

Article

An Automatic Diagnosis of Arrhythmias Using a Combination of CNN and LSTM Technology

Zhenyu Zheng ¹, Zhencheng Chen ^{1,2,3,*}, Fangrong Hu ¹, Jianming Zhu ^{2,3}, Qunfeng Tang ¹ and Yongbo Liang ^{1,2,3,*}

¹ School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China; zzy6951@163.com (Z.Z.); hufangrong@sina.com (F.H.); tangqunfeng1771@foxmail.com (Q.T.)

² School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, China; zjmcsu@126.com

³ Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin 541004, China

* Correspondence: chenzhcheng@163.com (Z.C.); liangyongbo001@gmail.com (Y.L.)

Received: 9 December 2019; Accepted: 4 January 2020; Published: 8 January 2020



Abstract: Electrocardiogram (ECG) signal evaluation is routinely used in clinics as a significant diagnostic method for detecting arrhythmia. However, it is very labor intensive to externally evaluate ECG signals, due to their small amplitude. Using automated detection and classification methods in the clinic can assist doctors in making accurate and expeditious diagnoses of diseases. In this study, we developed a classification method for arrhythmia based on the combination of a convolutional neural network and long short-term memory, which was then used to diagnose eight ECG signals, including a normal sinus rhythm. The ECG data of the experiment were derived from the MIT-BIH arrhythmia database. The experimental method mainly consisted of two parts. The input data of the model were two-dimensional grayscale images converted from one-dimensional signals, and detection and classification of the input data was carried out using the combined model. The advantage of this method is that it does not require performing feature extraction or noise filtering on the ECG signal. The experimental results showed that the implemented method demonstrated high classification performance in terms of accuracy, specificity, and sensitivity equal to 99.01%, 99.57%, and 97.67%, respectively. Our proposed model can assist doctors in accurately detecting arrhythmia during routine ECG screening.

Keywords: electrocardiogram; arrhythmia; automation; convolutional neural network; long short-term memory

1. Introduction

Electrocardiography provides abundant health and pathology information about the heart and is the main method of diagnosing heart disease [1]. Arrhythmia is an extremely common heart disease and is mainly diagnosed by doctors. However, misdiagnosis and missed diagnosis often occur in clinical practice due to differences in doctors' experiences and the randomness of arrhythmia events. At present, automatic detection and identification of arrhythmia events are urgently needed, as they can help doctors detect arrhythmia events earlier.

Traditionally, the study of arrhythmia diagnosis has mainly focused on the noise filtering of electrocardiogram (ECG) signals [2–4], signal segmentation [5–7], and manual feature extraction [8–11]. Osowski et al. [9] proposed a machine learning method that uses higher-order statistics (HOS) and Hermite functions to extract features, and a support vector machine (SVM) to classify heart diseases. De Chazal et al. [2] used morphological features and weighted linear discrete analysis (LDA)

combined with a packaging feature selection function to screen for heart disease. It is well known that the morphological approach is sensitive to ECG signal noise and has many limitations in the classification performance robustness of the model [12]. Thanks to the development of deep learning technology, many feature extraction processing tasks can be completed by convolutional computation. This method is superior to the morphological approach and has low requirements for signal quality in classification [13]. Kiranyaz et al. [14] introduced a one-dimensional convolutional neural network (1-D CNN) to identify and classify ventricular ectopic beats and premature ventricular contractions, and achieved good results. Yildirim et al. [15] proposed a deeper 1-D CNN classifier and was able to classify even more categories of heart disease and improve the classification performance. Although there are many references to ECG arrhythmia classification, there are still several limitations: (1) ECG signal information is lost during feature extraction or noise filtering, (2) ECG arrhythmia type has a limited number of classifications, and (3) the performance of the actual classification method is relatively poor.

Based on the abovementioned problems, a model based on the input of two-dimensional grayscale images is proposed in this paper, which combines a deep 2-D CNN with long short-term memory (LSTM). Some ECG signal information may be missed due to problems such as noise filtering, but this can be avoided by converting a one-dimensional ECG signal into a two-dimensional ECG image [16]. In most current studies, the data used are relatively limited. Many studies need to be very careful when preprocessing the one-dimensional ECG signals because the one-dimensional ECG signals are more sensitive and have a greater impact on the final accuracy. The conversion of one-dimensional ECG signals into two-dimensional ECG images can get more data and the data is effectively available. There is no need for very precise separation of individual beats when performing data conversion. Even if some adjacent signals are separated, the convolution layer of the model can ignore these small noise data. Using two-dimensional ECG images does not require noise filtering and manual feature extraction. Because the convolution and pooling layers of the model automatically ignore the noise data when acquiring the feature map, they avoid the problems of sensitivity to noise signals and accuracy being affected. Some researchers [17] tend to use images instead of one-dimensional signals as input data in other similar disease diagnosis studies. The use of two-dimensional ECG images for detection and classification is more like a way for cardiologists to diagnose arrhythmic diseases because the diseases are diagnosed and identified through the observation of the images. If one-dimensional ECG signals are applied to instruments such as ECG monitors, problems such as sampling rate and noise will inevitably occur, so two-dimensional ECG images can be further applied to ECG monitoring robots that can assist cardiac experts in diagnosing arrhythmic diseases. In addition, it is difficult to apply the data augmentation method used in previous studies due to the characteristics of the one-dimensional ECG signal. The ECG signal is augmented to enlarge the training data, which can effectively improve the classification accuracy. Therefore, in this study, we used different cropping methods to augment the two-dimensional ECG image, so as to help the 2-D CNN model train a single ECG image from different angles. The automatic extraction of ECG beats features using a 2-D CNN can solve the problem of current hand-designed waveform features that are not sufficiently robust to handle patient-to-patient differences in heart beats. In addition to the 2-D CNN model, there is another LSTM deep learning model, which is a time recurrent neural network (RNN). The status of each cell in the LSTM interacts with those of the others, and the time dynamics in the data are presented through the internal feedback state, which can avoid the problem of long-term dependence. The LSTM cells also have the capability of retaining and feeding back useful information of selectively stored information [18]. The combination of 2-D CNN and LSTM model features greatly improves the classification effect.

2. Materials and Methods

In this study, the datasets and annotations used were from the MIT-BIH arrhythmia database. The database included a total of 48 0.5 h long ECG signal records obtained from 47 subjects using two

leads [19]. Each signal record was sampled at 360 Hz with a set of beat markers presented at the R peak. These records were independently explained by multiple cardiologists. ECG signals were converted into ECG images as input data through data processing. In this paper, lead II signals of data were used in the experiments. Following the Association for the Advancement of Medical Instrumentation standard, according to the annotations provided by the MIT-BIH arrhythmia database, we selected “N” for normal sinus rhythm (NOR), “L” for left bundle branch block (LBBB), “R” for right bundle branch block (RBBB), “A” for atrial premature beat (APB), “V” for premature ventricular contraction (PVC), “/” for paced beat (PAB), “E” for ventricular escape beat (VEB), and “!” for ventricular flutter wave (VFW) for classification. Other types of arrhythmia were excluded in this paper, such as nodal escape beat, start of ventricular flutter, and other beats that cannot be classified. Those have been ignored by most ECG arrhythmia studies because these beats have relatively little research significance. The overall procedures are shown in Figure 1.

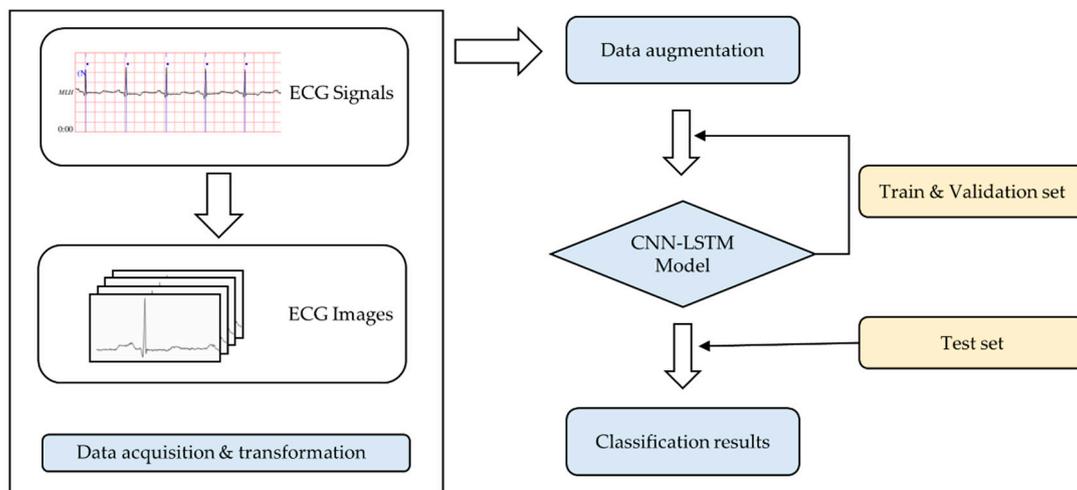


Figure 1. Overall procedures processed in ECG arrhythmia classification.

2.1. Data Preprocessing

In this study, the input data of the model were two-dimensional images. Most previous works have used one-dimensional ECG signals as the input data for the models, which then requires noise filtering and feature extraction of the data during the data preprocessing stage. Because of the time series characteristics of one-dimensional signals, some ECG signal information may be lost during the noise filtering and feature extraction process, which affects the integrity of the data and may also affect the accuracy of the final classification results. Therefore, in this paper, in the data preprocessing stage, we converted one-dimensional ECG signals into two-dimensional ECG images as classification data, which can ensure the integrity of the original ECG data to the greatest extent. We converted each ECG signal into a separate 192×128 grayscale image. From the ECG signals obtained from the database, the peak value of the R wave was used as a criterion for dividing each ECG beat according to the existing R wave peak markers in the database in order to locate each ECG signal. Then, 92 data points before and after the R wave peaks of the two ECG signals before and after were deleted, and then a single ECG image was cropped. This was accomplished using Equation (1):

$$T(\text{Rpeak}(n) - 92) \leq T(n) \leq T(\text{Rpeak}(n) + 92). \quad (1)$$

Finally, a total of 107,620 ECG image data points were obtained after conversion, and the categories were labeled respectively. From the transformation results, it can be seen that the amount of data was significantly improved through transformation, which also provided more data for subsequent model learning and training. Table 1 describes the information recorded by all ECG signals.

Table 1. A summary table of ECG signal description from the MIT-BIH arrhythmia database.

Type	Records	Beats
NOR	100,101,103,106,108,112,113,114,115,117,119,121,122,123, 203,205,219,230,234	75,016
PVC	105,116,200,201,202,208,210,213,215,221,228,233	7130
PAB	102,104,107,217	7024
APB	209,220,222,223,232	2544
LBBB	109,111,207,213	8072
RBBB	118,124,212,231	7256
VEB	207	106
VFW	207	472
Total		107,620

2.2. Data Augmentation

Because the database mostly contains the number of normal rhythm types, there is an imbalance in the amount of data obtained for each disease type. Due to the problem of unbalanced data volume in each category of data, data augmentation can increase the amount of data in a class with a small volume of data and effectively reduce the occurrence of overfitting problems [20]. Image enhancement can increase the amount of data. Most previous ECG arrhythmia studies were not able to manually add augmentation data to the training set due to the possibility of ECG signals being lost. The reason is that feedforward neural network (FFNN) [21] and SVM [22] classifiers assume that every ECG signal possess the same classification worth. In most studies with a large amount of data, the ECG signal segmentation method is used to divide a one-dimensional ECG signal into multiple ECG signal segments to expand the amount of data. However, since the input data of the model in this study were ECG images, the method of image enhancement would not modify the data, but it would increase the amount of data. This method draws on the idea of image processing and performs data enhancement on the converted two-dimensional ECG image. On the basis of the converted original ECG image, processing is performed in a certain manner, which increases the number of data samples, and at the same time, leaves the label value of the data unchanged. It can maximize the original qualities of the data while optimizing the data imbalance in the research. In this study, nine different clipping methods were used to increase the beat of the other seven ECG arrhythmia types, except the NOR class. Image cropping was performed on a specified area of the target image. The cropping method of the left top image is one example. The reference coordinates of the left top image were (0, 0). According to the cropping rule of 96 sizes, (0, 96), (96, 0), and (96, 96) coordinate points were used as the four vertex coordinates of the left top image. This method was used for image cropping and obtained a 96×96 left top image of the target image. The other eight images were cropped similarly. Among the other eight images, the reference coordinates of the center top image were (64, 0), the reference coordinates of the right top image were (96, 0), the reference coordinates of the left center image were (0, 16), and the reference coordinates of the center image were (64, 16), the reference coordinates of the right center image were (96, 16), the reference coordinates of the left bottom image were (0, 32), the reference coordinates of the center bottom image were (64, 32), and the reference coordinates of the right bottom image were (96, 32). By using this cropping method, all the augmentation images could be obtained. Finally, the entire enhanced image was adjusted to a size of 192×128 to ensure the uniformity of all sample data. This greatly increased the amount of data for a relatively small number of arrhythmia categories. The added image also retained the information contained in the original ECG image, which is of equal reference value. The data augmentation method was produced inside the model, which reduced the time spent between images in memory, thereby enhancing the learning speed of the model. The experimental data used in subsequent experiments in this paper were divided into 60%, 20%, and 20% of the training, validation, and test sets, respectively. All experimental data were randomly shuffled. According to different proportions, the disrupted experimental data were

randomly divided into different sets. There are 107,620 two-dimensional ECG image data in this paper. Among them, 64,572 data were divided into the training set. A total of 581,148 two-dimensional ECG image data were used for model training after data enhancement. The original PAB image and the nine cropped grayscale images are shown in Figure 2.

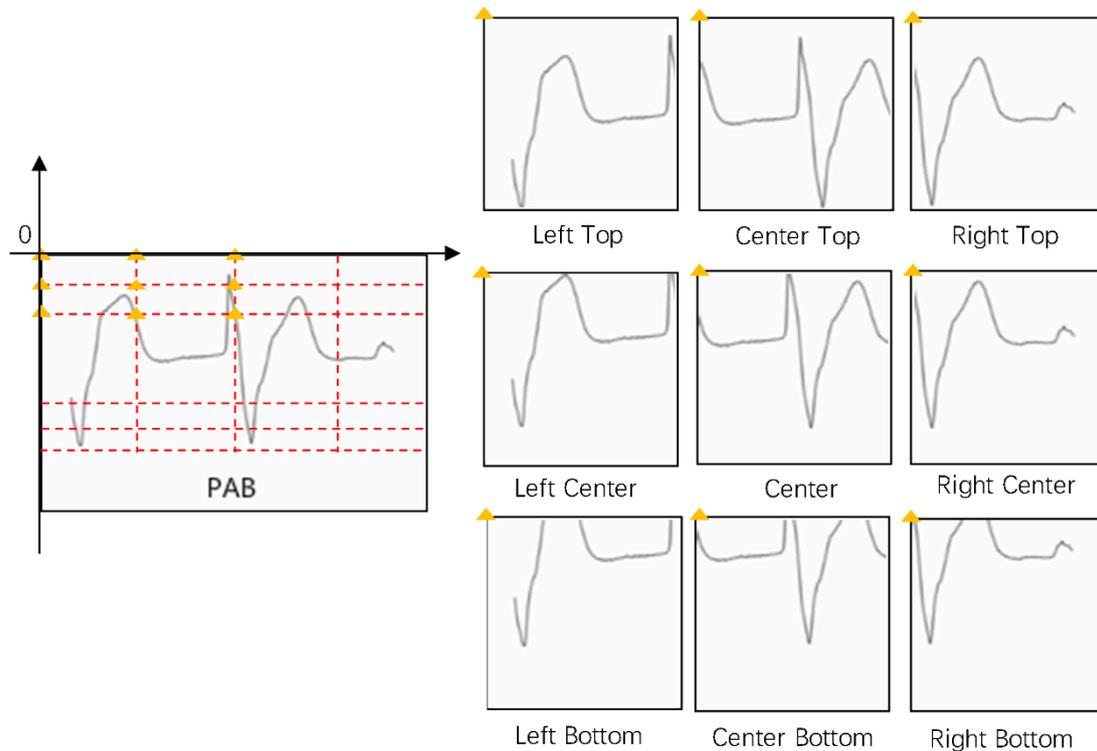


Figure 2. Original paced beat (PAB) image and nine cropped images.

2.3. CNN-LSTM Model

Deep learning [23–25] is a new technology that has become mainstream in the field of machine learning and pattern recognition. In this study, a new method for automatically detecting eight different types of ECG signal arrhythmias was developed. It uses a cross-learning model based on deep learning. The overall structure of the model is implemented by combining CNN and LSTM. Among them, CNN is suitable for processing spatial or locally related data, while LSTM is good at capturing the characteristics of data related to time series.

Layers 1–9 of the model are convolutional layers coupled to the largest collection layer, and layer 10 is the LSTM layer. The end of the network uses a fully connected layer for predicting the output. The spatial feature map can be well extracted by the convolutional layer. Subsequent LSTM layers help the model capture the temporal dynamics that exist in these signatures [26]. In the combination of CNN and LSTM, the output shape after the pooling layer of the model is (none, 16, 16, 256). We reshape the dimensions of the model through the reshape method, and the input size of the LSTM layer after reshaping is (256, 256). After analyzing the time characteristics of LSTM, the model finally sorts ECG signals through a fully connected layer. The training stages of the model can be improved by setting the optimizer and learning rate. So, we set and used a learning rate of 0.001 and the Adam optimizer for optimization. Figure 3 shows the proposed network model. A detailed overview of the structure is given in Table 2.

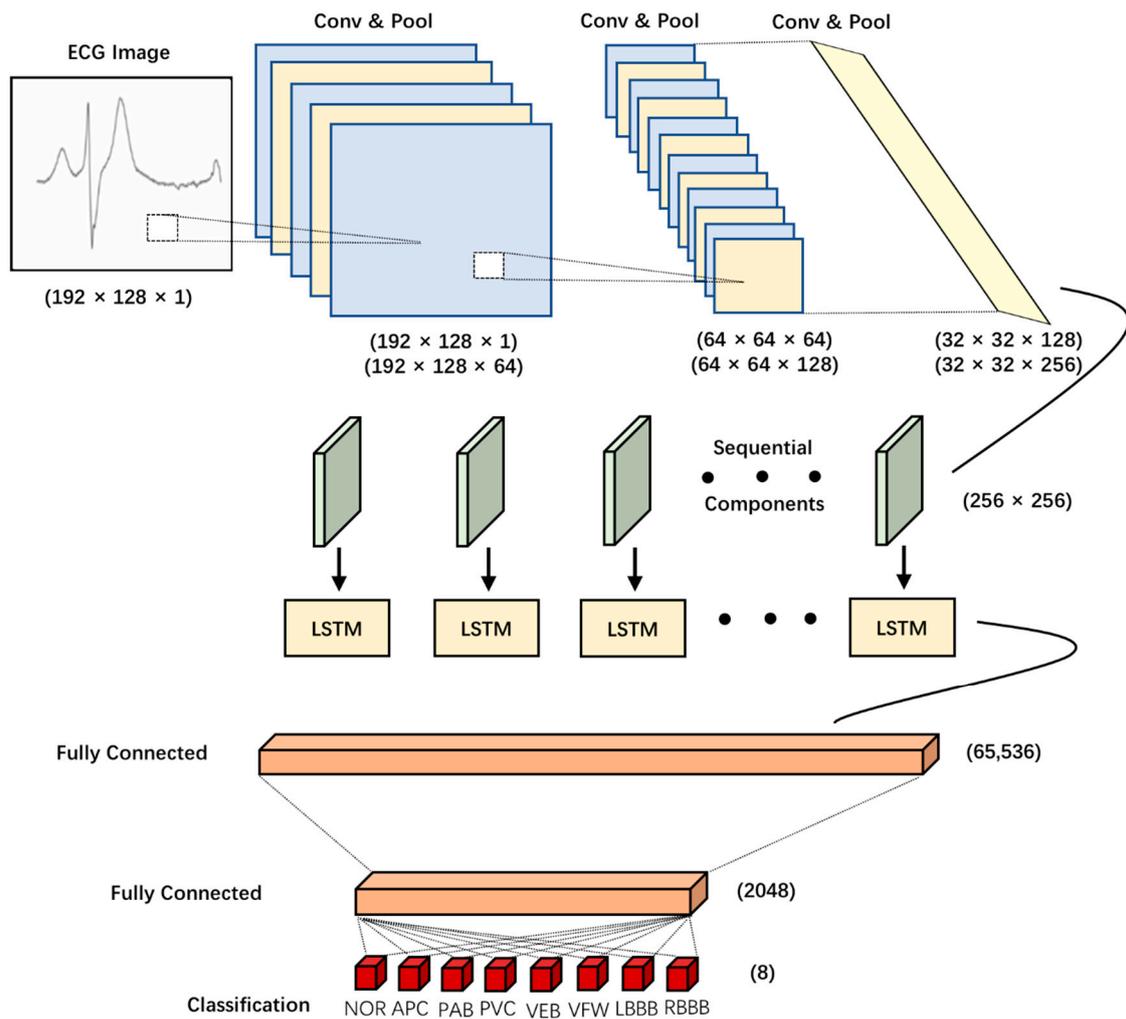


Figure 3. An illustration of the proposed CNN-LSTM architecture.

Table 2. Detailed overview of the proposed CNN-LSTM model.

Layer	Type	Kernel Size	Stride	Kernel	Input Size
1	Conv2D	3 × 3	1	64	192 × 128 × 1
2	Conv2D	3 × 3	1	64	192 × 128 × 64
3	Pool	2 × 2	2	-	192 × 128 × 64
4	Conv2D	3 × 3	1	128	64 × 64 × 64
5	Conv2D	3 × 3	1	128	64 × 64 × 128
6	Pool	2 × 2	2	-	64 × 64 × 128
7	Conv2D	3 × 3	1	256	32 × 32 × 128
8	Conv2D	3 × 3	1	256	32 × 32 × 256
9	Pool	2 × 2	2	-	32 × 32 × 256
10	LSTM	-	-	-	256 × 256
11	Fully-connected	-	-	2048	65,536
12	Fully-connected	-	-	2048	2048
13	Out	-	-	8	2048

2.4. VGGNet Model

Many pretrained models, such as VGGNet [27], GoogleNet, and so forth, could provide us with many solutions to the problem. In this study, we compared the proposed model with the well-known VGGNet model and other ECG arrhythmia classification studies. The VGGNet model is a deep convolutional neural network model composed of multiple convolution blocks. The model can extract

ECG deep features well, through convolution and pooling layers. It generates feature maps from the extracted features for learning and training. In the VGGNet model, we set and used a learning rate of 0.001 and the Adam optimizer for optimization too. A detailed overview of the structure is given in Table 3.

Table 3. Detailed overview of the VGGNet [27] model.

Layer	Type	Kernel Size	Stride	Kernel	Input Size
1	Conv2D	3×3	1	64	$192 \times 128 \times 3$
2	Conv2D	3×3	1	64	$192 \times 128 \times 64$
3	Pool	2×2	2	-	$192 \times 128 \times 64$
4	Conv2D	3×3	1	128	$64 \times 64 \times 64$
5	Conv2D	3×3	1	128	$64 \times 64 \times 128$
6	Pool	2×2	2	-	$64 \times 64 \times 128$
7	Conv2D	3×3	1	256	$32 \times 32 \times 128$
8	Conv2D	3×3	1	256	$32 \times 32 \times 256$
9	Conv2D	3×3	1	256	$32 \times 32 \times 256$
10	Pool	2×2	2	-	$32 \times 32 \times 256$
11	Conv2D	3×3	1	512	$16 \times 16 \times 256$
12	Conv2D	3×3	1	512	$16 \times 16 \times 512$
13	Conv2D	3×3	1	512	$16 \times 16 \times 512$
14	Pool	2×2	2	-	$8 \times 8 \times 512$
15	Fully-connected	-	-	4096	$8 \times 8 \times 512$
16	Fully-connected	-	-	4096	4096
17	Out	-	-	8	4096

2.5. Model Architecture and Details

An earlier part of the proposed model is a 2-D CNN structure, which is a combination of three convolution blocks with a step size of 1. There are two 2-D CNN layers and one maximum pooling layer consisting of each convolution block; it is activated using the exponential linear units (ELU) activation function. The batch normalization layer is used to batch normalize the activation output of the layer. In all convolution operations, by multiplying the superposition matrix, the convolution kernel is continuously extracted for each convolution feature. After two-dimensional convolution, the feature map of this layer uses a maximum pooled filter for feature extraction, and the step size of the filter is two. The feature map is propagated to the two-dimensional maximum pooling layer, and the maximum value of the specified area in the feature map is extracted and labeled to extract a new feature map. This continuously deepens the model network. The size of the feature map of each layer is gradually reduced to speed up the learning rate of the model structure.

Then, the feature map is passed to the LSTM layer in the latter part of the model to extract time information. The extracted features are sorted into sequential components after convolution and merging, and their time series prediction is performed by the LSTM circular chain structure. LSTM is different from the traditional RNN because it has a different structure to a single neural network. It consists of multiple cell states and gated modules. LSTM repeatedly combines these units to ensure that all information is cyclically learned throughout the network while remaining unchanged and persistent. The modules of this structure interact to resolve the disappearance of the gradient and avoid long-term dependence problems. After the LSTM layer, it is fed to the fully connected layer of the softmax layer with eight output neurons by a feature vector with representation and time-dependent features. Finally, arrhythmia prediction is performed by the outputs of the eight categories fed to the fully connected layer.

2.5.1. Activation Function

Activation functions are necessary to improve the approximation ability between each layer of the network to enhance the expressiveness of neural networks. Referring to other current related research,

nonlinear activation functions, including leakage rectified linear units (LReLU), ELU, and rectified linear units (ReLU), are widely used in CNN models. Most researchers use ReLU as the activation function of the model, but after analyzing the experimental results, when the input function gradient is too large, the neuron will lose the activation function after the network parameters are updated [28]. The ELU activation function was used in the experiments in this study, as it demonstrated better classification of ECG arrhythmia. ELU is shown in Equation (2):

$$\text{ELU}(x) = \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases} \quad (2)$$

2.5.2. Batch Normalization

In deep learning, with the deepening of the number of layers, the parameters of the layer in question are slightly changed, and the proportion of the input parameters of the latter layer have a more comprehensive impact. This phenomenon is called the internal covariate offset. To accelerate the convergence of the model during training and avoid the gradient expansion of the model, we added a batch normalization layer to the network model. In this way, normalizing the batch after each feature change in the network structure ensures that the conversion of different batches is kept within a certain range, thereby accelerating the convergence of the parameters [29]. Batch-normalized locations are typically applied before the activation function and after the convolutional layer. In the experiments in this study, the ELU function was placed before the batch normalization layer and achieved significant results. Therefore, there was an ELU function before the batch normalization layer in each convolution block. Behind each convolution block, there was a two-dimensional maximum pooling layer. The specific formula for batch normalization was calculated as

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i, \quad (3)$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m x_i - \mu, \quad (4)$$

$$x^{(i)} = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}}, \quad (5)$$

where $x^{(i)}$ is the standardized output; μ and σ represent the mean and variance of the same batch, respectively; and ε is a constant, with the value 0.001.

2.5.3. Dropout Regularization

Overfitting is a very important problem encountered during model training [30]. Therefore, to avoid overfitting problems, dropout regularization was used here to avoid overfitting of the model training. At the same time, we also conducted comparison experiments with models that did not use dropout regularization. Dropout regularization probabilistically discards some of the nodes in the same layer to reduce the dependencies between layers. The connection weight will be excluded when the neuron exits, which greatly improves the generalization capacity of the model. A model without dropout regularization adds all of the weights to the learning process during the training process, so the dependency between each layer of the model is greatly increased, which causes overfitting problems. In experiments using dropout regularization, it was placed before the last fully connected layer of the model. The rate of dropout was 0.5.

3. Results

The experimental data in this study came from the international standard ECG database MIT-BIH, which has accurate and comprehensive expert annotation and is widely used in current ECG research [19]. In the experiment, the experimental data were divided into 60%, 20%, and 20%

for the training, validation, and test sets, respectively. Among them, 21,524 data were used for testing. The number of epochs for training was 100. In each epoch, the batch size used for the dataset was 32, and it was extended over all input data. Two-dimensional ECG images were cropped to 96×96 ECG grayscale images as required. Finally, the enhanced image was adjusted to a size of 192×128 . All experiments were based on the deep learning framework Tensorflow. The working environment for training the network consisted of two NVIDIA Geforce RTX 2080 Ti GPUs with 64 GB of RAM. The entire training process took 16 h.

We compared two different experimental schemes and conducted experimental verification based on the presence or absence of dropout regularization. In Experiment A, we did not use dropout regularization, and the weights of the model during training were all involved in the learning process. In Experiment B, we added dropout regularization with a dropout rate of 0.5. That way, 50% of the information was discarded during training and 50% of the information was retained for learning. The comparison of the results of the two experimental schemes is shown in Figure 4. From the experimental results, we can see that the network after using dropout regularization always had a very stable state, and the accuracy rate gradually increased under the stable state, finally reaching the highest point. The network that did not use dropout regularization appeared to overfit, gradually stabilized after about 60 epochs, and showed very high accuracy.

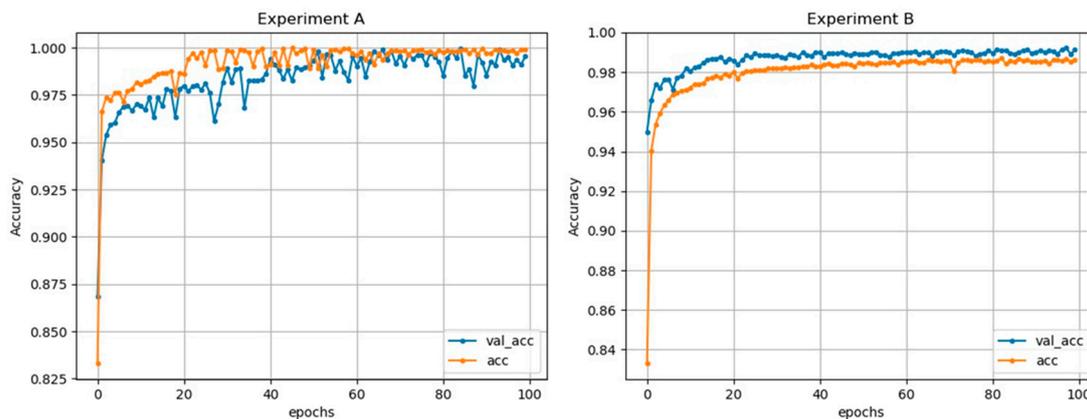


Figure 4. Accuracies of the two training models. Experiment A without dropout regularization. Experiment B with dropout regularization.

The accuracy and loss curves for training and verification are shown in Figure 5. Both the training and verification curves of the model increased in a stable state and stabilized at approximately 100 epochs. The classification evaluation of the model used the following evaluation metrics: accuracy (Acc), specificity (Spec), and sensitivity (Sen). The model combining CNN and LSTM achieved 99.01% accuracy, 97.67% sensitivity, and 99.57% specificity after experimental verification. The sensitivity indicates the ratio of normal ECG data detected by the system to the overall normal data. Specificity indicates the proportion of abnormal ECG data to total abnormal data. The accuracy rate represents the proportion of the data that determines the overall correctness of the data. The three metrics (Acc, Spec, and Sen) are defined as follows:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \times 100\%, \quad (6)$$

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%, \quad (7)$$

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \quad (8)$$

where TP indicates that normal ECG data are classified into normal categories; TN means classifying outlying data into exceptional categories (both TP and TN indicate accurate classification); FP indicates that abnormal ECG data are classified into normal categories; and FN means classifying normal data into exceptional categories (both FP and FN indicate a classification error). The three metrics can reflect the overall classification ability of the system as a whole. The larger the value, the better the classification effect. We also compared the evaluation indicators obtained from the two experiments with and without the dropout regularization model. The comparison results of the two experimental schemes are shown in Table 4. The model without dropout regularization showed high classification results due to overfitting, and obtained 99.87% Acc, 99.78% Spec, and 98.95% Sen. They were all higher than the experimental model using dropout regularization.

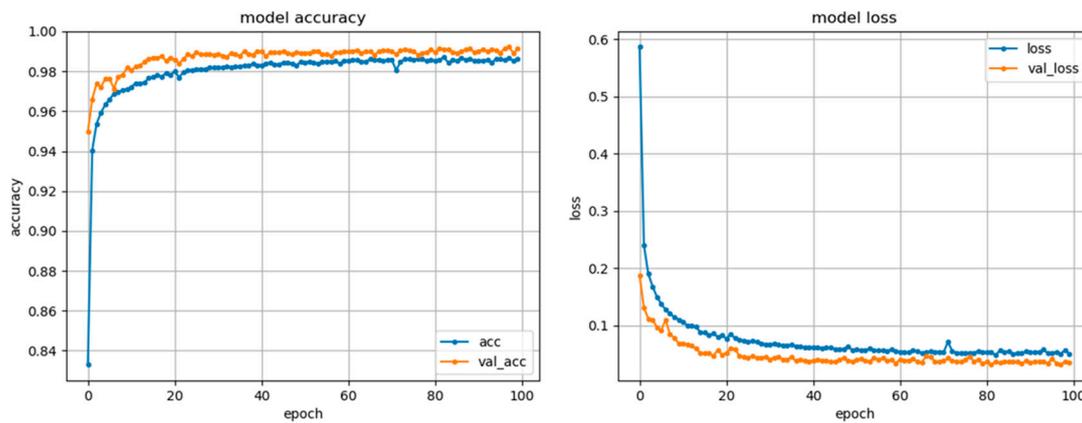


Figure 5. Accuracy and loss of CNN-LSTM training model.

Table 4. Average classification performances of the two experiments.

Experiment	Scheme	Acc	Spec	Sen
Experiment A	Without Dropout	99.87%	99.78%	98.95%
Experiment B	With Dropout	99.01%	99.57%	97.67%

Table 5 describes the confusion matrix for the training model classification results. It can be seen that the model performed better on the classification of PAB, LBBB, and VEB types, and the performance of the classification of APB types was average. This may have been caused by the small morphological differences of the waveforms during the learning process.

Table 5. Confusion matrix of the proposed CNN-LSTM model.

Predicted	NOR	PAB	APB	PVC	LBBB	RBBB	VEB	VFW
NOR	14,940	1	10	50	0	1	2	0
PAB	2	1404	0	0	0	0	0	0
APB	78	0	420	6	3	2	0	0
PVC	5	0	0	1420	0	0	0	1
LBBB	8	0	1	2	1602	0	1	1
RBBB	20	2	1	4	2	1422	1	0
VEB	1	0	0	0	0	0	21	0
VFW	8	0	0	0	1	0	0	86

Comparing the experiments with the same dataset, the results of 98.67% accuracy, 96.93% sensitivity, and 99.52% specificity were obtained by using the VGGNet model. The accuracy and loss curves of the VGGNet model for training and verification are shown in Figure 6. It can be seen from Figure 6 that the training accuracy and loss rate of the VGGNet model tend to stabilize after 20 epochs. The entire training process of VGGNet model took 27 h. Although the parameters of the internal

convolution layer are reduced in the VGGNet model, the actual internal parameter space is relatively large. Among them, most of the parameters come from the first fully connected layer, which consumes more computing resources. Therefore, it always takes longer training VGGNet models. Table 6 describes the confusion matrix for the VGGNet training model classification results. It can be seen that the performance of CNN-LSTM model in predicting PVC and RBBB types is better than VGGNet model by observing and comparing Tables 5 and 6. In the CNN-LSTM model, 2.1% of the subdivided categories were incorrectly classified into other categories, while in the VGGNet model, 3.5% of the subdivided categories were incorrectly classified into other categories. It can be seen that both the models performed better in the classification of PAB, LBBB, and VEB types. The comparison results of the two models are shown in Table 7. The two models differed in their numbers of convolutional and pooling layers and whether or not the LSTM layer was used. It can be seen from the results (Table 5) that the proposed model performed better than VGGNet.

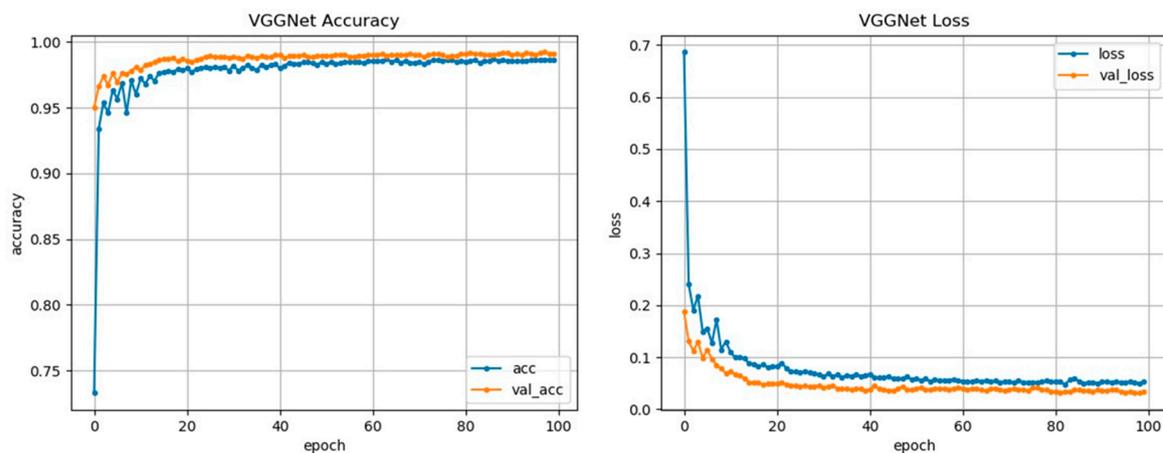


Figure 6. Accuracy and loss of VGGNet training model.

Table 6. Confusion matrix of the VGGNet model.

Predicted	NOR	PAB	APB	PVC	LBBB	RBBB	VEB	VFW
NOR	14,932	10	5	41	8	5	3	0
PAB	1	1402	1	1	0	0	0	0
APB	68	0	407	30	0	3	0	1
PVC	29	2	6	1387	0	0	0	2
LBBB	9	0	1	2	1603	0	0	0
RBBB	30	10	0	10	1	1401	0	0
VEB	1	0	0	0	0	0	21	0
VFW	6	0	0	2	0	0	0	87

Table 7. Comparison of the proposed model with VGGNet.

Model	Dataset	Method	Acc	Sen	Spec
Proposed Model	This study’s dataset	CNN-LSTM	99.01%	97.67%	99.57%
VGGNet [27]	This study’s dataset	CNN	98.67%	96.93%	99.52%

4. Discussion

With the continuous development of machine learning in recent years, the MIT-BIH arrhythmia database has been used by an increasing number of researchers in ECG research. Table 8 summarizes the study of the automatic detection of ECG arrhythmias. Compared with other related studies, the method of combining 2-D CNN and LSTM proposed in this paper was highly accurate. In most machine learning methods, there are often adaptability problems. Through experimental verification, we were able to provide a deeper comparison of the use of dropout regularization in the model.

Without dropout regularization, the training of a model is prone to overfitting, which seriously affects a model's classification ability. After using a 50% forgetting probability after the final batch normalization of the fully connected layer, good classification performance was obtained, which also greatly improved the generalization effect of the model. Most classification work requires noise removal and manual extraction of ECG signals, which inevitably leads to partial beat loss of ECG data. At the same time, most studies have limited data volume because of the different ECG signal segmentation methods. In this study, after converting one-dimensional ECG signals into two-dimensional ECG image data, we were able to avoid losing part of the data due to preprocessing problems. Moreover, data augmentation methods can also lead to an increase in the amount of data in relatively small categories. That further balances the different types of data and improves the classification performance of the model. It can be seen from Table 8 that the number of arrhythmia classifications obtained in each study differed, and the amount of data used varied. Osowski et al. [9] preprocessed the data by the HOS cumulant and Hermite coefficient of the QRS complex in ECG signals and combined the method of minimum mean square error with SVM, which obtained 98.71% accuracy. Martis et al. [31] also used HOS to preprocess the signals. They used 34,989 ECG signal data points and a least-squares SVM to classify the five arrhythmia types. The highest average was obtained, and the accuracy rate was 93.48%. Plawiak et al. [32] augmented the characteristics of ECG signals by spectral power density. He used ECG signal data to compare different machine learning models and finally used the support vector machine model to obtain the best classification of 17 arrhythmia diseases with 98.85% accuracy. Guerra et al. [33] also used SVM for classification, but they did not use a single specific SVM, instead, multiple SVMs, to achieve automatic classification. Their classification accuracy reached 94.50%. Summarizing the related research mentioned above, the research methods used are all traditional machine learning methods. In data processing, ECG signals need to be filtered and feature extracted by means such as HOS. At the same time, the use of models is also a form of traditional classification for machine learning.

Table 8. Correlational studies of ECG arrhythmia.

Author	Features Set	Classes	ECG Beats	Classifier	Accuracy
Machine Learning Methods					
Osowski et al. (2008) [9]	Higher-order statistics (HOS) cumulant and Hermite coefficient of QRS complex	13	12,785	Support Vector Machine	98.71%
Martis et al. (2013) [31]	Bispectrum and principal component analysis (PCA)	5	34,989	Least-Squares Support Vector Machine	93.48%
Plawiak et al. (2018) [32]	The spectral power density and genetic optimization of parameters	17	1000	Support Vector Machine	98.85%
Guerra et al. (2019) [33]	HOS and local binary patterns (LBPs)	4	49,691	Combination of Multiple Support Vector Machines	94.50%
Deep Learning Methods					
Kiranyaz et al. (2016) [14]	End-to-end	5	83,648	CNN	99.00%
Jun et al. (2018) [16]	End-to-end	8	106,501	CNN	99.05%
Shu Lih Oh et al. (2018) [18]	End-to-end	5	16,499	CNN + LSTM	98.10%
Acharya et al. (2017) [34]	End-to-end	5	109,449	CNN	94.03%
Yildirim et al. (2018) [35]	End-to-end	5	7326	Wavelet + Bi-LSTM	99.25%
This Study	End-to-end	8	107,620	CNN + LSTM	99.01%

In recent years, deep learning has also developed rapidly. Compared with machine learning, the results of deep learning are more significant. Deep learning models such as CNN and LSTM are used in the study of ECG arrhythmia classification by more and more researchers. Acharya et al. [34] constructed a nine-layer 1-D CNN model to automatically identify five different categories of heartbeats in ECG signals. The input data of the model were one-dimensional ECG signals. They filtered

the high-frequency noise of the signals and then detected and classified the noisy and non-noisy ECG signals through the model, which greatly improved the generalization ability of the model. The accuracy of the model for classifying original ECG signals was 94.03%. However, the ECG signals used for classification had a high degree of imbalance, and the classification accuracy of the data also decreased after noise filtering. Shu et al. [18] proposed a diagnostic model that combines 1-D CNN and LSTM. Input data of the model were also one-dimensional ECG signals. In the data processing stage, the ECG data were segmented into many ECG data segments of different lengths by positioning the waveforms, and then all ECG data segments were standardized to a uniform length. The model was able to classify ECG signals of different lengths into five categories and achieved an accuracy of 98.10%. Jun et al. [16] proposed a 2-D CNN model. Input data of the model were two-dimensional ECG data. The model used multiple convolution processing units to extract ECG deep features, and classified the extracted features. The proposed model achieved an accuracy of 99.05%. Although a single 2-D CNN model can learn the spatial characteristics of ECG data very well, the learning efficiency of the model is not high enough, and the convergence speed of model training accuracy is low. An LSTM layer is added after the 2-D CNN to learn the time series related features of the components decomposed into a convolutional feature sequence. That way, the temporal characteristics of the data can be better analyzed and further classified. Such training can improve the efficiency of the model, and at the same time, get a higher classification accuracy. Yildirim et al. [35] proposed a bidirectional LSTM (Bi-LSTM) model with wavelet sequences to analyze and classify ECG signal sequences in time series. Bi-LSTM adds more available information, including historical and new data, through a two-way network propagation, which can make the information of the data more fully used. In addition, the ECG data needed to be segmented at different scales to obtain 7326 ECG data segments, which were then used as data for the model. In the end, the proposed model achieved an accuracy of 99.25%.

According to the summary of the abovementioned machine learning and deep learning methods, the CNN-LSTM model proposed here demonstrated higher classification accuracy than other related studies and also showed better advantages in data processing. The quality of the data used often has a great impact on the final results of the model, so using ECG images for classification is also a novel idea. Therefore, the proposed model can be applied in clinics to help cardiologists objectively diagnose ECG heartbeat signals, or it can be used in new smart monitor applications.

5. Conclusions

Detection and identification of arrhythmias is an integral part of the early diagnosis of cardiovascular disease. This paper presented an effective arrhythmia classification method that combines 2-D CNN and LSTM and uses ECG images as the input data for the model. One-dimensional signals obtained from the MIT-BIH arrhythmia database were converted into 192×128 grayscale images. A total of 107,620 ECG images were obtained by processing the data acquired from the database. As a result, the accuracy of this method was 99.01%, the specificity was 99.57%, and the sensitivity was 97.67%. The classification results of ECG arrhythmia showed that the method of arrhythmia detection using a combination of ECG image data and CNN-LSTM can be useful for helping doctors better diagnose cardiovascular disease and can considerably reduce the workloads of doctors. In the future, this auxiliary diagnostic method could be used in connection with medical robots or medical monitors for diagnostic treatment.

Author Contributions: Conceptualization, Z.Z., Y.L., and Z.C.; methodology, Z.Z. and Y.L.; software, Z.Z.; validation, Z.Z.; formal analysis, Z.Z.; investigation, Z.Z.; resources, Z.Z., Y.L., and Z.C.; data curation, Z.Z.; writing—original draft preparation, Z.Z.; writing—review and editing, Z.Z., Y.L., and Z.C.; visualization, Z.Z.; supervision, Y.L. and Z.C.; project administration, F.H., J.Z., and Q.T.; funding acquisition, Y.L. and Z.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (grant numbers 61627807 and 81873913), the Natural Science Foundation of Guangxi (2017GXNSFGA198005), and the Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (YQ19112).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ye, C.; Coimbra, M.T.; Kumar, B.V. Arrhythmia detection and classification using morphological and dynamic features of ECG signals. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine Biology Society, Buenos Aires, Argentina, 31 August–4 September 2010.
2. De Chazal, P.; O'Dwyer, M.; Reilly, R.B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1196–1206. [[CrossRef](#)] [[PubMed](#)]
3. Mar, T.; Zaunseder, S.; Martinez, J.P.; Llamedo, M.; Poll, R. Optimization of ECG classification by means of feature selection. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 2168–2177. [[CrossRef](#)] [[PubMed](#)]
4. Shadmand, S.; Mashoufi, B. A new personalized ECG signal classification algorithm using Block-based Neural Network and Particle Swarm Optimization. *Biomed. Signal Process. Control* **2016**, *25*, 12–23. [[CrossRef](#)]
5. Hu, Y.H.; Palreddy, S.; Tompkins, W.J. A Patient-adaptable ECG Beat Classifier Using a Mixture of Experts Approach—Biomedical Engineering. *IEEE Trans. Biomed. Eng.* **1997**, *44*, 891–900.
6. El-Saadawy, H.; Tantawi, M.; Shedeed, H.A.; Tolba, M.F. Electrocardiogram (ECG) Classification Based On Dynamic Beats Segmentation. In Proceedings of the 10th International Conference on Informatics and Systems, Giza, Egypt, 9–11 May 2016.
7. He, H.; Tan, Y.; Xing, J. Unsupervised classification of 12-lead ECG signals using wavelet tensor decomposition and two-dimensional Gaussian spectral clustering. *Knowl. Based Syst.* **2019**, *163*, 392–403. [[CrossRef](#)]
8. Ye, C.; Kumar, B.V.; Coimbra, M.T. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2930–2941.
9. Osowski, S.; Hoai, L.T.; Markiewicz, T. Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 582–589. [[CrossRef](#)]
10. Rodriguez, J.; Goni, A.; Illarramendi, A. Real-Time Classification of ECGs on a PDA. *IEEE Trans. Inf. Technol. Biomed.* **2005**, *9*, 23–34. [[CrossRef](#)]
11. Singh, R.; Mehta, R.; Rajpal, N. Efficient wavelet families for ECG classification using neural classifiers. *Procedia Comput. Sci.* **2018**, *132*, 11–21. [[CrossRef](#)]
12. D'Aloia, M.; Longo, A.; Rizzi, M. Noisy ECG Signal Analysis for Automatic Peak Detection. *Information* **2019**, *10*, 35. [[CrossRef](#)]
13. Pourbabaee, B.; Roshtkhari, M.J.; Khorasani, K. Deep Convolutional Neural Networks and Learning ECG Features for Screening Paroxysmal Atrial Fibrillation Patients. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *48*, 2095–2104. [[CrossRef](#)]
14. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 664–675. [[CrossRef](#)] [[PubMed](#)]
15. Yildirim, O.; Plawiak, P.; Tan, R.S.; Acharya, U.R. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput. Biol. Med.* **2018**, *102*, 411–420. [[CrossRef](#)] [[PubMed](#)]
16. Jun, T.J.; Nguyen, H.M.; Kang, D.; Kim, D.; Kim, D.; Kim, Y.-H. ECG arrhythmia classification using a 2-D convolutional neural network. *arXiv* **2018**, arXiv:1804.06812.
17. Yildirim, O.; Talo, M.; Ay, B.; Baloglu, U.B.; Aydin, G.; Acharya, U.R. Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals. *Comput. Biol. Med.* **2019**, *113*, 103387. [[CrossRef](#)] [[PubMed](#)]
18. Oh, S.L.; Ng, E.Y.K.; Tan, R.S.; Acharya, U.R. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput. Biol. Med.* **2018**, *102*, 278–287. [[CrossRef](#)] [[PubMed](#)]
19. Moody, G.B.; Mark, R.G. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50. [[CrossRef](#)] [[PubMed](#)]
20. Bakkouri, I.; Afdel, K. Multi-scale CNN based on region proposals for efficient breast abnormality recognition. *Multimed. Tools Appl.* **2018**, *78*, 12939–12960. [[CrossRef](#)]
21. Jiménez-Serrano, S.; Yagüe-Mayans, J.; Simarro-Mondéjar, E.; Calvo, C.J.; Castells, F.; Millet, J. Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features. In Proceedings of the 2017 Computing in Cardiology Conference (CinC), Rennes, France, 24–27 September 2017.
22. Giorgio, A.; Rizzi, M.; Guaragnella, C. Efficient Detection of Ventricular Late Potentials on ECG Signals Based on Wavelet Denoising and SVM Classification. *Information* **2019**, *10*, 328. [[CrossRef](#)]

23. Zeiler, M.D.; Fergus, R. Visualizing and Understanding convolutional networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014.
24. Taigman, Y.; Yang, M.; Ranzato, M.A.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
25. Deng, L.; Abdel-Hamid, O.; Yu, D. A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013.
26. Khalifa, M.; Shaalan, K. Character convolutions for Arabic Named Entity Recognition with Long Short-Term Memory Networks. *Comput. Speech Lang.* **2019**, *58*, 335–346. [[CrossRef](#)]
27. Yao, Q.; Wang, R.; Fan, X.; Liu, J.; Li, Y. Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network. *Inf. Fusion* **2020**, *53*, 174–182. [[CrossRef](#)]
28. Urtnasan, E.; Park, J.U.; Joo, E.Y.; Lee, K.J. Automated Detection of Obstructive Sleep Apnea Events from a Single-Lead Electrocardiogram Using a Convolutional Neural Network. *J. Med. Syst.* **2018**, *42*, 104. [[CrossRef](#)] [[PubMed](#)]
29. Yang, S.; Hao, K.; Ding, Y.; Liu, J. Vehicle Driving Direction Control Based on Compressed Network. *Int. J. Pattern Recognit. Artif. Intell.* **2018**, *32*, 27. [[CrossRef](#)]
30. He, W.; Wang, G.; Hu, J.; Li, C.; Guo, B.; Li, F. Simultaneous Human Health Monitoring and Time-Frequency Sparse Representation Using EEG and ECG Signals. *IEEE Access* **2019**, *7*, 85985–85994. [[CrossRef](#)]
31. Martis, R.J.; Acharya, U.R.; Mandana, K.M.; Ray, A.K.; Chakraborty, C. Cardiac decision making using higher order spectra. *Biomed. Signal Process. Control* **2013**, *8*, 193–203. [[CrossRef](#)]
32. Pławiak, P. Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. *Expert Syst. Appl.* **2018**, *92*, 334–349. [[CrossRef](#)]
33. Mondéjar-Guerra, V.; Novo, J.; Rouco, J.; Penedo, M.G.; Ortega, M. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. *Biomed. Signal Process. Control* **2019**, *47*, 41–48. [[CrossRef](#)]
34. Acharya, U.R.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adam, M.; Gertych, A.; Tan, R.S. A deep convolutional neural network model to classify heartbeats. *Comput. Biol. Med.* **2017**, *89*, 389–396. [[CrossRef](#)]
35. Yildirim, O. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput. Biol. Med.* **2018**, *96*, 189–202. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).