

## Article

# Res-BiANet: A Hybrid Deep Learning Model for Arrhythmia Detection Based on PPG Signal

Yankun Wu <sup>1</sup>, Qunfeng Tang <sup>1</sup> , Weizong Zhan <sup>1</sup>, Shiyong Li <sup>2,\*</sup> and Zhencheng Chen <sup>1,\*</sup>

<sup>1</sup> School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, China; ykwu788@163.com (Y.W.); tangqunfeng@guet.edu.cn (Q.T.); 21122202029@mails.guet.edu.cn (W.Z.)

<sup>2</sup> School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China

\* Correspondence: lishiyong@guet.edu.cn (S.L.); chenzhcheng@guet.edu.cn (Z.C.)

**Abstract:** Arrhythmias are among the most prevalent cardiac conditions and frequently serve as a direct cause of sudden cardiac death. Hence, the automated detection of arrhythmias holds significant importance for assisting in the diagnosis of heart conditions. Recently, the photoplethysmography (PPG) signal, capable of conveying heartbeat information, has found application in the field of arrhythmia detection research. This work proposes a hybrid deep learning model, Res-BiANet, designed for the detection and classification of multiple types of arrhythmias. The improved ResNet and BiLSTM models are connected in parallel, and spatial and temporal features of PPG signals are extracted using ResNet and BiLSTM, respectively. Subsequent to BiLSTM, a multi-head self-attention mechanism was incorporated to enhance the extraction of global temporal correlation features over long distances. The model classifies five types of arrhythmia rhythms (premature ventricular contractions, premature atrial contractions, ventricular tachycardia, supraventricular tachycardia, and atrial fibrillation) and normal rhythm (sinus rhythm). Based on this foundation, experiments were conducted utilizing publicly accessible datasets, encompassing a total of 46,827 PPG signal fragments from 91 patients with arrhythmias. The experimental results demonstrate that Res-BiANet achieved exceptional classification performance, including an F1 score of 86.88%, overall accuracy of 92.38%, and precision, sensitivity, and specificity of 88.46%, 85.15%, and 98.43%, respectively. The outstanding performance of the Res-BiANet model suggests significant potential in supporting the auxiliary diagnosis of multiple types of arrhythmias.

**Keywords:** arrhythmia; photoplethysmography; ResNet; BiLSTM; attention mechanism



**Citation:** Wu, Y.; Tang, Q.; Zhan, W.; Li, S.; Chen, Z. Res-BiANet: A Hybrid Deep Learning Model for Arrhythmia Detection Based on PPG Signal.

*Electronics* **2024**, *13*, 665. <https://doi.org/10.3390/electronics13030665>

Academic Editor: Jichai Jeong

Received: 2 January 2024

Revised: 1 February 2024

Accepted: 4 February 2024

Published: 5 February 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Currently, cardiovascular disease is prevalent globally and is exhibiting a trend of striking at a younger age, imposing a substantial burden on global health and the economy [1]. Arrhythmias constitute a prevalent category within cardiovascular diseases [2]. Arrhythmias are mainly caused by the origin and conduction disorders of cardiac activity, leading to abnormal frequency and rhythm of heartbeats [3]. Arrhythmias can manifest with varying degrees of impact on patients. Mild cases may experience mild discomfort affecting daily life and exhibit symptoms like chest tightness and weakness. In severe cases, it can lead to heart failure and even sudden death [4]. The prevention and diagnosis of arrhythmias pose significant challenges, making the identification and detection of arrhythmias a current focal point in research. Presently, doctors utilize 12-lead electrocardiograms or 24 h dynamic electrocardiograms as the prevailing diagnostic methods for arrhythmia [5–8]. Additionally, researchers depend on these tools to automatically identify arrhythmias through modern technological means [9]. Electrodes used for electrocardiogram monitoring must be connected to the human body, and the presence of multiple lead wires can be burdensome and inconvenient for patients. In recent years, collecting

photoplethysmography (PPG) signals has gained prominence due to its advantages of low cost, portability, and wearability. Crucially, the PPG signal can convey heartbeat information [10], making it invaluable in clinical and biomedical signal processing and detection, with a wide range of applications [11,12].

Over the past decade, there has been a continuous increase in research on arrhythmia detection using PPG signals [13]. Currently, two primary methods for automatic recognition and detection are prevalent: machine learning (ML) and deep learning (DL). Machine learning methods include support vector machine (SVM) [14–17], random forest (RF) [16,18,19], decision tree (DT) [16,20], and artificial neural network (ANN) [16,21–23], etc. These machine learning methods heavily depend on manual feature extraction. The manually extracted PPG signal features primarily encompass two types: peak interval [24] and signal amplitude [25], along with their derived features [21,26]. Nevertheless, depending on manual feature extraction poses the risk of losing crucial signal information, potentially resulting in misdiagnosis. Consequently, deep learning methods, which eliminate the need for manual feature extraction, are extensively utilized in the realm of automatic arrhythmia detection [27–30]. At present, two primary deep learning methods used for arrhythmia detection are convolutional neural networks (CNN) [27–30] and recurrent neural networks (RNN) [28,29]. These studies encompass both single-type and multi-type arrhythmia detection. However, the majority of studies focus on single-type arrhythmia detection, with relatively fewer studies addressing multi-type arrhythmia detection. Shashikumar et al. [27] proposed a CNN-based deep learning model for detecting atrial fibrillation using PPG signals extracted from the wrist. This method converts PPG signals into two-dimensional spectrograms and inputs them into the model for feature extraction, integrating the characteristics of signal quality, achieving an accuracy of 91.8%. Cheng et al. [28] proposed a method that combines deep learning and time–frequency analysis to detect Atrial fibrillation (AF). They converted one-dimensional signals into two-dimensional time–frequency maps through time–frequency analysis, and then inputted the two-dimensional images into a model combining CNN and LSTM for feature learning and classification. The accuracy, sensitivity, specificity, and F1 score were 98.21%, 98.00%, 98.07%, and 98.13%, respectively. Aliamiri et al. [29] proposed a CRNN deep learning model combining CNN and RNN for detecting AF. This method employs a multimodal neural network to assess the quality of PPG signals, achieving an accuracy of 98.19% for detecting AF in 19 patients. Liu et al. [30] proposed a deep convolutional neural network based on 10 s PPG signals to classify six rhythms, including sinus rhythm (SR), premature ventricular contractions (PVC), premature atrial contractions (PAC), ventricular tachycardia (VT), supraventricular tachycardia (SVT), and AF. The overall accuracy of the VGGNet-16 model used for six heart rhythm types is 85.0%. The micro average area under the working characteristic curve of the subjects is 0.978, with an average sensitivity of 75.8%, specificity of 96.9%, positive predictive value of 75.2%, and negative predictive value of 97.0%. Our research dataset is sourced from this study.

CNN forms a deep network that can automatically learn features by stacking multiple layers of convolutional layers, and achieves end-to-end arrhythmia detection through dimensionality reduction. Long short-term memory (LSTM) [31] is a special type of RNN that has the ability to remember long and short-term information. It can selectively deliver information in long time sequences and does not cause useful information from long time ago to be forgotten. Nevertheless, LSTM can only extract sequence information from front to back and lacks the ability to capture information from back to front. Additionally, LSTM is still unable to solve long-range time-dependent problems to a certain extent [32].

In this paper, we propose a hybrid deep learning model, Res-BiANet, to automatically detect arrhythmias based on a dataset of multiple types of arrhythmia PPG signals. The method comprises three primary components: data input, feature extraction, and classification. The feature extraction module adopts a deep learning model based on improved ResNet and BiLSTM for parallel learning of features. BiLSTM has the capability to perceive both past and future information, capture richer contextual details, and enhance

the model's learning ability for PPG signals. After the BiLSTM module, a multi-head self-attention mechanism is added to enhance the ability of extracting temporal features over long distances and to alleviate the problem of time dependence over long distances. Finally, the features learned in parallel are fused and input into the fully connected layer for classification output.

## 2. Materials and Methods

### 2.1. Dataset

Liu et al. [30] disclosed the validation and test sets of their own PPG signal dataset. The publicly available data were from 91 patients with arrhythmia who had undergone radiofrequency ablation. Liu et al. [30] only disclosed 40% of the full data, but what were disclosed were the complete data they used for validation and testing, including complete arrhythmic rhythms for 91 patients. This article is based on the publicly available data from 91 patients as a dataset for the study. The 91 patients had a total of 46,827 PPG signal segments of 10 s each. The dataset contains one normal rhythm (SR) and five types of arrhythmic rhythms (PVC, PAC, VT, SVT, and AF). The labels corresponding to the six rhythms are 0, 1, 2, 3, 4, and 5, respectively. The description and quantity of each rhythm type are shown in Table 1.

**Table 1.** Dataset description and statistics.

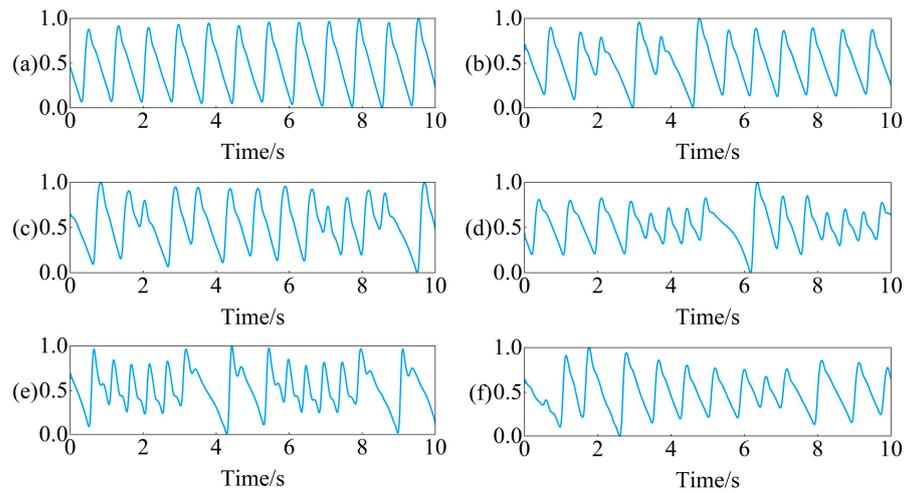
Type	Label	Description	Number of Samples
SR	0	Sinus rhythm	14,604
PVC	1	Premature ventricular contraction	4425
PAC	2	Premature atrial contraction	3773
VT	3	Ventricular tachycardia	2179
SVT	4	Supraventricular tachycardia	5677
AF	5	Atrial fibrillation	16,169
Total	/	/	46,827

Data preprocessing is essential before analyzing data and developing algorithms. As the publicly available datasets have already undergone data preprocessing and can be used directly for algorithm development studies, data preprocessing is not performed in this study.

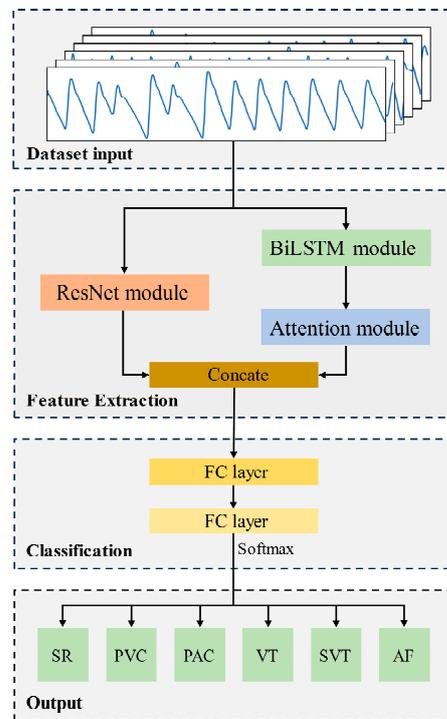
Introduction to data processing methods that data owners have already performed: Firstly, down sample the original collected PPG signal from 250 Hz to 100 Hz, and then use a 0.5–10 Hz bandpass filter to filter out the noise in the signal. After filtering, the PPG signal is divided into 10 s non overlapping segments. Finally, standardize the PPG signal every 10 s to ensure that all segments have the same scale. Examples of preprocessed PPG signals for six rhythms are shown in Figure 1.

### 2.2. Overview of Proposed Model

The proposed Res-BiANet model adopts a parallel connection method based on CNN and RNN to achieve arrhythmia detection of six rhythms. The main components of the model are composed of a feature extraction module and a classification module. The overall structure of the model is shown in Figure 2. The feature extraction module extracts the spatial and temporal features of PPG signals. The spatial feature extraction module is a residual network composed of nine residual blocks. The temporal feature extraction module is composed of four layers of BiLSTM, which integrate multi-head self-attention mechanism on the basis of BiLSTM to enhance the extraction of long-distance global temporal features and enhance the correlation between different features. Concatenate the extracted spatial and temporal features and input them into the classification module. The classification module maps the learned feature representations to the label space of the samples through two fully connected layers. Finally, the output layer outputs the predicted results for six rhythms.



**Figure 1.** Example of PPG signal segments for six rhythms: (a) SR, (b) PVC, (c) PAC, (d) VT, (e) SVT, (f) AF.



**Figure 2.** The overall structure of the proposed Res-BiANet model.

### 2.3. Feature Extraction

#### 2.3.1. Spatial Feature Extraction Module

Within the spatial feature extraction module, an improved ResNet is utilized to extract the spatial features of the PPG signals. The use of ResNet can enhance the network’s learning ability, attributed to its residual structure, subsequently improving the spatial feature extraction capability of the signal and enhancing the accuracy of heart rhythm detection.

Conventional convolutional neural networks are generally made by stacking a series of convolutional and down sampling layers, but the network degradation problem occurs when the stacking reaches a certain level [33,34]. In 2016, He et al. [35] proposed ResNet. The ResNet network introduces the concept of residual connections to address the challenge of deep network degradation. The residual structure retains the original features, ensuring smoother and more stable learning of the network, thus enhancing the accuracy and

generalization ability of the model. Additionally, to a certain extent, the complexity of learning can be reduced. The structure of the improved ResNet is shown in Figure 3. Adding a Batch Normalization (BN) [36] layer after a one-dimensional convolutional layer accelerates the convergence of the model during training. ReLU [37] is the activation function for each layer, and a Dropout [38] layer is added after the ReLU function to discard parameters and alleviate overfitting. Dropout is set to 0.5. Table 2 shows the configuration of the improved ResNet network structure parameters in the spatial feature extraction module. This module mainly consists of a pre-convolutional layer and 18 convolutional layers from 3 residual blocks. The final output dimension is 256.

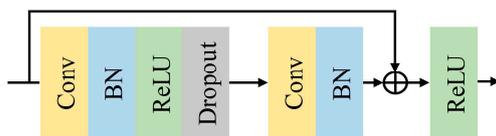


Figure 3. Improved ResNet unit structure.

Table 2. Network structure parameters of spatial feature extraction module.

Layers	Input Size	Output Size	Content
Layer 1	$1 \times 1000$	$64 \times 500$	$1 \times 7, 64, s^1 = 2, p^2 = 3$
Layer 2	$64 \times 500$	$64 \times 250$	Max pooling (3), $s = 2, p = 1$
Layer 3	$64 \times 250$	$64 \times 250$	$\begin{bmatrix} 1 \times 3, & 64 \\ 1 \times 3, & 64 \end{bmatrix} \times 3$
Layer 4	$64 \times 250$	$128 \times 125$	$\begin{bmatrix} 1 \times 3, & 128 \\ 1 \times 3, & 128 \end{bmatrix} \times 3$
Layer 5	$128 \times 125$	$256 \times 63$	$\begin{bmatrix} 1 \times 3, & 256 \\ 1 \times 3, & 256 \end{bmatrix} \times 3$

<sup>1</sup> stride; <sup>2</sup> padding.

### 2.3.2. Temporal Feature Extraction Module

Recurrent neural networks yield better results in handling sequence data. While RNNs can theoretically capture the contextual information of sequences, the information of long sequences tends to be lost as the backward propagation distance increases. Hoch Reiter et al. [31] proposed a special RNN called LSTM in 1997. In comparison to regular recurrent neural networks, LSTMs can propagate earlier information backward without forgetting it. LSTMs exhibit superior performance in tasks involving sequences with longer time spans. This study employed a bidirectional long short-term memory (BiLSTM) comprising both forward and backward propagating LSTMs. The structure of BiLSTM is shown in Figure 4, and the unit structure of LSTM is shown in the small window. The fundamental concept behind BiLSTM is that the features learned at the current moment encompass both past and future information. This accomplishes the simultaneous consideration of the global context, enhancing the extraction of time-dependent global features. In this study, a 4-layer BiLSTM with a hidden layer dimension set to 128 is employed.

LSTM conducts sequential computation and can only be computed either sequentially from front to back or back to front. This sequential computation poses two problems: (1) LSTM still encounters challenges in addressing long-range time-dependent problems to some extent. (2) The computation result at the current moment depends on the computation result at the previous moment, and this dependency limits the parallel computation capability of the model. The self-attention mechanism proposed by Ashish et al. [39] can alleviate the above problems. Self-attention mechanism is a special type of attention mechanism. When processing a sequence, it calculates the relative importance of each element and all other elements in the sequence, obtains distant dependencies between elements, and captures richer information. The self-attention mechanism performs well in handling tasks with long-distance global context. The self-attention mechanism can simultaneously

compute all positions in the input sequence, enhancing its efficiency, particularly when dealing with long sequences.

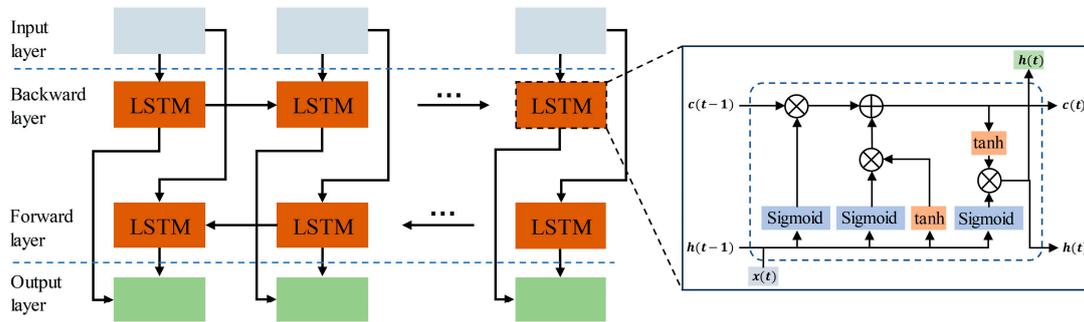


Figure 4. The overall structure of BiLSTM. The window displays the internal structure of the LSTM unit.

Specifically, for each element in the sequence, the similarity between it and other elements is first calculated and normalized to attention weights. Then, the attention weights are weighted and summed to obtain the output of the self-attention mechanism. The calculation of self-attention mechanism is shown in Equation (1).

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \times V \tag{1}$$

where  $Q \in R^{m \times d_k}$ ,  $K \in R^{n \times d_k}$  and  $V \in R^{n \times d_v}$  represent the three matrices *Query*, *Key*, and *Value* vector, respectively.  $m$  and  $n$  are the number of *Query* vector and *Key* vector, respectively.  $d_k$  and  $d_v$  represent the dimension of the *Key* vector and *Value* vector, respectively.

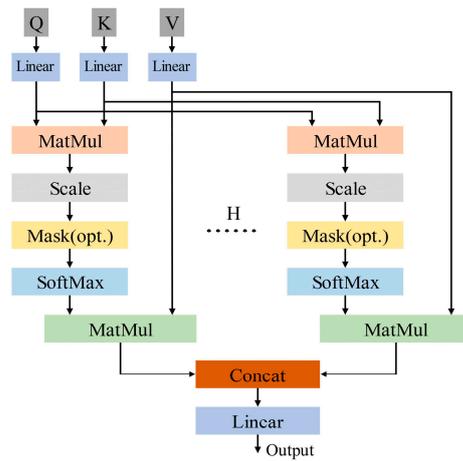
The dimensions of  $Q$ ,  $K$ , and  $V$  are the same in the self-attention mechanism. In the first step, each element of the input sequence is linearly transformed to obtain the respective  $Q$ ,  $K$ , and  $V$  vectors. In the second step, the dot products of the query vector  $Q$  and the key vector  $K$  are computed and then divided by the scaling factor  $\sqrt{d_k}$  to obtain the attention score. The purpose of dividing by the scaling factor is to avoid gradient vanishing and make the model converge better. In the third step, the attention score is mapped to a value between 0 and 1 using the Softmax function to obtain the attention weights. In the fourth step, a dot product operation is performed between the value vector of each element and its corresponding attention weight to obtain the attention output.

The multi-head self-attention mechanism is developed from the self-attention mechanism. The multi-head self-attention mechanism has multiple independent attention heads, which calculate attention weights separately and concatenate or weighted sum multiple results. Its structure is shown in Figure 5. Equations (2) and (3) are the calculation formulas for the multi-head self-attention mechanism.

$$\text{head}_i = \text{Attention}(Q_i, K_i, V_i) = \text{Softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) \times V, i \in [1, H] \tag{2}$$

$$\text{MultiHeadAttention}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_H)W^o \tag{3}$$

where  $H$  is the number of heads in the multi-head self-attention mechanism.  $W^o$  is the weight matrix. The  $H$  heads of the multi-head self-attention mechanism independently perform attention calculations to obtain their respective attention outputs, and concatenate the attention outputs of the  $H$  heads to obtain the output. Current study adopts the optimal self-attention mechanism with four heads.



**Figure 5.** The overall structure of multi-head self-attention mechanism.

#### 2.4. Experimental Environment

The model is implemented using the PyTorch 1.12.1 framework and Python 3.9. All code runs on a computer equipped with NVIDIA GeForce RTX 3060 GPU, Intel (R) Core (TM) i5-12490F CPU, and Windows 11 64-bit operating system.

#### 2.5. Experimental Metrics

To validate the performance of the model, this study recorded the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) results. Evaluate the performance of the model by using the above indicators to obtain the Precision (Pre), Sensitivity (Sen), Specificity (Spe), F1 score, and Accuracy (Acc) for each category. The definitions and calculation formulas for each evaluation indicator are as follows:

1. Pre: the proportion of all predicted positive samples to the actual positive samples. The calculation formula is shown in Equation (4).

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

2. Sen: the proportion of samples that are actually positive and predicted to be positive. The calculation formula is shown in Equation (5).

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

3. Spe: the proportion of samples that are actually negative and predicted to be negative. The calculation formula is shown in Equation (6).

$$\text{Spe} = \frac{\text{TN}}{\text{FP} + \text{TN}} \times 100\% \quad (6)$$

4. F1 score: the weighted average of model precision and recall. The calculation formula is shown in Equation (7).

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (7)$$

5. Acc: the proportion of correctly predicted quantities to the total sample size. The calculation formula is shown in Equation (8).

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \times 100\% \quad (8)$$

### 3. Results

#### 3.1. Model Training and Testing

The model proposed in this article is trained using the dataset described in Section 2.1 as input. Firstly, all the data in the dataset are randomly disrupted and divided into training set, validation set, and test set in the ratio of 6:2:2. The numbers of training set, validation set, and test set are 28,097, 9365 and 9365, respectively. The batch size for model training is set to 32. During the model training phase, the cross-entropy loss function is used to calculate the loss value, and backpropagation updates the model parameters. In addition, the Adam optimizer [40] is chosen to automatically adjust the learning rate by calculating the first-order and second-order moments of the gradient, and Adam also has the advantages of high computational efficiency and low memory usage. The initial learning rate of the Adam optimizer is 0.0001. Simultaneously using the learning rate exponential decay strategy, the decay factor  $\gamma = 0.95$ . Early stopping strategy is introduced in the training process; when the loss value of the validation set is not reduced for more than 10 epochs, the early stopping strategy will be triggered, which can effectively avoid the model overfitting problem.

Figure 6 shows the accuracy and loss curves during the model training process. It can be seen from the figure that the model triggered the early stopping strategy when it was trained to the 81st epoch, and the model has converged with good generalization. The training and validation accuracies of the model reached 92.29% and 91.81% with losses of 0.2225 and 0.2255, respectively. Figure 7 shows the confusion matrix of the classification results of the model on the test set. The numbers on the diagonal of the confusion matrix represent the number of each type of rhythm classified correctly. The overall accuracy of the model on the test set can be calculated to be 92.38%, indicating that the model has a good overall classification of the six rhythms. Other information can also be seen through the confusion matrix. The model classified SR and AF the best with an accuracy of 99.52% and 97.29%, respectively. However, it classified the PAC and VT categories the worst, with accuracies of 74.52% and 71.34%, respectively. More than 9.92% of PVC was misclassified as PAC, 8.17% of PAC was misclassified as PVC, 6.35% of VT was misclassified as SVT, and 7.11% of SVT was misclassified as VT.

In order to provide a more detailed evaluation of the model performance, Table 3 lists the Pre, Sen, Spe, F1 score, and accuracy of the model for each rhythm on the test set. From the table, it can be seen that almost all normal rhythms are recognized with accuracy of 99.52%. Pre, Sen, Spe, and F1 score are 98.04%, 99.52%, 99.10%, and 98.77%, respectively. The order of recognition ability for five types of arrhythmias is: AF > SVT > PVC > PAC > VT. The detection effect of AF is the best, with the accuracy of 97.29%, while Pre, Sen, Spe, and F1 score reached 94.55%, 94.55%, 97.12%, and 95.90%, respectively. Only 2.71% of AF were identified incorrectly. The recognition ability of VT is the worst, with accuracy of 71.34%, partly due to the minimum amount of VT data used for training. The average Pre of the six rhythms is 88.46%, the average Sen is 85.15%, the average Spe is 98.43%, the average F1 score is 86.88%, and the average accuracy is 85.60%.

**Table 3.** Detailed evaluation metrics for the model on the test set.

Type	Pre (%)	Sen (%)	Spe (%)	F1 Score (%)	Acc (%)	Overall Acc (%)
SR	98.04	99.52	99.10	98.77	99.52	92.38
PVC	83.89	82.78	98.29	83.33	82.78	
PAC	77.95	74.52	98.23	76.20	74.52	
VT	87.80	71.34	99.48	78.72	71.34	
SVT	88.53	88.16	98.35	88.34	88.16	
AF	94.55	94.55	97.12	95.90	97.29	
Average	88.46	85.15	98.43	86.88	85.60	

To visually represent the performance of the model on the test set, Figure 8a shows the ROC curves for the six different rhythms. The ROC curve uses the false positive rate

(FPR) as the horizontal coordinate and the true positive rate (TPR) as the vertical coordinate. The larger the area under the curve (AUC), the higher the accuracy, indicating better classification performance of the model. The average AUC for the six rhythms is 0.976. Similarly, AF has the highest AUC of 0.995. This indicates a high overall accuracy and good overall detection performance of the model for the six rhythms. Figure 8b shows the PR curves for the six different rhythms. The horizontal and vertical coordinates of the PR curve are the recall and precision rates, respectively, which can well represent the balance of precision and recall under different thresholds. Similarly, the larger the area under the curve (AUC), the higher the accuracy and recall, and the better the performance of the model. The micro average AUC area for the six rhythms is 0.956. This proves the feasibility of our model on imbalanced datasets.

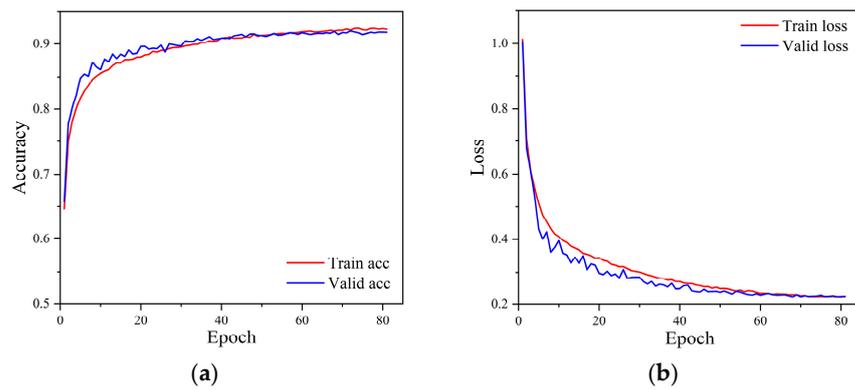


Figure 6. The accuracy and loss curve of the model training process. (a) Accuracy curves, (b) Loss curves.

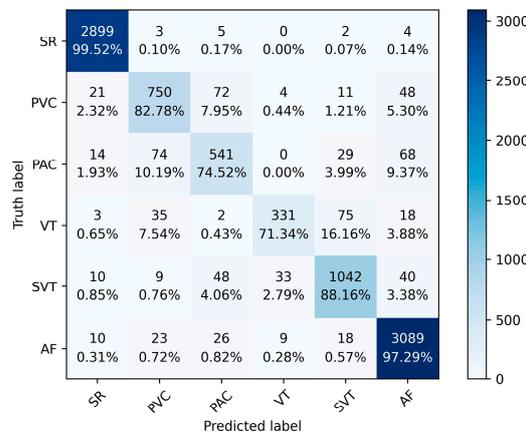


Figure 7. The confusion matrix of the model on the test set.

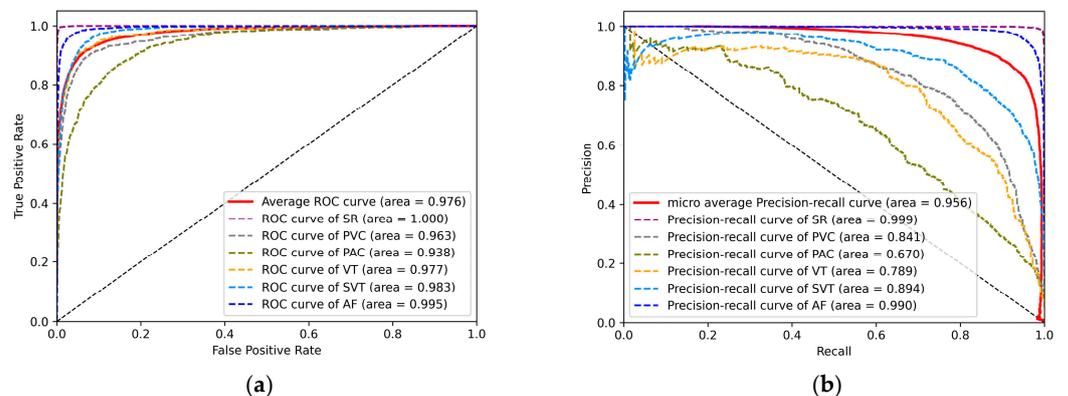


Figure 8. (a) ROC curves for six rhythms. (b) PR curves for six rhythms.

### 3.2. Result of Ablation Experiment

To verify the performance and advantages of our proposed Res-BiANet model, ablation experiments were designed. Each module of the model was gradually removed to obtain two models for ablation comparison experiments. And all experiments were subjected to five-fold cross-validation. The two models designed are as follows:

1. ResNet: Remove the temporal feature extraction module from the proposed model, leaving only the spatial feature extraction module.
2. Res-BiLSTM: Remove the attention mechanism from the temporal feature extraction module to form a parallel network of ResNet and BiLSTM.

Figure 9 shows the average confusion matrices of the ResNet, Res-BiLSTM, and Res-BiANet models on the test set. From the graph, ResNet has the worst overall classification performance, followed by Res-BiLSTM, and Res-BiANet performs the best. The ResNet model has 11.15% of PVC incorrectly identified as PAC, and 11.69% of PAC incorrectly identified as PVC. The Res-BiLSTM model reduced the error recognition rates of PVC and PAC to 8.17% and 11.50%, respectively. The Res-BiANet model reduced the error recognition rates of PVC and PAC to 7.48% and 10.56%, respectively. In the classification tasks of VT and SVT, the Res-BiANet model performs the best, with 15.57% of VT being recognized as SVT and only 3.15% of SVT being recognized as VT. In addition, the Res-BiANet model reduced the proportion of PVC recognized as AF to 7.18%.

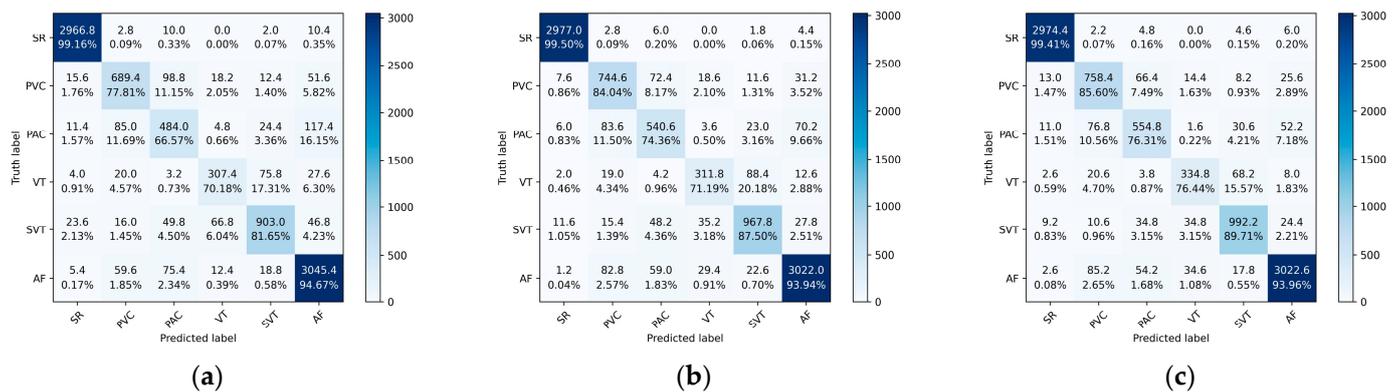


Figure 9. The average confusion matrices of ResNet, Res-BiLSTM, and Res-BiANet models after five-fold cross-validation. (a) ResNet. (b) Res-BiLSTM. (c) Res-BiANet.

Table 4 shows the average test results of three models in the ablation experiment. Among the three models, the Res-BiANet model performed the best, possessing the highest Pre, Sen, Spe, F1 score, and Acc, reaching 86.68%, 96.90%, 98.46%, 86.75%, and 92.22%, respectively. ResNet had the worst accuracy in classifying the six rhythms, with 89.64%. Next was Res-BiLSTM, with an accuracy of 91.43%. The F1 scores for the three models were 82.35%, 85.16%, and 86.75%, respectively. Res-BiANet had the highest overall classification accuracy, reaching 92.22%. After adding the BiLSTM module for extracting global temporal features on the basis of ResNet, the accuracy improved by 1.79%. The accuracy is improved by 0.79% with the addition of a multi-head self-attention mechanism for globally extracting temporal correlation features based on Res-BiLSTM. The results indicate that the added BiLSTM module has the greatest improvement on the model. The performance of the three models in the other three indicators is also ResNet < Res-BiLSTM < Res-BiANet. Overall, due to the combined effects of the spatial feature extraction module, temporal feature extraction module, and multi-head self-attention mechanism, our proposed Res-BiANet exhibits the best performance.

**Table 4.** Average results (Means + Sds) of ablation experiments.

Models	Pre (%)	Sen (%)	Spe (%)	F1 Score (%)	Acc (%)
ResNet	83.09 ± 0.62	81.67 ± 0.68	97.87 ± 0.65	82.35 ± 0.60	89.64 ± 0.37
ResNet-BiLSTM	85.33 ± 0.29	85.08 ± 0.45	98.30 ± 0.04	85.16 ± 0.32	91.43 ± 0.18
Res-BiANet	86.68 ± 0.42	86.90 ± 0.45	98.46 ± 0.32	86.75 ± 0.23	92.22 ± 0.15

### 3.3. Result of Contrast Experiment

A comparative experiment is also designed to verify the performance advantages of the model. There are the following explanations for the comparative experiment:

1. The dataset we received is not complete. The dataset publicly available from Liu et al. [30] accounts for 40% of all data. A comparison under the same amount of data is not feasible.
2. Prior to conducting this study, this publicly available dataset had not yet been found to be used in public studies.

Due to the above two issues, we also conducted comparative experiments using three classic deep learning models: AlexNet [41], VGG16 [42], and ResNet18 [35]. Table 5 presents the results of comparative experiments between the three classical models and our proposed model.

**Table 5.** Average results (Means + Sds) between three classic models and the proposed model.

Models	Pre (%)	Sen (%)	Spe (%)	F1 Score (%)	Acc (%)
AlexNet	69.90 ± 0.76	64.93 ± 0.62	95.54 ± 0.09	65.41 ± 0.48	79.89 ± 0.41
VGG16	80.35 ± 0.75	77.49 ± 0.94	97.36 ± 0.10	78.59 ± 0.86	87.34 ± 0.44
ResNet18	80.74 ± 0.19	81.02 ± 0.69	97.75 ± 0.04	80.82 ± 0.34	88.67 ± 0.14
Res-BiANet	86.68 ± 0.42	86.90 ± 0.45	98.46 ± 0.32	86.75 ± 0.23	92.22 ± 0.15

In the comparative experiment, the AlexNet model had the worst performance, with Pre, Sen, Spe, F1 score, and Acc of 69.90%, 64.93%, 95.54%, 65.41%, and 79.89%, respectively. Compared to AlexNet, VGG16 is a model with a deeper depth and uses small convolution kernels instead of large ones, which reduces parameters and improves classification accuracy. As a result, VGG16 is higher than AlexNet in all evaluation metrics, where the accuracy is improved to 87.34%. As the depth of the model increases, VGG will have the problem of vanishing gradients. In order to solve the problem of gradient vanishing, ResNet introduces a residual structure to further increase the depth of the model and improve accuracy. Consequently, ResNet18 performs better than VGG16, increasing the accuracy to 88.67%. Overall, with the same amount of data, the performance of the Res-BiANet model is significantly improved compared to AlexNet, VGG16, and ResNet18.

## 4. Discussion

In this article, our proposed Res-BiANet model achieved accurate recognition of six rhythms and demonstrated good performance. Res-BiANet is a parallel structured hybrid deep learning model that combines residual networks and a bidirectional long short-term memory network based on a multi-head self-attention mechanism. The model takes a 10 s PPG signal as input, and two channels extract spatial and temporal features of the PPG signal. The classifier accomplishes classification of six rhythms. The F1 score and overall accuracy of the six rhythm classifications reached 86.88% and 92.38%, respectively, while the Pre, Sen, and Spe were 88.46%, 85.15%, and 98.43%, respectively.

We conducted ablation experiments on the proposed model to assess the impact of each part of the model on its performance. The results of the five-fold cross-validation experiment showed that the addition of the BiLSTM module for extracting global temporal features improved the F1 score by 2.81% and overall accuracy by 1.79%. Upon incorporating a multi-head self-attention mechanism for extracting temporal correlation features globally, the F1 score and overall accuracy improved by 1.59% and 0.79%, respectively. The added BiLSTM module

exhibited the most significant improvement in the model's performance, followed by the multi-head self-attention mechanism. Subsequent comparisons are made with three classical deep learning models with the same amount of data. The results indicate that our proposed model maintains performance advantages even when using the same amount of data.

Table 6 shows the comparison between this study and previous studies. As can be seen in Table 6, different studies used different datasets and data volumes. Detection of atrial fibrillation was the most common among the single category of arrhythmia detection. Shashikumar et al. [27] proposed a CNN-based deep learning model for detecting AF from wrist-extracted PPG signals. The method converted ninety-eight 30 s PPG signals into 2D spectrograms as inputs to the model and achieved 91.8% accuracy in detecting AF in 98 patients. Cheng et al. [28] used 28,440 PPG signal segments from three datasets totaling 102 patients for AF detection. The one-dimensional PPG signals were converted into two-dimensional time–frequency maps by time–frequency analysis, and the two-dimensional images were input into a combined CNN and LSTM model for feature learning and classification. The accuracy, sensitivity, specificity, and F1 score of this study were 98.21%, 98.00%, 98.07%, and 98.13% for the AF detection task, respectively. Aliamiri et al. [29] collected a total of 1443 PPG signal segments from 19 patients using a wearable device. A hybrid CNN and RNN deep learning model, CRNN, was used, which detected AF with an accuracy of 98.19% in 19 patients. For the task of detecting multiple types of arrhythmias, Neha et al. [23] selected 670 PPG signal segments from 23 records in the publicly available dataset PhysioNet MIMIC-II. A dynamic time-warping (DTW) method was used to extract the warping features of the PPG signals as input to the ANN. The accuracy, sensitivity, specificity, F1 score, and precision for the detection of four types of rhythms (atrial flutter, normal, premature ventricular beats, and tachycardia) were 95.97%, 97%, 97%, and 96%, respectively. Liu et al. [30] used a self-constructed dataset including 228 patients totaling 118,217 PPG signal segments. They used a deep convolutional neural network for six types of arrhythmia detection, and the overall accuracy was 85.0%, while Pre, Sen, and Spe were 75.2%, 75.8%, and 96.9%, respectively. In contrast, our proposed Res-BiANet model shows better performance advantage with smaller data volumes, with an overall accuracy of 92.22% and an F1 score of 86.75%. The figures are 86.68%, 86.90%, and 98.46% for Pre, Sen and Spe, respectively. The results of our study are based on the data volume of 91 patients totaling 46,827 PPG signal segments. From Table 6, it can be seen that the dataset and data volume used in the arrhythmia detection task based on PPG signals are not the same. Therefore, there are certain limitations in comparing our research results with previous studies, especially the comparison with the data sources we used. The study by Liu et al. [30] was based on 228 patients totaling 118,217 PPG segments, and the patient data in the divided training, validation, and test sets were independent of each other, which is different from our dataset division strategy, where we took the approach of disrupting the data of the 91 patients used to divide the training, validation, and test sets.

We have tried to use the PPG signals in the publicly available dataset PhysioNet MIMIC II for arrhythmia identification and classification, but unfortunately the PPG signal segments about arrhythmias in PhysioNet MIMIC II are not annotated with labels, and we were not able to perform supervised learning. In addition, there is a single type of arrhythmia based on PPG signals in PhysioNet MIMIC II, and the goal of our study was to develop an automated detection method for multiple types of arrhythmias; therefore, we did not use PhysioNet MIMIC II or other publicly available datasets that are used by most researchers.

While our model has enhanced its recognition ability for arrhythmia types with lower recognition rates, such as PAC and VT, it still maintains an error recognition rate of nearly 25%. Moving forward, we will focus on these misclassified categories and continuously improve and optimize the network model. We will enhance the model's recognition ability for such arrhythmia signals by improving the feature extraction ability of peak intervals and amplitude sizes of PPG signals. Additionally, the method proposed in this article has not undergone clinical trials. In the future, we plan to apply this method to practical life

scenarios, such as using it as an auxiliary diagnostic tool for doctors and enabling portable detection of arrhythmias on wearable devices.

**Table 6.** Comparison with previous studies.

References	Database	Number of Subjects	Data Volume	Detection	Method	Result
Shashikumar et al. [27]	Self-generated database	98	98	AF	CNN	AUC: 0.95 Acc: 91.8%
Cheng et al. [28]	MIMIC-III waveform database; IEEE TAME Respiratory Rate Benchmark dataset; Synthetic dataset	102	28,440	AF	2D-CNN + LSTM	Sen: 98.00% Spe: 98.07% F1 score: 98.13% Acc: 98.21% AUC: 0.9959
Aliamiri et al. [29]	Self-generated database	19	1443	AF	CNN + RNN	AUC:99.67% Acc: 98.19%
Neha et al. [23]	PhysioNet MIMIC-II database	23	670	PVC, AFI, ST, and normal sinus rhythm	ANN	Pre: 96% Sen: 97% Spe: 97% F1 score: 96% Acc: 95.97%
Liu et al. [30]	Self-generated database	228	118,217	PVC, PAC, VT, SVT, and AF	DCNN	Pre: 75.2% Sen: 75.8% Spe: 96.9% Acc: 85.0%
This work	From the dataset in Ref. [30]	91	46,827	PVC, PAC, VT, SVT, and AF	Res-BiANet	Pre: 88.46% Sen: 85.15% Spe: 98.43% F1 score: 86.88% Acc: 92.38%

## 5. Conclusions

The automatic detection of multiple types of arrhythmias is currently a popular topic in auxiliary diagnosis, especially in the detection of arrhythmias based on portable PPG signals. This article introduces Res-BiANet, a deep learning model that integrates parallel CNN and RNN structures based on PPG signals for detecting multiple types of arrhythmias. The CNN employs an improved ResNet network to extract spatial features of PPG signals, while the RNN employs a BiLSTM network to extract global temporal features and incorporates a multi-head self-attention mechanism to enhance the extraction of global temporal correlation features. The experimental results demonstrated that Res-BiANet performed well in the classification task of five types of arrhythmias and normal heart rhythm. The F1 score and overall accuracy of six rhythm classifications reached 86.88% and 92.38%, respectively. Notably, with the addition of the temporal feature extraction module, the model's recognition rates for PVC, PAC, and VT increased by 7.79%, 9.74%, and 6.26%, respectively. This finding suggests that our model has significantly improved recognition abilities for arrhythmia categories with similar morphology. Therefore, our model has great potential in the auxiliary diagnosis of multiple types of arrhythmias detection based on PPG signals.

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

**Funding:** This research was funded by the project supported by the Joint Funds of the National Natural Science Foundation of China (U22A2092), the National Major Scientific Research Instrument and Equipment Development Project (61627807), and the Guangxi Science and Technology Major Special Project (2019AA12005).

**Data Availability Statement:** The data used in this study were obtained from publicly available datasets by Liu et al. [30].

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

## References

- Virani, S.S.; Alonso, A.; Aparicio, H.J.; Benjamin, E.J.; Bittencourt, M.S.; Callaway, C.W.; Carson, A.P.; Chamberlain, A.M.; Cheng, S.; Delling, F.N.; et al. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. *Circulation* **2021**, *143*, e254–e743. [[CrossRef](#)] [[PubMed](#)]
- Al-Khatib, S.M.; Stevenson, W.G.; Ackerman, M.J.; Bryant, W.J.; Callans, D.J.; Curtis, A.B.; Deal, B.J.; Dickfeld, T.; Field, M.E.; Fonarow, G.C.; et al. 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death. *Circulation* **2018**, *138*, e210–e271. [[PubMed](#)]
- Landstrom, A.P.; Dobrev, D.; Wehrens, X.H.T. Calcium Signaling and Cardiac Arrhythmias. *Circ. Res.* **2017**, *120*, 1969–1993. [[CrossRef](#)]
- Mehra, R. Global public health problem of sudden cardiac death. *J. Electrocardiol.* **2007**, *40*, S118–S122. [[CrossRef](#)] [[PubMed](#)]
- Enriquez, A.; Baranchuk, A.; Briceno, D.; Saenz, L.; Garcia, F. How to use the 12-lead ECG to predict the site of origin of idiopathic ventricular arrhythmias. *Heart Rhythm* **2019**, *16*, 1538–1544. [[CrossRef](#)]
- Chua, S.K.; Chen, L.C.; Lien, L.M.; Lo, H.M.; Liao, Z.Y.; Chao, S.P.; Chuang, C.Y.; Chiu, C.Z. Comparison of Arrhythmia Detection by 24-Hour Holter and 14-Day Continuous Electrocardiography Patch Monitoring. *Acta Cardiol. Sin.* **2020**, *36*, 251–259.
- Hammad, M.; Iliyasa, A.M.; Subasi, A.; Ho, E.S.L.; El-Latif, A.A.A. A Multitier Deep Learning Model for Arrhythmia Detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–9. [[CrossRef](#)]
- Daydulo, Y.D.; Thamineni, B.L.; Dawud, A.A. Cardiac arrhythmia detection using deep learning approach and time frequency representation of ECG signals. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 232. [[CrossRef](#)]
- Yao, Q.H.; Wang, R.X.; Fan, X.M.; Liu, J.K.; 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](#)]
- Alian, A.A.; Shelley, K.H. Photoplethysmography. *Best Pract. Res. Clin. Anaesthesiol.* **2014**, *28*, 395–406. [[CrossRef](#)]
- Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **2007**, *28*, R1–R39. [[CrossRef](#)]
- Deshpande, A.; Mandlik, S.A.; Lakhe, A.S.; Jethi, J.V.; Sinha, V. Photoplethysmography and Its Clinical Application. *MGM J. Med. Sci.* **2017**, *4*, 89–96. [[CrossRef](#)]
- Neha; Sardana, H.K.; Kanwade, R.; Tewary, S. Arrhythmia detection and classification using ECG and PPG techniques: A review. *Phys. Eng. Sci. Med.* **2021**, *44*, 1027–1048. [[CrossRef](#)] [[PubMed](#)]
- Shan, S.M.; Tang, S.C.; Huang, P.W.; Lin, Y.M.; Huang, W.H.; Lai, D.M.; Wu, A.Y.A. Reliable PPG-based algorithm in atrial fibrillation detection. In Proceedings of the 2016 IEEE Biomedical Circuits and Systems Conference (BioCAS), Shanghai, China, 17–19 October 2016; pp. 340–343.
- Schäck, T.; Harb, Y.S.; Muma, M.; Zoubir, A.M. Computationally Efficient Algorithm for Photoplethysmography-Based Atrial Fibrillation Detection Using Smartphones. In Proceedings of the 39th Annual International Conference of the IEEE-Engineering-in-Medicine-and-Biology-Society (EMBC), Jeju, Republic of Korea, 11–15 July 2017; pp. 104–108.
- Neha; Kanawade, R.; Tewary, S.; Sardana, H.K. Photoplethysmography based Arrhythmia Detection and Classification. In Proceedings of the 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 7–8 March 2019; pp. 944–948.
- Neha; Sardana, H.K.; Kanawade, R.; Dogra, N. Photoplethysmograph based arrhythmia detection using morphological features. *Biomed. Signal Process. Control* **2023**, *81*, 104422. [[CrossRef](#)]
- Eerikainen, L.M.; Bonomi, A.G.; Schipper, F.; Dekker, L.R.C.; de Morree, H.M.; Vullings, R.; Aarts, R.M. Detecting Atrial Fibrillation and Atrial Flutter in Daily Life Using Photoplethysmography Data. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 1610–1618. [[CrossRef](#)] [[PubMed](#)]
- Han, D.; Bashar, S.K.; Zieneddin, F.; Ding, E.; Whitcomb, C.; McManus, D.D.; Chon, K.H. Digital Image Processing Features of Smartwatch Photoplethysmography for Cardiac Arrhythmia Detection. In Proceedings of the 42nd Annual International Conference of the IEEE-Engineering-in-Medicine-and-Biology-Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 4071–4074.
- Fallet, S.; Lemay, M.; Renevey, P.; Leupi, C.; Pruvot, E.; Vesin, J.M. Can one detect atrial fibrillation using a wrist-type photoplethysmographic device? *Med. Biol. Eng. Comput.* **2019**, *57*, 477–487. [[CrossRef](#)] [[PubMed](#)]
- Solosenko, A.; Marozas, V. Automatic Premature Ventricular Contraction Detection in Photoplethysmographic Signals. In Proceedings of the IEEE Biomedical Circuits and Systems Conference (BioCAS), Lausanne, Switzerland, 22–24 October 2014; pp. 49–52.
- Solosenko, A.; Petrenas, A.; Marozas, V. Photoplethysmography-Based Method for Automatic Detection of Premature Ventricular Contractions. *IEEE Trans. Biomed. Circuits Syst.* **2015**, *9*, 662–669. [[CrossRef](#)] [[PubMed](#)]
- Neha; Sardana, H.K.; Dogra, N.; Kanawade, R. Dynamic time warping based arrhythmia detection using photoplethysmography signals. *Signal Image Video Process.* **2022**, *16*, 1925–1933. [[CrossRef](#)]

24. Suzuki, T.; Kameyama, K.i.; Tamura, T. Development of the irregular pulse detection method in daily life using wearable photoplethysmographic sensor. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 6080–6083.
25. Han, D.; Bashar, S.K.; Lazaro, J.; Ding, E.; Whitcomb, C.; McManus, D.D.; Chon, K.H. Smartwatch PPG Peak Detection Method for Sinus Rhythm and Cardiac Arrhythmia. In Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4310–4313.
26. Chen, X.; Huang, J.H.; Luo, F.F.; Gao, S.; Xi, M.; Li, J. Single channel photoplethysmography-based obstructive sleep apnea detection and arrhythmia classification. *Technol. Health Care* **2022**, *30*, 399–411. [[CrossRef](#)] [[PubMed](#)]
27. Shashikumar, S.P.; Shah, A.J.; Li, Q.; Clifford, G.D.; Nemati, S. A Deep Learning Approach to Monitoring and Detecting Atrial Fibrillation using Wearable Technology. In Proceedings of the 4th IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Orlando, FL, USA, 16–19 February 2017; pp. 141–144.
28. Cheng, P.; Chen, Z.C.; Li, Q.Z.; Gong, Q.; Zhu, J.M.; Liang, Y.B. Atrial Fibrillation Identification With PPG Signals Using a Combination of Time-Frequency Analysis and Deep Learning. *IEEE Access* **2020**, *8*, 172692–172706. [[CrossRef](#)]
29. Aliamiri, A.; Shen, Y. Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor. In Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Las Vegas, NV, USA, 4–7 March 2018; pp. 442–445.
30. Liu, Z.; Zhou, B.; Jiang, Z.; Chen, X.; Li, Y.; Tang, M.; Miao, F. Multiclass Arrhythmia Detection and Classification From Photoplethysmography Signals Using a Deep Convolutional Neural Network. *J. Am. Heart Assoc.* **2022**, *11*, e023555. [[CrossRef](#)] [[PubMed](#)]
31. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
32. Chien, H.-Y.S.; Turek, J.; Beckage, N.M.; Vo, V.A.; Honey, C.J.; Willke, T.L. Slower is Better: Revisiting the Forgetting Mechanism in LSTM for Slower Information Decay. *arXiv* **2021**, arXiv:2105.05944.
33. He, K.; Sun, J. Convolutional neural networks at constrained time cost. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5353–5360.
34. Srivastava, R.K.; Greff, K.; Schmidhuber, J. Training Very Deep Networks. *Comput. Sci.* **2015**, *2*, 2377–2385.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
36. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 448–456.
37. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
38. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
39. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4 December 2017; pp. 6000–6010.
40. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
41. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
42. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.