 doi:10.20944/preprints202204.0241.v2

A one-dimensional serial ECG diagnosis approach based on ensemble deep learning

Siqing Jing[1],Xiaojun Chen[2], Haojun Gao[1],Jian Wu[1,2,*]

[1] Zhejiang University School of Medicine, and department of Public Health, Zhejiang University, Hangzhou, Zhejiang, China;jsq522748392@126.com (S.J.); 22018666@zju.edu.cn (H.G.)
[2] Real-doctor Artificial Intelligence Research Center, Zhejiang University, Hangzhou, Zhejiang, China; chenxiaojun@zju.edu.cn (X.C.)
*: Correspondence: wujian2000@zju.edu.cn (J.W.)

Abstract: An electrocardiograph (ECG) reflects the health of the human heart and is used to help diagnose arrhythmia and myocardial infarction(MI) in clinical practice. Early diagnosis of arrhythmia helps implement preventive measures and plays a crucial role in saving a patient's life. With the increasing demand of clinicians for ECG analysis technology, intelligent detection and diagnosis of ECG signals has become a more efficient means to assist physicians in diagnosing cardiovascular diseases. This paper introduces an ECG diagnosis approach based on an ensemble deep learning combination of CNN(convolutional neural network) and SLAP(stacked-long short term memory architecture for prediction) architecture. ECG data is denoised and further divided into single heartbeats to achieve data standardization and sample diversity. Adam optimizer and BCEwithlogitsloss multi-classification loss function were used to enhance the model effect, and the system achieved the classification effect of 99.3% average accuracy, 99.0% F1-value, and 99.2% sensitivity in MIT-BIH standard database classification.

Key words: deep learning, ensemble learning, intelligent detection and diagnosis, multi-classification, preventive measures

1. Introduction

Cardiovascular disease (CVD) remains the leading cause of death and disability globally and will continue to grow[1]. An estimated 17.9 million people died from CVDs in 2019, representing 32% of all global deaths. Of these deaths, 85% were due to heart attack or stroke[2]. In addition, the world's population is aging faster these years; according to incomplete statistics, between 2015 and 2050, the proportion of the world's population over 60 years will nearly double from 12% to 22%[3], the prevalence of cardiac increases rapidly, and it will be a substantial medical resources burden of cardiovascular diseases[4] such as arrhythmia and myocardial infarction[5] (MI) to all the countries. However, the number of cardiovascular physicians is limited, as physicians cannot consistently maintain the best performance all the time due to fatigue, and the phenomenon of missed diagnosis and misdiagnosis may occur[6]. In addition, it isn't easy to achieve consistency in diagnostic standards of diagnosis due to the differences in expertise among physicians. Therefore, we proposed an electrocardiograph signal diagnosis system that can solve the problems existing in the medical electrocardiograph field. We need to first input the ECG signal voltage data through the ECG segmentation system for single heartbeat segmentation; in this phase, the abnormal location map is accessible, then shifted to the classification system to diagnose heart disease, which provides an auxiliary diagnostic basis for further clinical decision making. The innovation of this work lies in the ECG signal serial automatic segmentation approach and the ensemble learning of Resnet[7] and SLAP[8], which can make full use of the time series features by the LSTM (long short term

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

memory) network while extracting image features from the CNN (convolutional neural network). To achieve the effect of preferential classification, we give appropriate weight to the single learner with better performance and visualize the ECG signal by presenting segmentation maps, which assist doctors in interpreting each heartbeat more effectively.

As one of the most popular artificial intelligence technologies, deep learning(DL) has shown excellent performance in the research of electrocardiogram intelligent classification in recent years[9]. In the medical electrocardiograph field, CNN and RNN(recurrent neural network) are two significant deep learning methods for ECG detection and classification tasks. Two dimensions can represent the general characteristics of ECG signal data:

(1)One-dimensional electro-cardio signal formed by voltage data.

(2)Two-dimensional ECG diagram with time as X-axis and voltage as Y-axis.

In recent years, researchers have some advances in various ECG detection tasks. An 11-layer deep CNN for automatic arrhythmia detection using with or without augmented achieved an accuracy of 87.6%[10]. Zheng et al. [11] suggested a combination of CNN and LSTM methods learned from Nine-angles in each arrhythmia ECG image, which makes the learned information random to prevent the over-fitting caused by the fixed learned features and improves the generalization ability, whereas the weakness is that it adopted to connect CNN directly with LSTM in series, which would result in the length of the input part of LSTM being compressed by CNN and time-series features of the original information could not be well extracted, and result in inevitable information loss related to time series. In Niu's experiment[12], researchers proposed a novel method for ECG classification based on adversarial domain adaptation, Gaussian distribution adopted to deal with the random error of sample distribution, which improves the phenomenon of different data distribution caused by individual differences. Still, the disadvantage of adversarial domain adaptation is that it requires testing samples to participate in the training phase[13]. The researcher Ahmad[14] put forward ideas of two computationally efficient multimodal fusion frameworks for ECG heartbeat classification, the main feature of their experiment is transforming single-dimensional ECG signals into three types of two-dimensional images: Markov Transition Field (MTF), Recurrence Plot (RP), and Gramian Angular Field (GAF), respectively, a remarkable effect been achieved in the study, but only proposed that the three types of electrocardiogram be used in the experiment in two ways of combining features, whereas the specific cause for the combination is not expressly explained. In the machine learning of Emanet[15] and Urtnasan[16], Random Forest(RF) and SVM classifier are used respectively for electrocardiogram detection and diagnosis, which can both achieve good results in the case of small-size samples but lack an evaluation of the generalization ability in multiple datasets.

2. Materials and Methods

We pre-process the ECG signals to be input and then extract the features of the processed ECG data. And then, we make a classification by assigning weights to the target mapping of different models employing ensemble learning. The overall flow chart we proposed is shown in Figure 1.

Figure 1. Overall flow chart of this work

2.1. Dataset

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

We perform an experiment using the Massachusetts Institute of Technology-Beth Israel Hospital(MIT-BIH) Arrhythmia Database[17] and Tianchi Dataset[18]. The MIT-BIH database contains 48 fully annotated half-hour 2-lead ECG signals, collected at a frequency of 360 hertz with 11-bit resolution over a 10mv range, mainly used for training and validating in this work. In order to make these labels more practical and improve medical interpretability, the data classification catalog is set according to the international AAMI EC57 standard[19]. Therefore, the labels in the MIT-BIH arrhythmia database are overall divided into 5 categories:

- Supraventricular ectopic beat(SVEB, S)
- Ventricular ectopic beat(VEB, V)
- Fusion beat(F)
- Unknown beat(Q)
- Any heartbeat not in the SVEB, VEB, F, or Q classes(N)

Tianchi Dataset contained 55 categories of ECG labels, and 20662 samples were screened. Of these samples,8265 samples were selected as the final test sets by random sampling(1-20662 random numbers were used to generate 40% serial numbers of the original sample size). We read the voltage sequence of each heartbeat through the ECG segmentation system, and the fixed-length is set to 400 points. Samples of insufficient length automatically add '0' treatment. While reading each test sample, we extracted utilizing random sampling of three heartbeats (Setting these heartbeat labels to the same)as a test set. Before testing, only two of the 12-leads corresponding to the MIT-BIH Database are read, and the input voltage of the model is set as ×4.88mV to adapt to the input format. By comparing the MIT-BIH label types, we reserved 15 annotations as the original labels according to the AAMI EC57 category[19] classification standard. We finally divided these original labels into five categories for testing (Table 1).

Table 1. Mapping annotations with AAMI EC57[19] categories.

Categories	MIT-BIH	Annotations in Tianchi
N	. or N	Normal
	L/R	Left/Right bundle branch block
	j	Atrial escape
	E	Nodal escape
S	a	Atrial Premature
	J	Aberrant atrial premature
	A	Nodal premature
	S	Supra-ventricular premature
V	V	Premature ventricular contraction
	e	Ventrucular and normal
F	F	Fusion of paced and normal
Q	/	Paced
	Q	Fusion of paced and normal
	f	Unclassifiable

2.2. Data Pre-processing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

In the process of data set training, the effect of the training model largely depends on the quality of input data. The existence of noise in ECG signals affects the quality of data to impacts the learning ability of the model. Common noise in ECG signals includes myoelectric interference, power frequency interference, and baseline drift[21]. To unify the input data format, In this phase, we optimize raw record data quality and separate each heartbeat before training. The following steps are specific methods of noise elimination in this work(Figure 3):

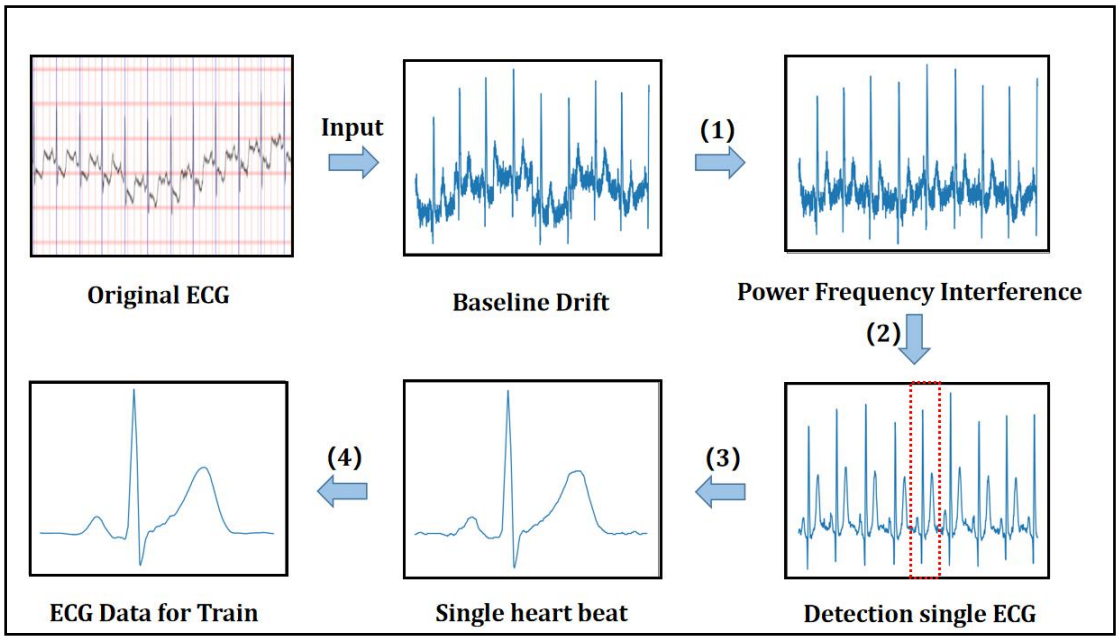


Figure 3. ECG signal detection and noise elimination.

(1) Handling baseline drift. Smoothing the ECG baseline drift signal containing waveform mutation to obtain smooth and slow baseline drift signals and then subtracting soft signals from the originally recorded data. [22]

(2) Control of power frequency interference. The power frequency interference is caused by the power supply environment of ECG signal acquisition equipment, the waveform like a sinusoidal signal. In this work, we need to subtract the sinusoidal filter from the ECG that we got in the step(1) to remove power frequency interference. Using band pass filter with a cutoff frequency of (0.6,35) and db8 to eliminate the influence of MA(muscle artifact noise) and EMG (electrode artifact noise).[23]

(3) Single heartbeat segmentation and alignment. The signal data points of each heartbeat *R* wave can be obtained from the annotation file in MIT-BIH arrhythmia database. Suppose R_i is the position of the i^{th} heartbeat *R* peak. According to the length of cardiac cycle of normal people,[24] the first position heartbeat is $\lfloor \frac{3}{8}(R_{i-1}+R_i) \rfloor$, the last position heartbeat is $\lfloor \frac{3}{8}(R_i+R_{i+1}) \rfloor$, where $\lfloor \mathrm{N} \rfloor$ means rounding down to N. Therefore, the present heartbeat sampling points is $K_i = \lfloor \frac{3}{8}(R_i+R_{i+1}) \rfloor - \lfloor \frac{3}{8}(R_{i-1}+R_i) \rfloor + 1$. In order to pass into the deep learning model, the heartbeats must be aligned. Suppose the ECG series points after unification is L^{align} (400 points in this paper); if K_i is less than L^{align}, then fill with zero to L^{align}, and if K_i is greater than L^{align}, crop to L^{align}, obtain the final heartbeat K_{final}. In order to eliminate offset and amplitude scaling problems in signal, the aligned heartbeats K_{final} are standardized by the formula:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

$$Z_{k-final} = \frac{K_{final} - \mu}{\sigma} \quad (1)$$

(4) Remove myoelectric interference. This step utilizes the wavelet transform tool PyWavelets[25] to analyze ECG signals in the time and frequency domain to obtain wavelet coefficients of each scale. After wavelet transform scale decomposition, the wavelet coefficients with larger amplitude are proper signals, while the ones with smaller amplitude are noise[26-27], eliminating the noise.

2.3. ECG diagnosis model

The diagnosis system mainly includes two parts, the ECG pre-processing system and the ECG classification system. The former is an auxiliary segmentation part of this system, which provides preconditions for subsequent multi-classification by inputting the original electrocardiograph signals, and this part will be mentioned in data pre-processing. The latter is used to diagnose heart disease by using ensemble learning. Figure 2 summarizes the main framework of the proposed ECG diagnosis system.

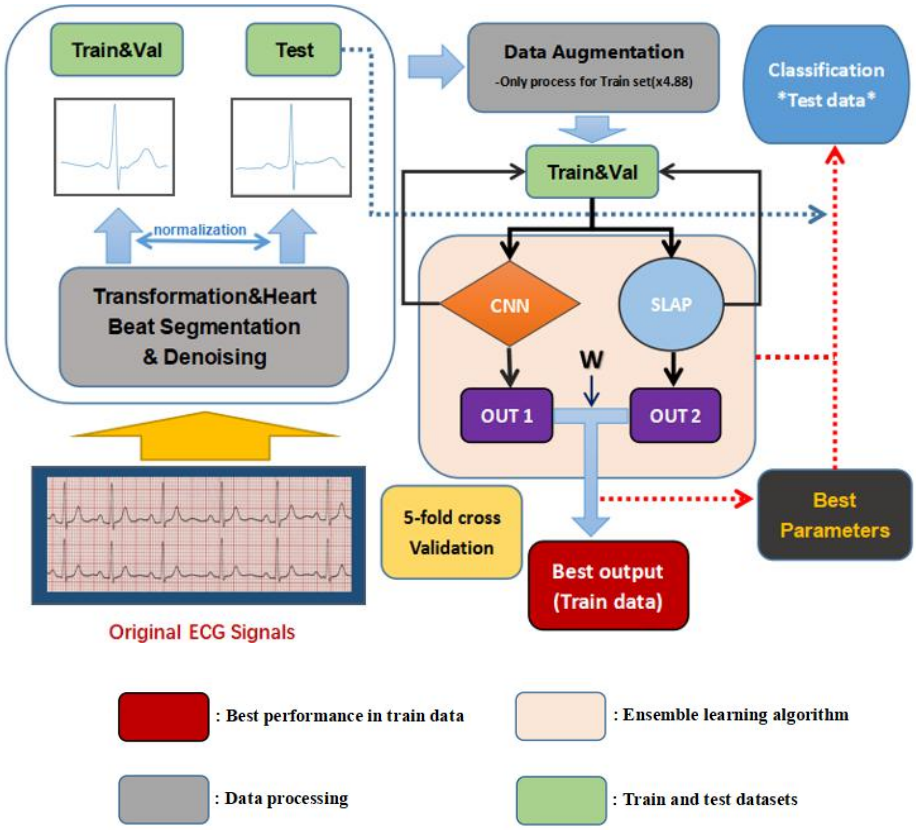


Figure 2. ECG diagnosis system

2.4. Ensemble Learning

One of the comprehensive learning methods used in this experiment is ensemble learning. It is a way to classify more effectively by constructing and combining multiple machine learning tasks and algorithms. A successful integration system needs a good learner and corresponding integration combination strategy[28]. In this study, the ensemble learning method adopted the bagging algorithm. According to the principle of random sampling, samples were random selected into two sub-training

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

datasets $S_n \in \{S_1, S_2\}$, respectively. The classifiers C_1 and C_2 represent ResNet and Stacked-LSTM classifiers, respectively. Further obtained a stronger classifier C^{strong} .

2.5. ResNet-34 Model

Convolutional neural networks can be used in one-dimensional to multi-dimensional situational tasks. The existing convolutional neural network is used primarily in two-dimensional images. As a two-dimensional image, an electrocardiograph can also be done using CNN technology to complete the classification tasks. We adopted the one-dimensional network structure with 34-layer ResNet, changed the original two-dimensional convolution into one-dimensional convolution, and the backbone framework is shown in Table 2.

We input the one-dimensional ECG data into the one-dimensional ResNet-34, setting the convolution kernel of the first layer as 15x1, several channels as 64, stride as 2, and the padding as 7 to adapt the following input size standards. The fully connected (FC) layer is 5x1, corresponding to five classification results, as shown in Figure 4. In this experiment, ResNet was used to achieve a good classification effect because the residual unit in ResNet could establish an associated channel identity between the input and output by short-circuiting. ResNet short circuit mechanism[29] solves the problem of information loss, which makes the result not only the new feature extracting, but to some extent, still retains the original features.

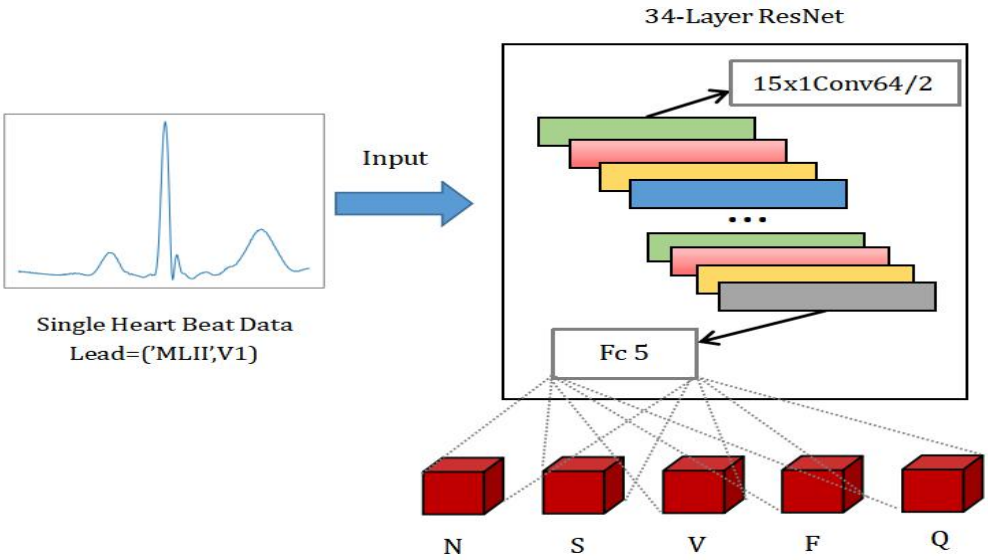


Figure 4. Heart beat classifier(34-Layer ResNet) schematic diagram.

Tabel 2. One-dimensional ResNet-34 backbone block.

Block	Layer	Output (N, C_{out}, L_{out})
Input	Input layer	Nx2x400
Conv1_1d	K=15,c=64,p=7,s=2	Nx64x200
Conv2_x_1d	K=3,max pool,s=2	Nx64x100
	$\begin{pmatrix} k=3,\ out\_c=64 \\ k=3,\ out\_c=64 \end{pmatrix} \times 3$	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

Conv3_x_1d	$\begin{pmatrix} k=3, out\_c=128 \\ k=3, out\_c=128 \end{pmatrix} \times 4$	Nx128x50
Conv4_x_1d	$\begin{pmatrix} k=3, out\_c=256 \\ k=3, out\_c=256 \end{pmatrix} \times 6$	Nx256x25
Conv5_x_1d	$\begin{pmatrix} k=3, out\_c=512 \\ k=3, out\_c=512 \end{pmatrix} \times 3$	Nx512x13
Average pool	AdaptiveAvgPool1d	Nx512x1
Fc,Sigmoid	Linear	Nx5x1

K: Kernel_size, P: Padding, S: Stride, C_{out}/out_c: Out channel.

N: batch size , L_{out}: Out length.

2.6. SLAP Model

A recurrent neural network is used to solve the sequential tasks with time series characteristics. We utilize the LSTM[30] network of the recurrent neural network as the classifier, which has better memory efficiency than the RNN(recurrent neural network) structural network and can solve the problem of gradient explosion and gradient disappearance. The structure of the LSTM unit mainly includes three parts: input gate, output gate, and forget gate.

The input gate mainly dealing with the input of time series data by assigning weights w_i and w_c to h_{t-1} and x_t , adding bias vectors b_i and b_c , obtaining the following formula through their respective activation functions.

$$i_t = f_\sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

The forget gate, a processing structure of long-term memory signal C_t in LSTM, which can selectively and selectively forget some invalid information in h_{t-1} and x_t . By weighting W_f , adding the paranoid vector b_f , through the activation function:

$$f_t = f_\sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$

C_t was obtained by mediation variable processing,this structure can selectively remember valid information and retain the main information for a long time:

$$C_t = C_{t-1} \otimes f_t + i_t \otimes \tilde{C}_t \tag{5}$$

The output gate determined which outputs will be considered for the current state:

$$o_t = f_\sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t \otimes \tanh(C_t) \tag{7}$$

Stacked neural network has at least two LSTM sub-layers, which using different types of weight parameters(Figure 5), the output of a upper hidden layer is the input of the next hidden layer, which improves the effect of feature extraction, especially for ECG signal.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

Formula 8 defines the data transformation process.

$$i^{next} = f(\sum_{n=1}^{Num} W_n^{next} \times o_n + b^{next}) \tag{8}$$

Where i^{next} is the input data of a next hidden layer in SLAP; W_n^{next} is the weight of edge that connects the previous output and the next layer's input; *Num* is the size of the previous layer's output; o_n is the output value for one cell; b^{next} is the bias and $f(\bullet)$ is the activation function.

The SLAP not only allows for different weight matrices and biases in ECG data, but improves the capacity of time series feature extraction. Table 3 shows the parameter definitions for training an SLAP. The length of ECG series received by input layer is 400 (means each ECG series is 400 unit time), and the output of predicted series is 5. There are two hidden LSTM layers in the SLAP, and 64 cells in each layer. The learning rate of the model is 10e−4, the error of termination training is 10e−6, and the beta coefficient for calculating F score is set to 0.15.

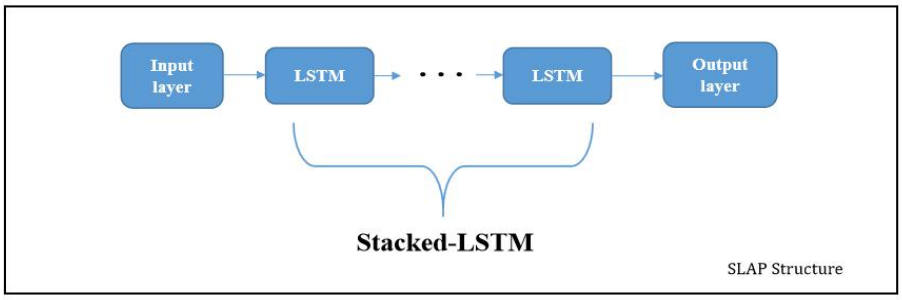


Figure 5. Stacked-LSTM architecture for prediction.

Tabel 3. Parameter presupposition for SLAP

Parameters	Value	Parameters	Value
Input size	400	Training times	873
Cells number	128	Learning rate	10e-4
Layers	2	Stopping value	10e-6
Output size	5	Beta coefficient	0.15

2.7. Process Details

We used RTX2080Ti GPU to complete the training and testing, which took nearly 30 hours for training and validation (parameter tuning and model changing), and approximately 45min for testing.We set the number of training cycles as 240, and each training was conducted 200 samples in batches. The early stop operation was conducted if the loss in 15 training cycles does not drop, and the memory on GPU released if the best model was not stored during the implement of early stop mechanism.

2.7.1. Loss and regularization

BCEwithlogitsloss function is used for multi-classification tasks,For the training set $D = \{(x_i, y_i)\}_{i=1}^{m}$, the calculation method is:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

$$\sigma = Sigmoid()\tag{9}$$

$$l_i = -w_i[y_i \cdot \log(\sigma(x_i)) + (1-y_i)\cdot \log(1-\sigma(x_i))]\tag{10}$$

$$L = -\frac{1}{m}\sum_{i=1}^{m} l_i\tag{11}$$

We set the dropout function (0.5) and add L2 loss regularization (0.04) term to the loss function, as shown in the following formula:

$$L_{regularization} = L + \frac{\lambda}{2}\cdot\frac{1}{m}\sum_{i=1}^{m}\theta_i^2\tag{12}$$

2.7.2. Optimizer and Learning rate

Adam Optimizer was adopted as our learner to prevent the loss of important features due to fast learning. We set the initial learning rate as 1e-3 and set the learning rate to decrease in a certain epoch, the formula is:

$$LearingRat\,e_n = LearingRat\,e_{n-1} * 0.9^{[(Epoch/5)]}\tag{13}$$

2.7.3. Fusion methods

In order to synthesize and reduce the errors of each learner as much as possible, the results are obtained by assessing the effectiveness of each model, when each model works best and relatively has the fewest errors, adding an appropriate weight vector. We found that $W = [0.792, 0.208]$ to achieve the best results:

$$EnsembleOut = \sum_{i}^{m=2}\sum_{j}^{l=2} Model_i(x^i{}_n)\cdot W_j\tag{14}$$

2.7.4. Cross-validation

Different distribution of datasets may result in other effects, which may affect the evaluation of the actual model effectiveness. Therefore, we divided the original data set into K-equal parts with entirely different data, extracted each of which in turn as the verification set $E_n \in \{E_1, E_2, ..., E_k\}$, and the corresponding rest as the training set. The model is applied to conduct k times training and once validation to evaluate the overall effect of this system. In this paper, we set k=5 while the proportion of each data type is uniformly distributed; the average value is taken as the model verification effect index. This method is called K-fold cross-validation (Figure 7).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

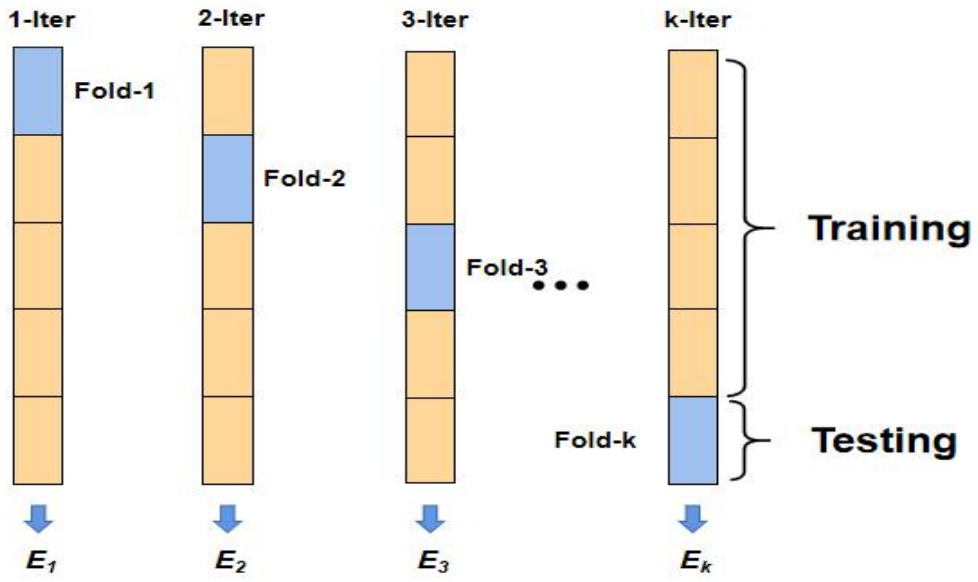


Figure 6. K-fold cross validation.

3.Result

3.1. Data cleaning and distribution.

To ensure that heartbeats are sinus rhythm, the records 102,104,107,217 are eliminated[20]. We divide the datasets into three categories: Training data, validating data, and Testing data. Table 4 summarizes the distribution of data sets in the experiment.

Table 4. The distribution of different datasets.

	MIT-BIH Database	Tianchi (12-Lead)	Amount
Train set	78836	-	78836
Validation set	19709	-	19709
Test set	10949	8265	19214

3.2. Evaluation metrics

We use five indicators commonly used for multi-classification tasks to evaluate the effectiveness of models: Positive Predictive value(PPv%), Accuracy(Acc%), Specificity (Spec%), Sensitivity (Sens%), and F1-score(F1%). Accuracy is used to evaluate the proportion of incidents classified as correct. A positive predictive value indicates the percentage of events that are actually true when the predictive method estimates them as positive. F1-score is used for comprehensive evaluation of PPV and Sens, and is defined as follows：

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

$$Spec = \frac{TN}{TN + FP} \tag{16}$$

$$Sens = \frac{TP}{TP + FN} \tag{17}$$

$$PPv = \frac{TP}{TP + FP} \tag{18}$$

$$F1 = \frac{2 \times PPv \times Sens}{PPv + Sens} \tag{19}$$

where,

--True Positive(TP): Correctly detected as positive.

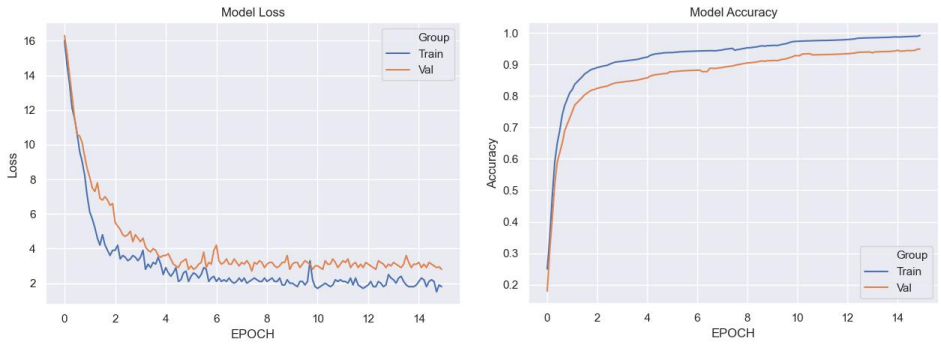
--True Negative(TN): Correctly detected as negative.

--False Positive(FP): Incorrectly detected as positive.

--False Negative(FN): Incorrectly detected as negative.

3.3. Training process

The training situation is shown in Figure 7. The convergence rate is fast, and the loss curve decreases steadily in the training sets means the optimizer keeps updating the parameters by the reverse transmission of neurons. The accuracy curve rises steadily, indicating that the training process of this model is stable and the system precision keeps improving.



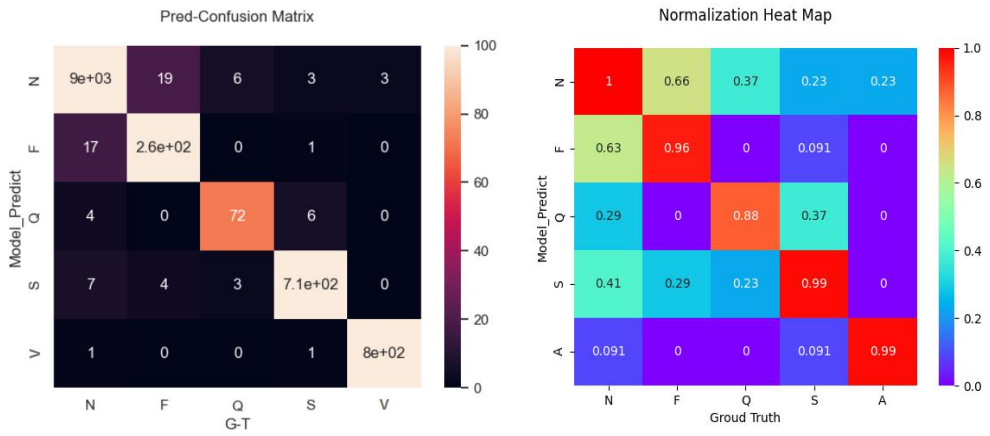
(a) **(b)**

Figure 7. Training and validating process of proposed model:(a)Loss, (b)Accuracy.

3.4. The performance in test sets

The training data in this experiment were divided into 72%, 18%, and 10% for the training, validation, and testing, respectively. A 5-fold cross-validation was performed. The final effect of the model was predicted and evaluated in 10% of the MIT-BIH arrhythmia test set. We set the maximum value of color to 100-sample for better visualization; therefore, the cells over 100 samples will be treated as the brightest color (Figure8(a)). The normalization heat map is shown in Figure 8(b).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2



(a) (b)

Figure 8. Prediction performance of proposed model in the MIT-BIH arrhythmia test set:(a)Confusion Matrix, (b)Normalization Heat Map

We recognize the model's generalization ability through the method of model testing. The proposed model was used to test its effect on different data sets in this work. Table 5 shows that the F1 value reached 99.0% and 95.9% in the two datasets. N-label performed 99.3% specificity and 99.2% sensitivity in the MIT-BIH data set, with 97.6% average accuracy in the Tianchi dataset. Compared with other labels, the main reason for the high performance of N-label is the imbalanced label distribution in the original training samples, the sufficient number of N-label leads to better test results. The testing effect of the MIT-BIH Database is better than that of the Tianchi Dataset because the training process extracts more features from the MIT-BIH database. On the contrary, the Tianchi dataset is just used for testing, and its features are not adequately removed, so the better effect is more inclined to the MIT-BIH database.

Table 5. The test results for five tags in different test sets.

Data	**MIT-BIH Database** (10949 samples)					**TianChi Dataset** (8265 samples)				
	N	**F**	**Q**	**SVEB**	**VEB**	**N**	**F**	**Q**	**SVEB**	**VEB**
Acc	99.3%	**99.6%**	98.8%	99.1%	98.3%	**97.6%**	97.0%	91.9%	97.3%	96.6%
PPv	**98.8%**	97.5%	94.9%	98.6%	96.7%	**95.2%**	93.5%	90.7%	94.8%	92.7%
Sens	**99.2%**	96.8%	98.0%	98.9%	97.5%	96.3%	**96.7%**	87.8%	96.1%	95.7%
Spec	**99.3%**	98.1%	96.7%	97.3%	98.3%	**96.1%**	95.1%	78.1%	95.2%	93.3%
F1	**99.0%**	97.3%	96.4%	98.8%	97.2%	95.7%	95.1%	89.5%	**95.9%**	94.2%

Acc:accuracy; PPv:positive predictive value;Sens:sensitivity;Spec:specificity;F1:f1-score.

The highest average result of each classifier is highlighted in bold.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

4. Discussion

In other studies, they merely train or test the electrocardiograms that had already been segmented, and doctors can not directly use such results for lack of practicality. This work proposed a function that reads images according to frames and automatically divides them into single cardiac beats, which can help doctors to interpret each cardiac rhythm more effectively. Bypassing electrocardiograph records into the ECG pre-processing system, each picture was read at the speed of two frames per second with 400 signal points as the field of vision. Meanwhile, the serial number of the heartbeat was recorded. The voltage sequence corresponding to the heartbeat can be obtained by serial number, automatically input into the ensemble learning system for training. Finally, the ECG classification system can assist doctors by automatically diagnosing the type of heartbeat and presenting different kinds of heartbeat in different colors(Figure 9).

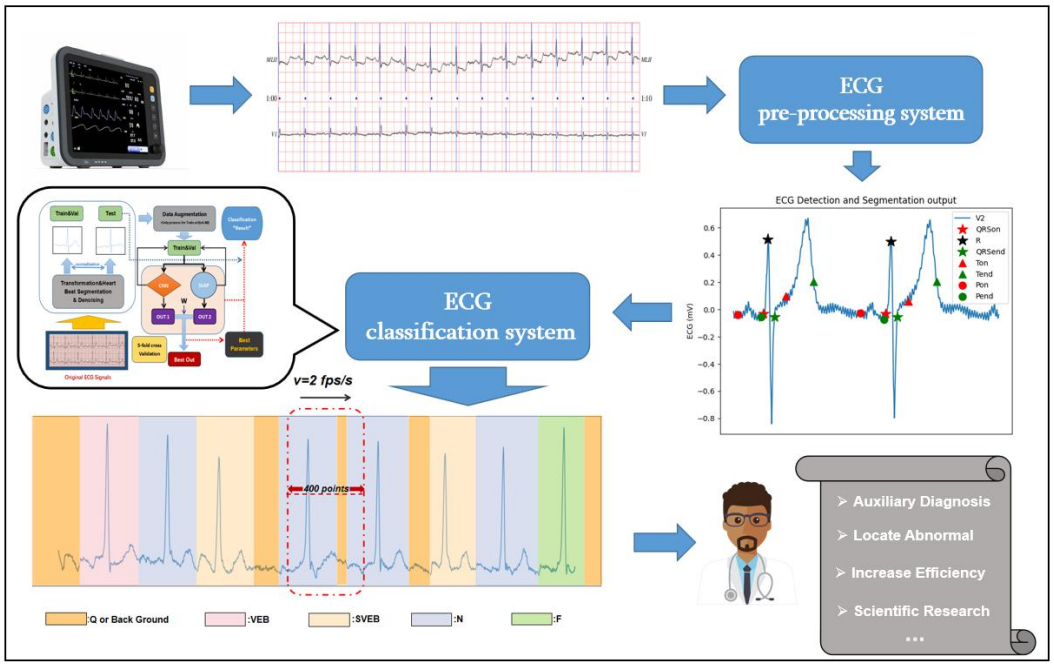


Figure 9. Single cardiac beats auxiliary diagnosis system.

Transfer learning and different backbone collocation with or without data augmented were applied to train the system (adjust input/output dimensions and initialize the parameters randomly). The results of different models are shown in Table 6. We conduct ten training repetitions on the model due to random error subject to a normal distribution in the training process. Record the value of each training performance, using the mean value and the corresponding standard deviation to represent the result.

The proposed model achieves better performance than other single classifiers with 99.13% average accuracy, 97.64% specificity, 97.82% sensitivity, and 97.41% F1-value. The results of experiments with data enhancement were generally better than those without data enhancement. Except for our model, the Resnet model gets the highest 98.33% average accuracy, 96.87% specificity, and 97.68% sensitivity when the learning rate is 9e-5, suggesting that the model can reduce the rate of missed diagnosis and misdiagnosis. Vision transform(VIT),[31] as a novel experimental method, was applied to ECG and achieved 97.48% average accuracy. The model considers the location information of ECG and is divided into nine patches, which are classified by the MLP (multilayer perceptron) classifier. Generally, electrocardiographs segmented by time are interpretable. The limitation of VIT is the uninterpretability of dividing ECG into patches without continuity. In ensemble learning, it is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 May 2022 doi:10.20944/preprints202204.0241.v2

necessary to specify the combination strategy by considering the error rate of each model. F1 value is one of the comprehensive evaluation indexes for systematically improving the model's performance. LSTM model has the best 96.72% PPV, which plays an essential role in enhancing the model's total index F1 value.

Table 6. Effects experiment of different classifiers.

Backbone	Augment	LR	Acc(%)	Spec(%)	Sens(%)	PPv(%)	F1(%)
DenseNet[32]	-	2e-4	96.81±1.17	95.27±0.93	97.15±0.81	93.27±1.26	95.56±0.37[a]
	+	1e-4	97.23±0.96	95.81±1.02	96.74±0.45	91.47±0.68	94.13±0.62[bc]
ResNet	-	4e-4	98.26±0.23	96.58±0.51	97.32±0.73	94.35±0.34	95.84±0.18
	+	9e-5	98.33±0.63	96.87±0.43	97.68±0.56	95.61±0.78	96.78±0.31[c]
VGGNet[33]	-	1e-3	94.72±1.44	87.61±1.83	94.89±0.76	91.37±0.74	93.23±1.16[a]
	+	3e-4	95.81±1.21	89.23±1.78	94.20±1.05	91.66±0.97	92.94±1.24[b]
LSTM	-	4e-4	95.47±1.62	93.24±1.43	93.26±1.07	95.68±0.89	94.42±0.97[a]
	+	7e-5	96.03±1.16	94.11±0.93	91.32±1.39	**96.72±0.64**	93.77±0.75[b]
VIT	-	1e-4	96.97±0.36	95.26±0.63	97.08±0.22	94.88±0.51	95.94±0.41
	+	5e-4	97.48±0.83	95.33±0.72	97.63±0.41	95.11±0.44	96.40±0.57[b]
Proposed	-	3e-5	98.86±0.31	**97.64±0.52**	97.73±0.34	95.41±0.46	96.86±0.37
	+	2e-5	**99.13±0.19**	97.12±0.33	**97.82±0.67**	96.59±0.29	**97.41±0.34**[c]

LR:The best learning rate in the model, '+'or '-':With or without augmentation, VIT:Vision transformer.[31]

a:F1(This model without augmented)compared with F1(Proposed model without augmented),P<0.05.

b:F1(This model with augmented)compared with F1(Proposed model with augmented),P<0.05.

c:F1(This model with augmented)compared with F1(This model without augmented),P<0.05.

Moreover, we have compared our proposed model with some existing models evaluated over the same datasets. We adopted the weighted average method to assess the effects of various studies by calculating the weights assigned to tasks with different classification categories—the impact of additional studies as shown in Table 7. In the ECG detection and diagnosis tasks, some deep learning studies based on convolutional neural network achieves better results than machine learning. The ensemble learning method proposed by us based on deep learning has higher accuracy than two other multi-mode fusion of ECG sequence and CNN-LSTM series studies. Generally, the classification accuracy will decrease with the number of classifications, whereas the k-nearest Neighbor method, with better performance in machine learning, achieves 97.00% average accuracy in 17 categories. The reason is that the samples quality of this experiment is good, and the distribution of different labels is balanced so that the classifier pays equal attention to each category and has a better generalization ability during testing. Multi-modal fusion adopts two test sets, and its generalization effect is good, with 98.60% average accuracy. Our proposed system shows the best result in the studies we listed, with a 99.13% average accuracy close to 100%.

Table 7. Performance comparison with other studies.

Studies*	DL/ML	Test samples	Accuracy
[§]Random Forest / Gradient Boost Tree[34]	Machine Learning	32168	96.75%[a] / 97.98%

§PSO-Support Vector Machines[35]	Machine Learning	40438+-	94.16%a
ΦMulti-modal Fusion[14]	Deep Learning	21892+2911	98.60%
ΦCNN - LSTM[11]	Deep Learning	21524+-	99.01%
ΦRRHOS - LSTM[37]	Deep Learning	49603+-	95.81%a
§K-Nearest Neighbor[36]	Machine Learning	**109702**	97.00%a
Our Proposed	Deep Learning	10949+8265	**99.13%**

§:Study was published earlier than 2015; ф:Study was published later than 2015.

+-:The distribution of positive and negative samples is imbalanced; DL:deep learning; ML:machine learning.

a:Average accuracy(This model) compared with average accuracy(Proposed model),P<0.05.

5. Conclusion

An auxiliary diagnosis of clinical electrocardiogram was studied in this paper. ECG intelligent detection and classification tasks are carried out using the integration method combined with a strategy based on the deep learning of recurrent neural networks (Stacked-LSTM) and convolutional neural network(1-dimensional ResNet). The innovation lies in combining the method of artificial intelligence with the traditional diagnostic way. The main feature of this study is that the system can realize the segmentation of heartbeat visualization and heartbeat classification by reading the ECG voltage sequence into our system. To some extent, it can assist doctors in improving their diagnostic efficiency and effectively reducing the rate of missed diagnosis and misdiagnosis, which has practical application value in the medical field. In addition, the prediction of models in other studies is generally completed in the MIT-BIH dataset. In contrast, the Tianchi dataset was introduced in our research as a new test set, which fully confirmed the generalization ability of these proposed models.

In the follow-up study, on the premise of improving the model classification categories and model generalization ability, further enhancing the prediction accuracy and practicability of this system to better assist doctors in electrocardiograph auxiliary diagnosis, which helps physicians reduce their diagnostic workload and improve the diagnostic efficiency while reducing missed diagnosis and misdiagnosis.

Author Contributions: Conceptualization, S.J., X.G., H.G., and J.W.; methodology, S.J. and H.G.; software, S.J.; validation, H.G.,X.G.; formal analysis, S.J. and H.G.; resources, J.W.; data curation, S.J.; writing—original draft preparation, S.J.; writing—review and editing, S.J. and J.W.; visualization, H.G.; supervision, J.W.; project administration, J.W.,X.G.; funding acquisition, J.W.; 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 62176231 and 62106218), the Natural Science Foundation of Zhejiang(LGF20F020013), and the National Key Research and Development Program of China (2019YFB1404802).

Data Availability Statement: The relevant codes on the website: http://github.com/Jiangsiqing/1-d-ecg-diagnose.

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

References

1. Bansilal S, Castellano JM, Fuster V. Global burden of CVD: focus on secondary prevention of cardiovascular disease. *Int J Cardiol.* 2015 Dec; 201 Suppl 1: S1-7.
2. World Health Organization (2019).Cardiovascular diseases (CVDs). http://www. who.int/mediacentre/factsheets/fs317/en/ Accessed 11 June 2021
3. World Health Organization (2018).Ageing and health.https://www.who.int/news-room/fact-sheets/detail/ageing-and-health Accessed 4 Oct. 2021
4. Bhatnagar P, Wickramasinghe K, Wilkins E.Trends in the epidemiology of cardiovascular disease in the UK. *Heart* 2016 Dec 15; 102(24): 1945-1952.
5. Wolf MM, Varigos GA, Hunt D. Sinus arrhythmia in acute myocardial infarction. *Med J Aust* 1978 Jul 15; 2(2): 52-3.
6. Bertrand PB, Levine RA, Isselbacher EM. Fact or Artifact in Two-Dimensional Echocardiography: Avoiding Misdiagnosis and Missed Diagnosis. *J Am Soc Echocardiogr* 2016 May; 29(5): 381-91.
7. He K , Zhang X , Ren S. Deep Residual Learning for Image Recognition[J]. *IEEE* 2016.
8. Sun L , Wang Y , He J. A stacked LSTM for atrial fibrillation prediction based on multivariate ECGs[J]. *Health Information Science and Systems* 2020; 8(1): 1-7.
9. Hinton G E , Salakhutdinov R R . Reducing the Dimensionality of Data with Neural Networks[J]. *Science* 313.
10. Jun T J , Nguyen H M , Kang D , et al. ECG arrhythmia classification using a 2-D convolutional neural network[J]. 2018.
11. Zheng Z , Chen Z , Hu F , et al. An Automatic Diagnosis of Arrhythmias Using a Combination of CNN and LSTM Technology[J]. *Electronics* 2020; 9(1): 121.
12. Niu L, Chen C, Liu H, Zhou S, Shu M. A Deep-Learning Approach to ECG Classification Based on Adversarial Domain Adaptation. *Healthcare (Basel)* 2020; Oct 27; 8(4): 437.
13. Zhang Q , Zhao Z , Zhang X , et al. Conditional Adversarial Domain Generalization with A Single Discriminator for Bearing Fault Diagnosis[J]. *IEEE Transactions on Instrumentation and Measurement* 2021; PP(99): 1-1.
14. Ahmad Z , Tabassum A , Guan L , et al. ECG Heartbeat Classification Using Multimodal Fusion[J]. *IEEE Access* 2021; 9: 100615-100626.
15. Emanet N . ECG beat classification by using discrete wavelet transform and Random Forest algorithm[C] *International Conference on Soft Computing. IEEE* 2009.
16. Urtnasan E., Park J.U., Joo E.Y., et al. Automated Detection of Obstructive Sleep Apnea Events from a Single-Lead Electrocardiogram Using a Convolutional Neural Network. *J. Med. Syst* 2018; 42: 104.
17. Moody G B , Mark R G , Goldberger A L . PhysioNet: a Web-based resource for the study of physiologic signals[J]. *Engineering in Medicine & Biology Magazine IEEE* 2001; 20(3): 70-75.
18. Alibaba Cloud.Tianchi Competition,ECG and Human machine Intelligence Competition.https://tianchi.aliyun.com/competition/entrance/231754/information?lang=en-us(2019, accessed 19 November 2019).

19. Association for the Advancement of Medical Instrumentation; American National Standards Institute. Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms; AAMI: Arlington, VA, USA, 2012.
20. Sujit S J , Coronado I , Kamali A , et al. Automated image quality evaluation of structural brain MRI using an ensemble of deep learning networks[J]. *Journal of Magnetic Resonance Imaging* 2019.
21. YE C， KUMAR B V K V， COIMB R A M T, et al. Heartbeat classification using morphological and dynamic features of ECG signals [J]. *IEEE Transactions on Biomedical Engi- neering* 2012; 59(10) : 2930-2941.
22. Barbara N , Camilleri T A , Camilleri K P. A comparison of EOG baseline drift mitigation techniques[J]. *Biomedical Signal Processing and Control* 2020; 57: 101738.
23. Lian-Gang L U . Filter circuits design for filtering 50 Hz power frequency interference[J]. *Journal of Liaoning Teachers College(Natural Science Edition)* 2012.
24. Narayana K , Rao A B . Wavelet based QRS detection in ECG using MATLAB[J]. *Innovative Systems Design & Engineering* 2011; 2(7).
25. Lee G , Gommers R , Waselewski F , et al. PyWavelets: A Python package for wavelet analysis[J]. *The Journal of Open Source Software* 2019; 4(36): 1237.
26. V. Vijendra, M. Kulkarni. "ECG signal filtering using DWT haar wavelets coefficient techniques," 2016 International Conference on Emerging Trends in Engineering, *Technology and Science (ICETETS)* 2016; pp.1-6.
27. L. Hargrove, P. Zhou, K. Englehart. "The Effect of ECG Interference on Pattern-Recognition-Based Myoelectric Control for Targeted Muscle Reinnervated Patients," *IEEE Transactions on Biomedical Engineering* vol. 56, no. 9, pp. 2197-2201, Sept. 2009.
28. Dietterich T G . Ensemble Methods in Machine Learning[J]. *proc international workshgp on multiple classifier systems* 2000.
29. He K , Zhang X , Ren S , et al. Identity Mappings in Deep Residual Networks[C]// *Springer, Cham. Springer, Cham* 2016.
30. Hochreiter, Sepp, Schmidhuber, et al. Long short-term memory.[J]. *Neural Computation* 1997.
31. Dosovitskiy A , Beyer L , Kolesnikov A , et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J]. 2020.
32. Huang G , Liu Z , Laurens V , et al. Densely Connected Convolutional Networks[C]// *IEEE Computer Society. IEEE Computer Society* 2016.
33. Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. *Computer Science* 2014.
34. Melgani F, Bazi Y (2008). Classification of electrocardiogram signals with support vector machines and particle swarm optimization. *IEEE transactions on information technology in biomedicine* 12(5): 667-677
35. D Kopiec, Martyna J . A Hybrid Approach for ECG Classification Based on Particle Swarm Optimization and Support Vector Machine[C]// *International Conference on Hybrid Artificial Intelligence Systems. Springer, Berlin, Heidelberg* 2011.

 doi:10.20944/preprints202204.0241.v2

36. Park J, Lee K, Kang K. (2013). Arrhythmia detection from heartbeat using k-nearest neighbor classifier. *IEEE International Conference on Bioinformatics and Biomedicine* pp.15-22
37. E. Essa and X. Xie, "An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification," *IEEE Access* vol. 9, pp.103452-103464, 2021.