second ECG segment is converted into RP again. Converting the ECG segments to RP in the second stage requires the application of the R peaks detection algorithm [47] since the Noise and VF data has been classified in the first stage. To make the 2-second segments of the second stage, the 1-second equivalent data segment is taken from the ECG signal before and after the detected R-peak. This is done to make it easier for the model to distinguish between classes. A detailed explanation for the process is provided in the seven steps below.

The steps for the proposed method

- Segmenting ECG signals into 2-sec segments;
- Converting the 2-sec segments to recurrence plot for the first classification stage;
- Training the classifier to separate labels between Noise, VF and Other segments;
- Applying the R-peak detection algorithm on the labels classified as Others;
- Taking 1-sec equivalent of data before and after the R-peak to make a 2-sec segment of the second stage;
- Converting the 2-sec segments of the second stage to recurrence plot;
- Training the classifier for the classification of AF, Normal, PAC and PVC.



**Figure 5.** Flowchart of the proposed approach to classify different types of arrhythmia segments. AF: atrial Fibrillation; VF: ventricular fibrillation; PVC: premature ventricular contraction; NSR: normal sinus, PVC premature atrial contraction, Noise.



**Figure 6.** Typical signal samples for different classes. (a) Electrocardiogram (ECG) signal including different type of arrhythmia. (b-d) Zoomed-in non-overlapping segments (others) of the first stage. (e-g) Zoomed-in overlapping segments of the second classification stage of arrhythmia from (c). (e) Normal, (f) PVC, and (g) Normal. The black circle marks the beginning and end of each second stage segment.

Figure 6 shows a representation of ECG data from one of the subjects that took part in the research. The figure summarizes the procedure followed when taking the ECG segments in preparation for the signal to be converted into the images for the first and second classification stage using RP method. (a) is a 30-second representation of the subject's ECG signal with different types of arrhythmia. The 2-second segment is applied on the data to prepare it for the application for the first stage RP as marked by the dotted lines. The different colors represent different segments. The application of the 2-second segments in (a) results in different classes as represented in (b), (c), and (d). If classified as Others in the first stage, the R peak detection algorithm begins the process of preparing the data for the second stage. After the detection of the R peak, one-second equivalent of data is taken before and after the R peak to make a 2-second segment of the second stage. As shown in Figure 6, if the first segment is classified as Others, the R peak detection will not be applied on the first segment due to edge effect. When applied to the next segment (c) of the first stage, the first segment for the second-stage (e) will have data from both (b) and (c) as shown by the blue dotted lines. The starting and end time for each data segment is marked by a circle (O) on each segment. It is worth mentioning that it is possible that one segment can have more than one arrhythmia type. To avoid confusion, the segment is labeled using the annotation of the beat in the detected R-peak.

#### 4.2. Training on supercomputer

Training CNN can be very difficult and time-consuming, so high computing power is required for both accuracy and time-saving. Training CNNs with CPU is almost impossible depending on the size of the network and the data required. However, employing GPUs in training CNNs has a great effect on the desired accuracy and the time required for the process. Latter-days GPUs are very efficient in image processing and computer graphics manipulation for large data due to their parallel structure. Nvidia Tesla K40 GPU is used hosted in a supercomputer provided by Lenovo for testing the first stage of the classification. The supercomputer provided us with access to a GPU that was able to achieve a very appealing computational efficiency which allows training a model with 250 epochs within 8 hours. For the second classification stage, 2x Nvidia Tesla V100 is used with 32G vRam from Lenovo after they upgraded their supercomputer which allowed us to train one cross-validation in each GPU at a time and further reduce the computation time.

#### 4.3. Classification metrics

The classification evaluation of the proposed ECG arrhythmia discrimination is validated in six ECG labels from the MIT-BIH arrhythmia database including Normal, AF, PAC, PVC, Noise and VF samples. The labels are randomly selected into three subsets that include training, validation, and testing. Three measures are used to evaluate the performance of the proposed classification method, which includes sensitivity, positive predictive value (PPV), and Accuracy and defined in equations (3)-(5) below. The positive predictive value is the percentage of the total relevant results that are correctly classified, which is the number of correctly predicted labels divide by the total number of labels. On the other hand, sensitivity is the percentage of the results that is relevant, which is the number of correctly predicted labels divide by the number of generated labels. The accuracy is calculated as the sum of the correctly classified data divided by the total number of data.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$\text{PPV} = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (5)$$

where TP is the true positive rate, FN is the false-negative rate, TN is the true negative rate, and FP is the false positive rate.

## 5. Results

### 5.1. Choosing the optimum number of epochs

The convolutional network training procedure implemented in (Karpagachelvi et al., 2010) is adopted. A batch size of 128 is set throughout the training of the two stages. The first classification stage involves the classification of images into one of the three classes mentioned. A total of 29217 images are used for the first classification stage; 20531 Others, 4256 Noise, and 4430 VF. The second classification stage involves the classification of images into AF, Normal, PAC, and PVC. On the other hand, 19640 images are used in the second stage, from which 6488 from AF, 7228 for normal, 2559 for PAC and 3365 for PVC. Each class of images is trained and tested together, and performance evaluation is measured based on the accuracy of the predicted results. Two measurements have usually reported loss and accuracy. An image is correctly classified if the predicted class is the same as the label. To find the best model to use with the classification, the number of epoch is varied and the optimum is selected by comparing the performances. Number of epochs for training started at 100 epochs to 250 epochs and 350 epochs for first and second classification stage respectively. We used an arithmetic progression sequence with a common difference of 50 epochs to help us

find the optimum number of epochs. This section also shows data imbalance in the first stage, thus the reason for training only up to 250 epochs in the first stage.

**Table 3.** Evaluation of first stage model using the training, validation and testing accuracies to choose the optimum number of epochs

Epoch	Training			Validation			Testing		
	Alex	VGG16	VGG19	Alex	VGG16	VGG19	Alex	VGG16	VGG19
100	98.06%	96.91%	97.55%	94.45%	93.70%	94.21%	91.09%	87.35%	81.01%
150	99.16%	98.42%	97.01%	95.01%	94.04%	93.94%	89.20%	84.95%	<b>87.93%</b>
200	99.62%	99.63%	99.29%	95.39%	93.81%	94.17%	<b>92.61%</b>	81.08%	83.55%
250	99.47%	99.67%	99.71%	95.20%	93.92%	94.73%	91.83%	<b>89.92%</b>	84.21%

The bold numbers suggest the optimum number of epochs using the testing accuracy

### 5.1.1. Results for finding the optimum number of epochs for the first stage

Table 3 shows the performance evaluation of the models used to find the number of epochs that have a better classification performance. A comparison of the sensitivity and the positive predictive value for the number of epochs used is shown in Table 4. The highest performance model is selected for the first stage and cross-validation. The results for training and testing using a different number of epochs suggests that 200 epochs are the preferred optimum number of epochs to be used in the first classification stage using AlexNet model. The 200 epochs give better training, validation and testing performance in comparison with the 100, 150, and 250 epochs tried. On the other hand, results for testing VGG-16 and VGG-19 using different number of epochs suggests that 250 and 150 epochs are the preferred optimum number of epochs, with testing accuracy of 89.92% and 87.93% respectively. Comparison between architectures suggest that AlexNet is the preferred model architecture for the first classification stage of arrhythmia giving a better performs than the other two architectures (Table 3).

### 5.1.2. Results for finding the optimum number of epochs for the second stage

Table 5 shows the performance of the models used to find the number of epochs that performs better in the second classification stage. A comparison of the sensitivity and the positive predictive value for the above-varied epochs is shown in Table 6. The model with the highest performance is chosen for the second classification stage and did cross-validation. From the results in Table 5, it is clear that 150 epochs for AlexNet are the optimum number of epochs since it gives the highest testing accuracy over the other number of

**Table 4.** Evaluation of first stage model using the sensitivity and PPV to choose the optimum number of epochs.

Epoch	Sensitivity								
	AlexNet			VGG16			VGG19		
	Noise	Others	VF	Noise	Others	VF	Noise	Others	VF
100	97.50%	94.70%	48.02%	<b>99.22%</b>	<b>87.65%</b>	<b>63.28%</b>	99.22%	79.92%	57.63%
150	57.86%	99.93%	97.65%	97.79%	84.68%	63.84%	<b>96.56%</b>	<b>86.84%</b>	<b>81.92%</b>
200	<b>68.35%</b>	<b>99.68%</b>	<b>96.18%</b>	99.06%	80.28%	55.65%	99.84%	82.86%	60.17%
250	65.74%	99.62%	94.69%	97.50%	92.14%	56.78%	99.37%	84.03%	58.47%
PPV									
100	64.63%	99.22%	100%	<b>55.66%</b>	<b>99.93%</b>	<b>95.73%</b>	45.25%	99.88%	97.14%
150	99.53%	89.72%	77.36%	52.17%	99.73%	86.92%	60.55%	100.00%	76.12%

200	<b>97.65%</b>	<b>91.56%</b>	<b>92.64%</b>	45.87%	99.88%	89.55%	49.23%	99.92%	94.67%
250	96.09%	93.79%	79.23%	63.06%	99.48%	86.27%	<b>50.12%</b>	<b>99.88%</b>	<b>95.39%</b>

The bold numbers suggest the optimum number of epochs using the mean for specificity and the positive predictive value  
**Table 5.** Evaluation of second-stage model using the training, validation and testing accuracies to choose the optimum number of epochs

# of Epochs	Training			Validation			Testing		
	Alex	VGG16	VGG19	Alex	VGG16	VGG19	Alex	VGG16	VGG19
100	99.53%	98.80%	98.88%	99.26%	97.34%	97.45%	98.11%	86.86%	94.09%
150	99.77%	99.20%	99.11%	99.44%	97.32%	96.96%	<b>98.53%</b>	87.15%	95.13%
200	99.58%	99.33%	99.23%	98.20%	97.42%	97.45%	98.36%	<b>96.03%</b>	93.88%
250	99.70%	99.52%	99.47%	98.47%	97.46%	97.20%	98.11%	<b>96.03%</b>	<b>95.92%</b>
300	99.70%	99.55%	99.34%	98.12%	97.48%	97.12%	98.23%	90.83%	92.61%
350	99.74%	99.57%	99.58%	98.34%	97.17%	97.56%	98.44%	91.52%	89.70%

The bold numbers suggest the optimum number of epochs using the testing accuracy

epochs. For VGG-16 and VGG-19, 200 and 250 epochs respectively give the highest testing accuracy for both architectures, thus can be referred to as the optimum number of epochs for both architectures. The AlexNet model architecture give a better prediction alternative for the second classification stage. This is evidence from the results of the optimum number of epochs presented in Table 5.

Initially, AF data is utilized from AFDB together with Normal, PAC and PVC from MITDB for the second classification stage. After carefully considering that the MIT database also contains data with AF and the fact that the AF from AFDB (AF\_AFDB) is more serious than that from MITDB (AF\_MITDB), AF images are tested from MITDB with the model trained only on AF images from AFDB. This was done to test if AF\_MITDB will affect the accuracy of the model. Upon testing the model, a reduced testing accuracy of 63% is obtained on the AF\_MITDB, which proved the need to include AF\_MITDB on the training set. After adding the AF data from MITDB on the training set, and five classes on the testing sets, which included Normal, AF\_MITDB, AF\_AFDB, PAC, and PVC.

In a real-life scenario, types of arrhythmia are not limited to exactly the same kind that is experienced in every individual. To see if the model that is trained with data from a few subjects in the MITDB can perform well to any individual that is experiencing the same kind of arrhythmia, the performance of the model is evaluated by testing it on all the patients in the MITDB. Since the MITDB contains a large number of classes, some of which are not part of the study, specific focus is given to the performance of the model based on the classes of interest. Finally, the second stage is trained once more using part of the data in every patient in the MITDB database. Out of the 48 subjects in the MITDB, data from 44 subjects are used to train and test the second stage of arrhythmia. Four subjects including mitdb102, mitdb104, mitdb107, and mitdb217 did not take part in the study because they used a pacemaker.

## 5.2. Performance evaluation

This section summarizes the performance evaluation of the classification using the results obtained when training and testing the two stages. Data for the training, validation, and testing were distributed at 70%, 15% and 15% of the total data respectively. The testing results are used to evaluate the performance of the models of the two stages. The confusion matrix table is used to chart the predicted vs actual results of the classification. The confusion matrix table describes an output of negative vs positive outcome of the testing results. Table 7 shows the summary of the training time, and the memory requirements of the three architectures for the arrhythmia classification based on ECG RP images. The AlexNet model offers a better alternative in terms of computational cost and the memory requirements than both VGG models. Although

AlexNet is preferred in this study, it is hindered in small hardware because it is still needed high computational cost. In this regard, high performance hardware accelerators are needed for reducing both execution time, learning accuracy and power consumption [72].

**Table 6.** Evaluation of second-stage model using the sensitivity and PPV to choose the optimum number of epochs

Epoch	Sensitivity											
	ALEX				VGG16				VGG19			
	AF	Normal	PAC	PVC	AF	Normal	PAC	PVC	AF	Normal	PAC	PVC
100	95.92%	99.08%	94.72%	99.40%	91.93%	75.37%	77.84%	98.08%	95.09%	92.10%	86.54%	98.19%
150	<b>97.41%</b>	<b>99.89%</b>	<b>96.57%</b>	<b>99.37%</b>	90.02%	78.62%	76.78%	97.98%	95.01%	95.17%	86.54%	98.49%
200	97.61%	99.44%	97.55%	98.11%	<b>95.26%</b>	<b>99.00%</b>	<b>85.75%</b>	<b>97.28%</b>	92.35%	92.85%	88.13%	99.19%
250	97.09%	99.33%	97.06%	98.43%	91.18%	72.80%	73.61%	97.68%	<b>96.59%</b>	<b>95.01%</b>	<b>90.50%</b>	<b>98.29%</b>
300	98.85%	98.76%	97.06%	95.60%	95.42%	86.77%	81.53%	96.37%	86.02%	95.01%	88.13%	99.40%
350	98.13%	99.66%	97.55%	96.54%	95.42%	86.77%	81.53%	96.37%	96.42%	78.04%	82.85%	98.29%
PPV												
100	98.29%	98.67%	93.73%	97.43%	78.48%	97.11%	86.76%	88.94%	90.93%	98.14%	96.19%	92.85%
150	<b>97.41%</b>	<b>99.89%</b>	<b>96.57%</b>	<b>99.37%</b>	77.67%	98.03%	97.00%	86.94%	93.68%	96.46%	97.62%	94.49%
200	97.61%	99.44%	97.55%	98.11%	<b>94.01%</b>	<b>95.28%</b>	<b>99.39%</b>	<b>98.37%</b>	94.31%	97.72%	90.27%	90.61%
250	97.09%	99.33%	97.06%	98.43%	74.41%	98.54%	92.08%	87.22%	<b>92.29%</b>	<b>97.44%</b>	<b>98.00%</b>	<b>97.99%</b>
300	98.85%	98.76%	97.06%	95.60%	82.64%	96.93%	96.56%	96.47%	97.73%	93.91%	89.30%	87.49%
350	98.13%	99.66%	97.55%	96.54%	82.64%	96.93%	96.56%	96.47%	80.10%	99.26%	96.32%	92.24%

The bold numbers suggest the optimum number of epochs using the mean for specificity and the positive predictive value

**Table 7.** Training time and memory requirements of the three CNN architectures on the arrhythmia classification using 150 epochs

	AlexNet	VGG16	VGG19
Time (hrs)	5.37	7.6	7.75
Memory (MB)	228	525	545

### 5.2.1. Evaluation of the first classification stage

The evaluation of the first classification stage is based on the accuracy of the training, validation, and testing of the model. The accuracy of the testing results, the PPV and the sensitivity of the model are calculated from the predicted results using the classification matrices table. Table 8 illustrates the evaluation of the first stage cross-validation using the optimum number of epochs in AlexNet. The evaluation involves the training, validation and testing accuracy for the optimum number of epochs obtained earlier after the training and testing process. The first columns represent the cross-validation number and the rest of the columns are the accuracies for training, validation, and testing respectively. The last row gives the mean and standard deviation for the accuracy of the training, validation, and testing which show  $99.7 \pm 0.08\%$ ,  $97.60 \pm 0.25\%$ , and  $95.30 \pm 1.27\%$  respectively. Table 9 gives the evaluation of the performance of the first stage using the sensitivity and the PPV for the fivefold cross-validation. The evaluation also includes the mean and standard deviation for the sensitivity and PPV respectively. The mean and standard deviation of the sensitivities for the first classification stage are  $94 \pm 2.04\%$ ,  $92.3 \pm 3.62\%$ , and  $97.6 \pm 0.88\%$  for VF, Noise, and Others. The results for the mean and standard deviation of the PPV are  $93.4 \pm 2.82\%$ ,  $84 \pm 5.39\%$ , and  $99.9 \pm 0.06\%$  for VF, Noise, and Others respectively

### 5.2.2. Evaluation of the second classification stage

Table 10 illustrates the evaluation of the fivefold cross-validation of the second classification stage. It shows the, accuracy of the training, validation, and testing of the fivefold cross-validation for the second

**Table 8.** The evaluation of the first stage of arrhythmia classification using the training, validation and testing results of the fivefold cross-validation.

Cross validation	Training	Validation	Testing
CV_1	99.79%	97.84%	93.35%
CV_2	99.60%	97.32%	96.10%
CV_3	99.78%	97.58%	94.85%
CV_4	99.74%	97.49%	96.59%
CV_5	99.68%	97.91%	95.68%
Mean $\pm$ SD	99.7 $\pm$ 0.08%	97.6 $\pm$ 0.25%	95.3 $\pm$ 1.27%

classification stage using the optimum number of epochs obtained earlier in the training and testing process for AlexNet model. The first column represents the cross-validation number and the rest of the column represents the accuracies for training, validation, and testing respectively. The last row gives the mean and standard deviation for the accuracy of the training, validation, and testing with 99.61  $\pm$  0.09%, 98.75  $\pm$  0.41%, and 98.41  $\pm$  0.11% respectively.

**Table 9.** Shows the evaluation of the performance of the first stage using the PPV and the sensitivity for the fivefold cross-validation.

Labels	CV_1	CV_2	CV_3	CV_4	CV_5	Mean $\pm$ SD
Sensitivity						
VF	90.70%	95.85%	94.68%	95.18%	93.36%	94 $\pm$ 2.04%
Noise	95.77%	86.38%	92.25%	92.25%	94.60%	92.3 $\pm$ 3.62%
Others	96.63%	98.20%	98.44%	97.90%	96.59%	97.6 $\pm$ 0.88%
PPV						
VF	95.35%	88.50%	94.06%	93.78%	95.25%	93.4 $\pm$ 2.82%
Noise	77.13%	90.00%	87.14%	86.00%	79.64%	84 $\pm$ 5.39%
Others	99.95%	99.80%	99.90%	99.85%	99.90%	99.9 $\pm$ 0.06%

**Table 10.** Illustrates the evaluation of the fivefold cross-validation of the second classification stage.

Cross validation	Training	Validation	Testing
CV_1	99.77%	99.44%	98.53%
CV_2	99.53%	98.72%	98.53%
CV_3	99.61%	98.63%	98.32%
CV_4	99.58%	98.38%	98.32%
CV_5	99.56%	98.58%	98.36%
Mean $\pm$ SD	99.61 $\pm$ 0.09%	98.75 $\pm$ 0.41%	98.41 $\pm$ 0.11%

Table 11 gives the evaluation results of the fivefold cross-validation for the second classification stage using the sensitivity and the PPV method. The evaluation includes results of cross-validation for each of the four labels and the mean and standard deviation for the sensitivity and PPV of the fivefold cross-validation respectively. The mean and standard deviation for the sensitivity of the fivefold cross-validation of the second classification stage is 97.81  $\pm$  0.01%, 99.44  $\pm$  0.00%, 96.28  $\pm$  0.02%, and 97.99  $\pm$  0.01% for AF, Normal, PAC and PVC respectively. The mean and standard deviation of for the PPV of fivefold cross-validation gives a 98.75  $\pm$  0.00%, 99.19  $\pm$  0.00%, 96.98  $\pm$  0.01% and 95.53  $\pm$  0.01% for AF, Normal, PAC and PVC respectively.

### 5.3. Evaluation of the approach in preparation for real-life application

The classification method is used to prepare for real-life application by evaluating the performance of the two-classification stage on new data. 2800 ECG data segments from the six labels are used, two in the first stage and four from the second stage. 400 ECG data segments per label are applied, which includes Normal,

**Table 11.** Sensitivity and PPV for the second stage cross-validation.

Labels	CV_1	CV_2	CV_3	CV_4	CV_5	Mean $\pm$ SD
Sensitivity						
AF	97.09%	97.71%	98.23%	97.82%	98.23%	97.81 $\pm$ 0.00%
Normal	99.33%	99.44%	99.33%	99.78%	99.33%	99.44 $\pm$ 0.00%
PAC	97.55%	97.55%	94.61%	94.61%	97.06%	96.28 $\pm$ 0.02%
PVC	97.80%	99.06%	98.11%	98.11%	96.86%	97.99 $\pm$ 0.01%
PPV						
AF	98.84%	99.26%	98.34%	98.85%	98.44%	98.75 $\pm$ 0.00%
Normal	99.21%	99.10%	99.10%	99.00%	99.55%	99.19 $\pm$ 0.00%
PAC	94.31%	98.03%	97.97%	98.47%	96.12%	96.98 $\pm$ 0.02%
PVC	95.11%	95.17%	96.30%	94.83%	96.25%	95.53 $\pm$ 0.01%

AF\_AFDB, and AF\_MITDB, PAC, PVC, Noise, and VF. Since there are not enough ECG segments to cover the preferred amount of segments, part of the segments is included in the training process for other labels, which included PAC; it is decided to choose random subjects to cover the set number of segments. The ECG signals are applied to the first stage algorithm. The first stage algorithm discriminates the ECG signal into two-second segments, converts them to recurrence plot, and tests them using the best model of the first stage cross-validation. The first stage model discriminates the ECG segments into one of the first stage labels. If the segments were predicted as noise or VF, the algorithm takes the next ECG segment into the first stage, whereas if the segment is predicted as others, the algorithm sends the segment to the second stage. In the second stage, the algorithm detects the R-peak and takes one-second data before and after the R-peak and converts it to recurrence plot to classify it between Normal, AF, PAC or PVC. After classifying the segment into one of the labels in the second stage, the algorithm goes back to take as an input the next two-second segment of the ECG. If there are no segments left, the algorithm terminates the process.

To improve the PPV for noise, the voting method is used to test the results of the first stage and the best model for the second stage. The voting method in testing the first stage allows the application of the five models of the cross-validation to vote for the best likely to be a classified label in the first stage. After advice from the Food and Drug Administration [48], the PPV for noise and VF are improved by merging the results from VF and Noise since they are both discriminated in the first stage. Results of AF\_MITDB and AF\_AFDB are merged to have one AF label. Table 12 gives the evaluation of the three approaches in a real-life scenario. The evaluation includes the results of the sensitivity and PPV for the best models of the first stage and the second stage. It also gives the evaluation of the results after applying the voting method in the first stage and the best model in the second stage. Lastly, it gives the results of the evaluating of the two stages after merging the labels including merging VF and Noise in the first stage. The results give the testing accuracy of 85.29 % for the best model. The results of the sensitivity are 96.50%, 90.38%, 75%, 99.50%, and 72.63% for Normal, AF, PAC, PVC, and VF/Noise respectively. The results of the PPV are 92.79%, 87.53%, 95.24%, 94.31%, and 71.05% for Normal, AF, PAC, PVC, and VF/Noise respectively. The testing accuracy after applying the voting method in the first stage was improved to 88.28% and sensitivities are 96.50%, 87.75%, 91.25%, 96.50%, 99.25%, and 73.25% respectively. The PPV gives 93.92%, 98.08%, 96.74%, 94.98%, and 69.15% respectively. The classification metrics after merging the noise with VF from the first stage gives an improved testing accuracy of 92.96%. The sensitivities become 97.25%, 85%, 96.5%, 98.75%, and 94.13% for Normal, AF, PAC, PVC, and VF/Noise respectively. The PPV become 93.51%, 97.7%, 96.26%, 94.72%, and 86.55% for Normal, AF, PAC, PVC, and VF/Noise respectively.

The MITDB data was annotated by a slope sensitive QRS detector for normal beats, and further annotated by two cardiologists for other additional labels like those that could not be identified by the detector and included rhythms, comments, and signal quality. The MITDB contains more labels than were involved in the study including Aberrated atrial premature beat (a), Nodal (junctional) premature beat (J), Supraventricular premature beat (S), Nodal (junctional) escape beat (j), Left bundle branch block (L) and Right

**Table 12.** Evaluation of the second stage model using the best model, voting method between the models of the fivefold cross-validation and merging the VF and noise label. The evaluation gives sensitivity and PPV.

Method	Normal	AF	PAC	PVC	VF/Noise	Normal	AF	PAC	PVC	VF/Noise	Accuracy
	Sensitivity					PPV					
Best model	96.50%	90.38%	75.00%	99.50%	72.63%	92.79%	87.53%	95.24%	94.31%	71.05%	85.29%
Voting	96.50%	89.50%	96.50%	99.25%	73.25%	93.92%	98.08%	96.74%	94.98%	69.15%	88.28%
Merging VF& Noise	97.25%	85.00%	96.50%	98.75%	94.13%	93.51%	97.70%	96.26%	94.72%	86.55%	92.96%

bundle branch block (R) and Fusion of paced and normal beat (f) to mention just a few. The same performance evaluation procedures were carried out to the second stage analysis of arrhythmia to validate the algorithm before real-life application in electronic devices. This time the second classification stage is tested using the MITDB database without labels and cross-checked the results of the classification with the annotation types provided in Physio-Net. Since the MITDB contains a number of labels for the ECG, some of which are not among the labels included in the study, data with labels that were not part of the study are ignored. Upon evaluating the performance of the second classification stage by testing the second stage model on the MITDB database, the model performed poorly on the segments that had Left bundle branch block (L) and Right bundle branch block (R) as part of the Normal label. There are four subjects in MITDB that have L labels which include mitdb109, mitdb111, mitdb207, and mitdb214 and there are six that contain R labels with mitdb118, mitdb124, mitdb207, mitdb212, mitdb231, and mitdb232 among them.

To improve the performance, the Left Bundle Branch Block and the Right Bundle Branch Block are included as independent labels in the training sets to have six labels. The classification evaluation for the second stage after having tested the model using the MITDB database with six labels gives a gross, which is used to calculate the sensitivity, PPV and the accuracy of the classification. Calculating the gross involves summing up the predicated results of all the labels separately and finding the sensitivity, PPV and the testing accuracy of the labels. The confusion matrix for the gross is given in Table 13. Table 14 gives the sensitivity and the PPV of the gross. The sensitivities are 79.83%, 98.97%, 96.35%, 95.81%, 95.10%, and 99.24% for AF, L, Normal, PAC, PVC, and R respectively. The PPVs are 94.98%, 91.06%, 98.88%, 64.84%, 91.37%, and 87.81% for, AF, L, Normal, PAC, PVC, and R respectively. The testing accuracy of the classification is 95.10%.

Table 15 compares the results of the evaluation of the classification before and after introducing the Left and the Right bundle branch blocks in the training set. The table gives the results of the testing using the confusion matrix used to calculate the gross. For comparison purposes, the results for L and R labels are merged with Normal labels to have the same number of labels in the second stage. The evaluation of the before including the L and R labels are 86.79%, 87.83%, 97.68%, and 97.31% for AF, Normal, PAC, and PVC respectively, while the sensitivity after including the L and R labels as part of the normal label were found second stage using the whole MITDB database includes the calculation of the Gross from the confusion matrix. Table 16 gives the comparison of the sensitivity and PPV for the evaluation of the second classification stage using the MITDB before and after including the L and B label as part of the Normal label. The sensitivity

**Table 13.** The gross to evaluate the performance of the second stage using the six labels after including L and R labels in the training set.

Predicted	AF	L	Actual Normal	PAC	PVC	R
-----------	----	---	------------------	-----	-----	---

AF	7500	1	288	21	82	4
L	374	7975	270	5	126	8
Normal	622	0	62151	31	52	2
PAC	439	0	863	2449	19	7
PVC	347	1	223	13	6546	34
R	113	81	709	37	58	7188

**Table 14.** Performance of the second stage using sensitivity and PPV of the six labels of the MITDB data

Evaluation	AF	L	Normal	PAC	PVC	R	Accuracy
Sensitivity	79.83%	98.97%	96.35%	95.81%	95.10%	99.24%	95.10%
PPV	94.98%	91.06%	98.88%	64.84%	91.37%	87.81%	

**Table 15.** Confusion matrix for the gross of the testing data from the whole MITDB database before and after including l and r in the training data

Predicted	Actual			
	AF	Normal	PAC	PVC
Training without L and R				
AF	8304	2786	99	92
Normal	557	68316	62	74
PAC	374	5355	2481	28
PVC	333	3093	7	6698
Training with L and R				
AF	7921	1988	54	95
Normal	916	76552	83	271
PAC	385	826	2441	18
PVC	346	242	8	6512

**Table 16.** Comparing the performances of stage 2 using sensitivity and PPV of the four labels of the MITDB data

Label	Sensitivity	PPV	Sensitivity	PPV
	Training without L and R		Training with L and R	
AF	86.79%	73.61%	82.79%	78.75%
Normal	85.88%	99.00%	96.16%	98.37%
PAC	93.66%	30.12%	94.39%	66.51%
PVC	97.19%	66.11%	94.43%	91.62%
Accuracy	86.97%		94.70%	

to be 82.79%, 98.22%, 96.10%, and 94.61% respectively. The PPVs before including L and R labels are 87.08%, 99.06%, 29.72%, and 88.65% for AF, Normal, PAC and PVC respectively, while the PPVs for the classification after including the L and R labels are 95.28%, 98.41%, 65.69%, and 91.45% for AF, Normal, PAC, and PVC respectively. The testing accuracy for the second classification stage before and after introducing the L and R labels as part of the Normal label are 88.65% and 96.41% respectively.

## 6. Discussion

This paper introduces an approach for ECG arrhythmia classification that can be used in electronic devices with no prior feature extraction on the ECG signal. Although RP is one form of feature extraction, it only shows the image not the details and mathematical calculation of the features. This extensive approach utilizes data from four databases accessible from the Physio-Net Physio-Bank archive, which includes the AFDDB, CUDB, MITBD, and VFDB [41, 46]. It is believed that making use of different databases and making sure that almost every subject in the databases takes part in the work will strengthen the ability

of the classifier to function properly in real-life scenarios. The other reason is this; more than one database is used due to the fact that the MITDB does not include all the arrhythmia under this study.

Past research in this field applied short and long duration ECG signal segments successfully, ranging between 0.33s to 10s ECG signal segments [55,67,69,70]. To our knowledge, there is no prior research in the study that focuses on the analysis of 2-second ECG segments. We chose the short duration ECG signal that focuses on the same segment length (2s) throughout the classification. The reason for the 2-second segment was in consideration to some of the features in the types of arrhythmia found in the structures of ECGs [65]. We aimed at applying the duration of a segment short enough to expose most of the features that are required to discriminate the different types of arrhythmia under study. AF and VF [71], which form an essential part of the arrhythmia under study, are annotated using rhythms and require a few more beats to be identify from the ECG signals, while others (Normal, PAC and PVC) are annotated using beats and can be discriminated in one beat. As mentioned in section 1, the normal heart rate for a healthy person is between 60-100 beats per minutes. The 2-second segment is enough to expose about three R-peaks in each segment, which is enough to classify between the types of arrhythmia that have rhythms of AF and VF on ECG signals.

This work approach is based on two-stage classification, which uses a segment-by-segment classification in the first stage and a beat-by-beat classification with an overlapping segment in the second stage. The reason for not using a direct classification approach in this work was the necessity to apply the R-peak detection algorithm to strengthen the ability for the model to effectively discriminate between different types of arrhythmias. However, since VF and Noise ECG signals does not always have R-peak, we designed the segment-by-segment classification in the first stage to discriminate the two arrhythmia from the rest before we applied the R-peak detection algorithm in the second stage. Another reason was that we wanted to design a model that will be able to be applied to patients without prior knowledge of the type of arrhythmia they are experiencing. On the other hand, direct approaches of one stage are mostly designed for specific arrhythmias.

As mentioned in the previous section, the proposed approach uses RP to represent the ECG segments in the classification [32]. The first stage uses segment-by-segment approach to convert the ECG signal to RP while the second stage used beat-by-beat approach. Since there are not enough segments for balanced data set for all the labels of arrhythmia under study. We took overlapping windows for the ECG segments of the second stage to have near-balanced data sets. Occasionally, some record of ECG signals includes other arrhythmia in the same segment as seen in Figure 6. This is because the 2-second segments are big enough to allow a number of beats in one segment depending on the nature of the signal in the subject. In many cases, we could not avoid that due to insufficient data in other arrhythmia. As much as having overlapping windows seems like a challenge on the selection of the window size, it also presents some positive aspects. It allows the model to give predictions to every beat in the second classification stage of arrhythmia. Another positive thing about having an overlapping window that contains more than one type of arrhythmia in one segment is that it strengthens the ability of the model to classify similar case if present in the subject's ECG signal. The approach takes 2-second ECG segments, converts it to RP, and classifies it in two stages. The fact that the approach takes two-second segments and that this is no time lag between each data segment shows that there is less classification time needed compared to other methods [49].

This paper applies data reconstruction method (RP) to prepare the data for training, while also reducing the need for complex feature processing and calculations. We tried different CNN architectures to explore the reliability and effectiveness of this approach. Three CNN architectures were compared in this work with the usage of GPU environment, which help reduce the computation time. However, since this main focus was not on designing a new CNN architecture, we modified existing CNN model [28] [64] to adapt to the ECG arrhythmia classification in this work. The AlexNet model presents a better classification accuracy compared to the other two models applied earlier in the study (Tables 3 and 5). For that reason, we chose the AlexNet model for both the first stage and the second classification stage. Table

7 provides more evidence that AlexNet is a preferred architecture for the classification of arrhythmia in the study. It shows that AlexNet require less training time and less memory for the storage of the model which reduces the cost of acquiring costly devices. The results from the research also shows that CNN applied on RP based ECG segments provides an efficient approach in arrhythmia classification field, with a mean for the second stage cross validation accuracy of  $98.41\% \pm 0.11\%$  (Table 10). From the second stage evaluation in Table 15, it is worth mentioning that the classifier does well in a balanced number of data. The sensitivity and PPV for the balanced data in the training sets provide accuracies that are above 85% before and after including the L and R labels as part of the normal label except for the PAC. PAC label provided lower PPV accuracies of 29.72% and 65.69% before and after introducing the L and R label respectively. The reason behind the low accuracies is due to the very low number of PAC labels available in the databases.

## 7. Conclusions

In this paper, an effective way of converting 1D ECG signals is proposed using RP, method to 2D segments to enhance arrhythmia classification using a convolutional neural network. Two-second ECG segments were converted into recurrence plot in the first classification stage. The second stage applies the R-peak detection algorithm on the two-second segments and converts the ECG to recurrence plot to improve the classification.

To evaluate the classification, part of the data was set apart for discriminating the two types of arrhythmia in the first stage and the entire MITDB database was used to discriminate between the four types of arrhythmia in the second stage. The average testing accuracy for the fivefold cross-validation is  $95.3 \pm 1.27\%$  and  $98.41 \pm 0.11\%$  for the first and second stage respectively. The classification results performed better compared to the authors' previous work and other works that use CNN to discriminate between ECG abnormalities using 1D ECG data [50-54]. This confirms that converting the 1D ECG signal to 2D segments provides deep learning models with better access to learn the important features of the ECG data. This work also had better performance than [67-69], and the same average performance as [55], which used 2D ECG segments to classify arrhythmia despite the fact that they used larger ECG segments and [55] used more classification stages than the current work. The results of this work provide evidence that converting ECG data to images using recurrence plot plays a vital role in improving the accuracy of arrhythmia classification.

While the aim of developing an algorithm is expressed to provide a comprehensive arrhythmia discrimination approach for electronic devices, this approach encountered three limitations that influence the effectiveness of the pursuit. First, the data imbalance of the MITDB database is widely used in similar studies, and the number of arrhythmia types is not equally distributed in the database. PAC episodes for instance in Table 11, are relatively small in numbers compared to the rest in this study. Second, different forms of annotation of AF episodes are annotated using rhythm while the others are annotated using beat-by-beat annotation. As shown in Table 11, the short duration classification for AF is not enough to discriminate AF effectively from the rest of the arrhythmia types. Third is the computational cost and memory requirement. In future work, we are interested in applying transfer learning [66], which can provide better classification accuracies using relatively small amount of data compared to training from scratch. This will help us in maintaining the balance between datasets in all the classes. We are also interested in adding the third classification stage that will introduce long duration ECG segments for a better discrimination between AF and PAC [67]. To address the concerns of high computational cost and memory requirement, which limits the deployment of effective wide and deep models in resource-constrained environments, we are interested in applying different architecture designs like GoogleNet and ResNet that are capable of overcoming the shortcomings of the proposed.

**Acknowledgments:** This work was financially supported by AI R&D Department, New Era AI Robotic Inc., Taipei City, Taiwan. This work was also supported in part by the Ministry of Science and Technology, Taiwan.

**Author Contributions:** B.M.M., C.-H.L., and Y.-T.L. developed the algorithms and analyzed the data. B.M.M. wrote the paper. J.-S.S., M.F.A., evaluated and supervised the study.

**Conflicts of Interest:** The authors declare no conflict of interest

## References

1. D. L. Longo, A. S. Fauci, D. L. Kasper, S. Hauser, J. L. Jameson, and J. Loscalzo. "Harrison's principles of internal medicine." NEW YORK: McGraw-Hill CO 6, no. 1 (2012): 312.
2. I. Graham, D. Atar, K. Borch-Johnsen, G. Boysen, G. Burell, R. Cifkova, J. Dallongeville, G. De Backer, S. Ebrahim, B. Gjelsvik, and C. Herrmann-Lingen. European guidelines on cardiovascular disease prevention in clinical practice: executive summary: Fourth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (Constituted by representatives of nine societies and by invited experts). *European heart journal*, 28(19), pp.2375-2414, 2007.
3. A. S. Go, D. Mozaffarian, V. L. Roger, E. J. Benjamin, J. D. Berry, W. B. Borden, D. M. Bravata, S. Dai, E. S. Ford, C. S. Fox, and S. Franco, 2013. Executive summary: heart disease and stroke statistics—2013 update: a report from the American Heart Association. *Circulation*, 127(1), pp.143-152.
4. G. E. A. P. A. Batista, E. J. Keogh, A. Mafra-Neto, and E. Rowton, "Sensors and software to allow computational entomology, an emerging application of data mining," in Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011, pp. 761–764.
5. D. F. Silva, V. M. Souza, G. E. Batista, and R. Giusti, "Spoken digit recognition in Portuguese using line spectral frequencies," in Advances in Artificial Intelligence – IBERAMIA 2012, ser. Lecture Notes in Computer Science, J. Pavan, N. Duque-Mendez, and R. Fuentes-Fernandez, Eds. Springer Berlin Heidelberg, 2012, vol. 7637, pp. 241–250.
6. Y. Hao, B. J. L. Campana, and E. J. Keogh, "Monitoring and mining insect sounds in visual space." in Proceedings of the 12th SIAM Conference on Data Mining, 2012, pp. 792–803
7. A. Bagnall, L. M. Davis, J. Hills, and J. Lines, "Transformation based ensembles for time series classification," in Proceedings of the 12th SIAM International Conference on Data Mining, 2012, pp. 307–318
8. J. Lines, L. M. Davis, J. Hills, and A. Bagnall, "A shapelet transform for time series classification," in The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, pp. 289–297.
9. S. Sinha, P. S. Routh, P. D. Anno, & J. P. Castagna, Spectral decomposition of seismic data with continuous-wavelet transform. *Geophysics*, 70(6), P19-P25, 2005.
10. J. M. Anumonwo, and J. Kalifa, "Risk factors and genetics of atrial fibrillation," *Heart failure clinics*, 12(2), 2016, pp.157-166.
11. T. N. Nguyen, S. N. Hilmer, R. G. Cumming, "Review of epidemiology and management of atrial fibrillation in developing countries". *International Journal of Cardiology*. 167 S. Hajeb-Mohammadalipour, M. Ahmadi, R. Shahghadami, and K. Chon, "Automated Method for Discrimination of Arrhythmias Using Time, Frequency, and Nonlinear Features of Electrocardiogram Signals," *Sensors*, 18(7), p.2090, 2018. (6), 2412–20, 2013.
12. T. M. Munger, L. Q. Wu, W. K. Shen, "Atrial fibrillation". *Journal of Biomedical Research*. 28, (1), 1–17, January 2014.
13. J. M. Anumonwo, J. Kalifa, "Risk factors and genetics of Atrial Fibrillation". *Cardiology Clinics*. 32(4): November 2014.
14. T. N. Nguyen, S. N. Hilmer, R. G. Cumming, "Review of epidemiology and management of atrial fibrillation in developing countries". *International Journal of Cardiology*. 167 (6): 2412–20, September 2013.
15. P. Nickolls, R. M. T. Lu & K. A. Collins, "Apparatus and method for antitachycardia pacing using a virtual electrode" Nickolls, P., Lu, R.M., Collins, K.A., McCulloch, R.M., Cheatle, L.M. and Cleland, B., *Teletronics*

- Pacing Systems Inc, 1993. Apparatus and method for antitachycardia pacing using a virtual electrode. U.S. Patent 5,181,511.
16. B. Akdemir, H. Yarmohammadi, M. C. Alraies, W. O. Adkisson, "Premature ventricular contractions: Reassure or refer?" *Cleveland Clinic Journal of Medicine*, 83 (7): 524–530, 2016.
  17. K. Najarian, R. Splinter, "Biomedical Signal and Image Processing," second ed., CRC Press, Taylor and Francis Group, Boca Raton, 2012.
  18. U. R. Acharya, J. S. Suri, J. A. E. Spaan, S. M. Krishnan, "Advances in Cardiac Signal Processing," Springer-Verlag Berlin Heidelberg, New York, 2007.
  19. A. L. Goldberger, "Clinical electrocardiography: A simplified approach," Mosby, St. Louis, MO, USA, 2012
  20. S. Kiranyaz, T. Ince, M. Gabbouj, "Real-time patient-special ECG classification by 1-D convolutional neural networks," *IEEE Transactions on Biomedical Engineering* 63(3):664-675, 2016.
  21. P. Rajpurkar, A. Y. Hannun, M. Haghpanshi et al, "Cardiologist-level arrhythmia detection with convolutional neural networks," arXiv preprint arXiv:1707.01836, 2017.
  22. T. J. Jun, H. M. Nguyen, D. Kang, D. Kim, D. Kim, and Y. H. Kim, "ECG arrhythmia classification using a 2-D convolutional neural network," arXiv preprint arXiv:1804.06812, 2018.
  23. S. Anwar, K. Hwang, and W. Sung, "Fixed point optimization of deep convolutional neural networks for object recognition," In 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (pp. 1131-1135) April. 2015,.
  24. M. Liang, and X. Hu, "Recurrent convolutional neural network for object recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3367-3375), 2015.
  25. T. Ishii, R. Nakamura, H. Nakada, Y. Mochizuki, and H. Ishikawa, "Surface object recognition with CNN and SVM in Landsat 8 images," In 2015 14th IAPR International Conference on Machine Vision Applications (MVA) (pp. 341-344), May. 2015, IEEE.
  26. S. Z. Mahmoodabadi, A. Ahmadian, and M. D. Abolhasani, "ECG feature extraction using Daubechies wavelets," In Proceedings of the Fifth IASTED International Conference on Visualization, Imaging and Image Processing (pp. 343-348), September 2005.
  27. S. Z. Mahmoodabadi, A. Ahmadian, M. D. Abolhasani, M. Eslami, and J. H. Bidgoli, "ECG feature extraction based on multiresolution wavelet transform," In 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (pp. 3902-3905), January 2006. IEEE.
  28. S. Karpagachelvi, M. Arthanari, and M. Sivakumar, "ECG feature extraction techniques-a survey approach," arXiv preprint arXiv:1005.0957, 2010.
  29. P. M. Agante, and J. M. De Sá, "ECG noise filtering using wavelets with soft-thresholding methods," In *Computers in Cardiology* 1999. Vol. 26 (Cat. No. 99CH37004) (pp. 535-538). 1999 IEEE.
  30. G. Lu, J. S. Brittain, P. Holland, J. Yianni, A. L. Green, J. F. Stein, T. Z. Aziz, and S. Wang, "Removing ECG noise from surface EMG signals using adaptive filtering," *Neuroscience letters*, 462(1), pp.14-19, 2009.
  31. K. M. Chang, and S. H. Liu, "Gaussian noise filtering from ECG by Wiener filter and ensemble empirical mode decomposition," *Journal of Signal Processing Systems*, 64(2), pp.249-264, 2011.
  32. M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, 321, pp.321-331, 2018.
  33. C. Wigington, S. Stewart, B. Davis, B. Barrett, B. Price, and S. Cohen, "Data augmentation for recognition of handwritten words and lines using a CNN-LSTM network," In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 639-645) November. 2017. IEEE.
  34. J. Guo, and S. Gould, "Deep CNN ensemble with data augmentation for object detection," arXiv preprint arXiv:1506.07224, 2015.
  35. A. T. Monk and A. H. Compton, "Recurrence phenomena in cosmic ray intensity," *Reviews of Modern Physics*, vol. 11, no. 3–4, pp. 173–179, 1939.
  36. J. P. Eckmann, O. S. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, vol. 4, no. 9, pp. 973–977, 1987.

37. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," In *Advances in neural information processing systems*, pp. 1097-1105, 2012.
38. A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks" *Communications of the ACM*. 60 (6): 84–90, (2017-05-24)..
39. V. Romanuke, "Appropriate number and allocation of ReLUs in convolutional neural networks," *Naukovi Visti NTUU KPI*, (1), pp.69-78.
40. A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks" (PDF). *Advances in Neural Information Processing Systems*. 1: 1097–1105, 2012.
41. A. L. Goldberger, L. A. 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. *Circulation*, 101, e215–e220, 2000.
42. G. B. Moody, R. G. Mark, "The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* 2001, 20, 45–50. [CrossRef] [PubMed]
43. G. B. Moody, R. G. Mark, "A new method for detecting atrial fibrillation using R-R intervals," *Computers in Cardiology*. 10:227-230 (1983).
44. F. M. Nolle, F. K. Badura, J. M. Catlett, R. W. Bowser, M.H. Sketch, CREI-GARD, "A new concept in computerized arrhythmia monitoring systems," *Comput. Cardiol*, 13, 515–518, 1986.
45. G. B. Moody, W. Muldrow, R. G. Mark, "A noise stress test for arrhythmia detectors," *Comput. Cardiol.*, 11, 381–384, 1984.
46. Association for the Advancement of Medical Instrumentation. *Testing and Reporting Performance Results of Cardiac Rhythm and st Segment Measurement Algorithms; ANSI/AAMI EC57; Association for the Advancement of Medical Instrumentation: Arlington, VA, USA, 2012.*
47. J. T. Tang, X. L. Yang, J. C. Xu, Y. Tang, Q. Zou, and X. K. Zhang, "The algorithm of R peak detection in ECG based on empirical mode decomposition," *Fourth International Conference on Natural Computation (Vol. 5, pp. 624-627)*, 2008.
48. Association for the Advancement of Medical Instrumentation, 1998. "Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms," *ANSI/AAMI EC38*, 1998.
49. Hajeb-Mohammadalipour, M. Ahmadi, R. Shahghadami, and K. Chon, "Automated Method for Discrimination of Arrhythmias Using Time, Frequency, and Nonlinear Features of Electrocardiogram Signals," *Sensors*, 18(7), p.2090, 2018.
50. M. Sadrawi, C. H.Lin, Y. T.Lin, Y. Hsieh, C. C. Kuo, J. Chien, K. Haraikawa, M. Abbod, and J. S. Shieh, "Arrhythmia evaluation in wearable ECG devices," *Sensors*, 17(11), p.2445, 2017.
51. T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Transactions on Industrial Electronics*, 63(11), pp.7067-7075, 2016.
52. P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, and A.Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.
53. U.R. Acharya, H. Fujita, O.S. Lih, M. Adam, J.H. Tan, and C.K. Chua, "Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network," *Knowledge-Based Systems*, 132, pp.62-71, 2017.
54. U.R. Acharya, H. Fujita, S.L. Oh, Y. Hagiwara, J.H Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Information Sciences*, 415, pp.190-198, 2017.
55. U.R. Acharya, H. Fujita, O.S. Lih, Y. Hagiwara, J.H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network.," *Information sciences*, 405, pp.81-90, 2017
56. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, *Recurrence plots for the analysis of complex systems. Phys. Rep.* 438(5–6), 237–329 (2007)
57. G. Robinson, M. Thiel, *Recurrences determine the dynamics. Chaos* 19, 023104 (2009)
58. Y. Hirata, S. Horai, K. Aihara, *Reproduction of distance matrices from recurrence plots and its applications. Eur. Phys. J. Spec. Top.* 164(1), 13–22 (2008)

59. J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, Recurrence plots of dynamical systems. *Europhys. Lett.* 4, 973–977 (1987)
60. N. Marwan, How to avoid potential pitfalls in recurrence plot based data analysis. *Int. J. Bifurcat. Chaos* 21(4), 1003–1017 (2011)
61. L. Matassini, H. Kantz, J.A. Hołyst, R. Hegger, Optimizing of recurrence plots for noise reduction. *Phys. Rev. E* 65(2), 021102 (2002)
62. S. Schinkel, O. Dimigen, N. Marwan, Selection of recurrence threshold for signal detection. *Eur. Phys. J. Spec. Top.* 164(1), 45–53 (2008)
63. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., "ImageNet large scale visual recognition challenge", *International Journal of Computer Vision*, vol. 115, no. 3, pp. 221-252, 2015.
64. K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556. 2014 Sep 4.
65. Z. Binici, T. Intzilakis, O. W. Nielsen, L. Kober, A. Sajadieh, "Excessive supraventricular ectopic activity and increased risk of atrial fibrillation and stroke," *Circulation* 121, no. 17, p. 1904, 2010
66. W. Dai, Q. Yang, G. R. Xue, and Y. Yu, "Boosting for transfer learning." In *Proceedings of the 24th international conference on Machine learning*, pp. 193-200. 2007.
67. O. Yıldırım, P. Pławiak, R. S. Tan, and U.R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals." *Computers in biology and medicine* 102, pp. 411-420, 2018,
68. P. Pławiak, "Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system," *Expert Syst. Appl.* 92, pp. 334–349, 2018
69. P. Pławiak, "Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals," *Swarm and Evolutionary Computation* 39 pp. 192–208, 2018
70. M. Llamedo, J.P. Martinez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 58, pp. 616–625, 2011
71. L. Sörnmo, editor, "Atrial Fibrillation from an Engineering Perspective," Springer; May 15. 2018
72. X. Geng, J. Lin, B. Zhao, A. Kong, M. M. Aly, V. Chandrasekhar, "Hardware-Aware Softmax Approximation for Deep Neural Networks." In *Asian Conference on Computer Vision*, pp. 107-122, Springer, Cham, Dec 2, 2018.