A Multi-Task Cardiovascular Disease Classification Method via Adaptive Update

Renjie Tang, Junzhou Luo and Junbo Qian

June 23, 2021

A Multi-task Cardiovascular Disease Classification Method via Adaptive Update

Renjie Tang
[1st]School of Computer science and engineering, Southeast University,
Nanjing, Jiangsu Province, China
[2nd]China Mobile Group Zhejiang Co., Ltd.
Hangzhou, Zhejiang Province, China
Corresponding author: Tangrj@seu.edu.cn

Junzhou Luo
School of Computer science and engineering, Southeast University,
Nanjing, Jiangsu Province, China
jluo@seu.edu.cn

Junbo Qian
China Mobile Group Zhejiang Co., Ltd.
Hangzhou, Zhejiang Province, China
qianjb@zj.chinamobile.com

***Abstract*—**Federate learning is a generic method to augment medical data by training a global model with distributed data located in regional medical environment. Nevertheless, there exists a significant distinction with respect to the distribution of medical cases between primary hospitals and tertiary hospitals, which lead to the state that the global model is incapable of satisfying individual medical needs. We propose a novel multi-task cardiovascular disease classification method based on adaptive update of sample masking and the adaptive optimization of weights. First, a mask is designed to fit local data, and it can guide the weight updated with accordance to hard samples. Then, the global model is focused on specific task with adaptive optimization of weights. Experimental evaluation on our collected data show great improvement. The F1 metric increases from 49.54% to 75.5% by adopting the proposed method.

Keywords*—Personalized model, Federated learning, Adaptive update*

I. Introduction

Cardiovascular disease is one of the major diseases threatening human life and health. ECG analysis is the widely used method for cardiovascular disease classification. Traditional methods generally depend on human detection, and their accurary highly depends on the doctor's efficiency, work intensity, and professional level. Compared with manual diagnosis, intelligent diagnosis of cardiovascular diseases has natural advantages. However, it is difficult for primary medical institutions to train a cardiovascular disease classification model that meets the needs of the real medical environment due to data scale and other reasons in practical work.

In order to solve the problem of ECG data scarcity, the generic way is to augment medical data with federated learning framework, which can train a global model with data located on distributed regional node. Then each local hospital downloads the global model for private ECG classification. Unfortunately, after evaluation, it is found that the global model has a low classification accuracy rate in some specialty local hospitals. The classification rate is high when the ECG data sample distribution of a medical institution is close to the training data of the global model. On the contrary, model performance descreases drasticlly when local data is different from global data. When the global model is updated based on a small amount of unbalanced data, the evaluation performance of the model is not the optimal result. One of the important factors is that data distribution may quite different between local and global, and this makes it difficult for a unified global model to meet the individual needs of characteristic hospitals. Therefore, it is necessary to design a general model to solve the imbalance problem. In this paper, we propose a novel cardiovascular disease multi-task classification method base on adaptive update. Adaptive update of sample masks and adaptive weights optimization are used to improve the model classification ability for both global and local data.

The organization of this article is as follows: Chapter 2 introduces the related works. Chapter 3 gives the multi-task classification method for cardiovascular diseases based on adaptive update. Experimental results are given in Chapter 4. Chapter 5 summarizes this paper.

II. Related Work

Automatic processing and diagnosis methods of ECG signals have made rapid progresses. Ref.[1] designed filters for ECG signal preprocessing including low-pass, high-pass, and time-space domain. A discrete cosine transformation was proposed in ref.[2]. Ref.[3] utilized discrete wavelets and the non-local mean to filter and reconstruct ECG signal. Wavelet transform was used to locate the main wave position in ECG, and locatef the position of R waveform[4]. In addition to traditional methods, some improved techniques have been proposed, such as the analysis based on the ordinary spatial spectrum mode[5], etc. However, these ways still have many missed and false detections in clinical ECG signal analysis.

Machine learning and statistic based methods are also widely used in ECG analysis, such as bayes' theorem, k-nearest neighbor, decision tree and linear discriminator[6-8], etc. Ref.[9] used fuzzy neural networks for extracting features including the period in ECG and the high-order quantities of the QRS wave. Ref.[10] leveraged multi-layer perceptrons by adopting features with regard to the wavelet transform. Ref.[11-13] proposed a method that classify bioelectric signals via support vector machines, least squares support vector machines and twin support vector machines. These methods are data-driven with clear and efficient processes, but are limited by feature extraction and expression capabilities.

As the explosive increase of data and computing power, deep learning has gradually become the mainstream method. It can automatically represent the complex features of ECG signals with non-linear fitting, and then the performance can be greatly improved. Ref.[17] proposed a convolutional neural network for multi-lead ECG classification. Deep belief network was used to diagnosis of ECG arrhythmia in ref.[14-16]. Adversarial networks were used for data enhancement in ref.[18]. High error recognition rate associated with the high accuracy was discussed in ref.[19], and CNNs model was corrected. Depth factorization was used to deal with the complex noise in ECG[20]. Ref.[21] used autoencoder model for ECG feature extraction. Graph regularization non-negative matrix decomposition was used for dimension reduction, and sparse representation is used for feature representation[22]. However, most of the studies are based on public data sets and the sample size is small. It is not possible to deal diversity, variability, and randomness of clinical data.

The unbalanced distribution of ECG data is one of the most challenge problems. It has severely affected the performance of the global model in local inference. Generally, reducing the impact of data imbalance by changing the distribution of samples in the data procesing, such as over sampling, under sampling and mixed sampling, etc. Ref.[23] proposed the SMOTE algorithm that synthesizes minority samples through heuristic strategies. Ref.[24] proposed the Borderline-SMOTE algorithm, which leveraged the SMOTE algorithm to synthesize new samples for minority at the border. Ref.[25] proposed ADASYN, which is based on the adaptive synthetic sampling of imbalanced data. A cluster-based over-sampling algorithm that combines clustering algorithm and over-sampling technology was proposed in ref.[26]. It ensured synthetic sample is always located in the minority area. Ref.[27] combined SMOTE with TomekLink to avoid noise data. Oversampling was adopted for the minority class of boundary samples in ref.[28]. It can satisfy the balance of training samples after processing noise data and redundant data in majority class. Ref.[29] studied the best class distribution from the perspective of data sets. Ref.[30] used optimization of KL divergence for post-training priori and rebalancing. Ref.[31] proposed a novel focal loss, which focused on hard samples during training. However, case distribution of each medical node actually reflects the medical health characteristics of the population in a region, and it cannot be simply adjusted at the data level or loss function. ECG data has special biological characteristics, and it is very difficult to up-sampling to construct samples. Therefore, we proposes a personalized update method of cardiovascular disease classification model based on mixed loss. Through the adaptive weight update of sample masking and loss, we can explore the personalized update of the global model in private hospitals.

III. Multi-task Classification Method For Cardiovascular Diseases

Equation 1 gives the private data D_p, which is collected in a hospital node.

$$Dp = \{Dp_i \,|\, 1 < i < N\},\ \ Dp_i = \{x_i, [y_i^0, y_i^1, \cdots y_i^T]\} \tag{1}$$

D_{pi} is the i-th data tuple. x_i represents the i-th ECG data, and $[y_i^0, y_i^1, ..., y_i^T]$ indicates the type label. It is supposed that there are T tasks in total. In this paper, a multi-task classification method is proposed based on adaptive update. Figure 1 gives the framework of the proposed method. The backbone network for ECG feature extraction is shared. Then a mask is assigned to each task, and a weight adaptive adjustment is conducted with final loss function.

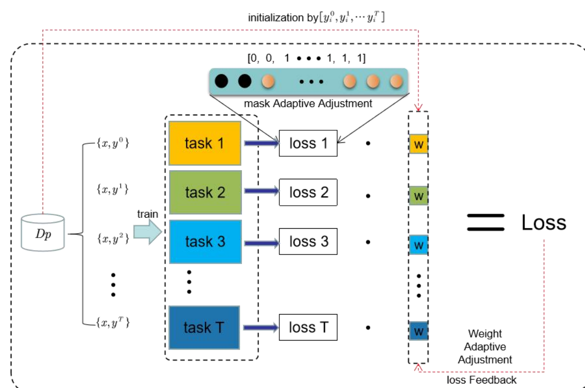

FIGURE I. Multi-task classification method for cardiovascular diseases based on adptive update

However, there is a serious imbalance between the types of ECG symptoms in different local node, which leads to huge discrepancies in local data evaluation with the same global model. Data types of global node could be rare in local node, so the local training data may fail to meet the needs of model updating. The proposed method has two main components: adaptive sample masking weight update and adaptive loss cost weight update. The adaptive sample masking weight update generates adaptive sample masks via the local data, diagnostic models and dynamically adjusted thresholds. After the model iteration, the mask will be adaptively updated at the same time to ensure that it is on a single task. The model adaptively select samples that are difficult to infer for update accurately. The adaptive loss cost weight update the model's loss weights by leveraging the local case distribution. At the same time, the model will adaptively adjust the loss weights of multiple tasks according to the loss changes of different tasks to ensure that the model can focus on tasks with weaker classification capabilities. Details are described as follows:

A. The Adaptive Sample Masking Update

The global model can be trained well with federate learning framework across distributed data. For model personalization, we design a mask for local samples. Easy samples are masked, and samples that are not recognized by the global model are selected to update the personalized model.

Let *y* represents the true lable of an ECG data sample, *y'* represents the predicted value of the model that data. *gap* is the absolute value of the difference between the predicted value and the true label, as shown in equation 2.

$$gap = \{ |y_i - y_i'| \mid 1 < i < N \} \tag{2}$$

When the global model is iteratively updated locally, it is necessary to calculate the difference between the predicted value and the true label *gap*. When *gap* is lower than the dynamic threshold, it indicates that the model can accurately predict the sample. We generate a mask that represents the predictive ability of the model based on *gap* and the dynamic threshold. The *mask* can be computed as equation 3.

$$mask = \{ (gap_i > th, 1, 0) \mid 1 < i < N \} \tag{3}$$

The dynamic threshold at the beginning of training should not be large because the accuracy of the model is poor. In the subsequent iterative training, the dynamic threshold is gradually increased to a higher value. In the final tuning stage, the model will only focus on samples that cannot be accurately predicted. Figure 2 shows the trend of the dynamic threshold versus the number of iterations. The abscissa represents iteration rounds during the model updating, and the ordinate represents the threshold value.

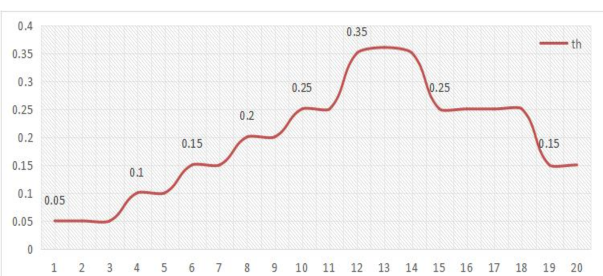

FIGURE II. Trend Of Dynamic Threshold With The Number Of Iterations

The dynamic threshold is set to 0.05 at the beginning of the iteration, then the threshold will be gradually increased to 0.35 in subsequent iterations, and it will fall back to 0.15 finally. We found that if the model pay too much attention to extremely hard samples, it may lead to a decline in the overall evaluation of the model in later stages of iterative training. Therefore, the dynamic threshold needs to return to a relatively low value.

The basic loss function we used is focal loss, which is improved on the basis of standard cross-entropy loss. Let p_t represents the probability that the sample belongs to the positive sample. In order to distinguish the difficulty of the prediction accuracy of the sample, a small weight is added. Increasing the adjustment factor $(1\text{-}p_t)^r$. The representation of focal loss is given in equation 4:

$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{4}$$

When the model is iteratively trained, the range of the dynamic mask is adjusted through the dynamic threshold, and the appropriate sample is selected to update the model using focal loss. The final *maskloss* is given as equation 5:

$$mask\,loss = mask \cdot \mathrm{FL}(p_t) \tag{5}$$

As shown in Figure 3, the abscissa is p_t, and the ordinate is loss value. It shows the change curve of the loss value of the standard cross entropy formula CE and the mask loss according to the p_t. For simply sample, the p_t is very high. The model will adjust the range of the mask through the dynamic threshold, and more suitable samples to update weights can be selected.

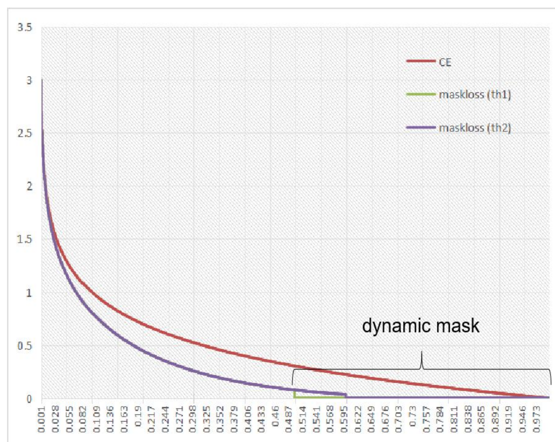

FIGURE III. Loss Function Graphs Under Different Thresholds Adaptive Sample Mask Weight Update Method

With the adaptive sample masking update, model performance for a locally single task can be improved.

B. Multi-task Classification With Self-optimized Weights

The adaptive weight update of sample masking solves the problem of unbalanced samples. It can adjust the loss weights of different classification tasks when the model updates weights of the task locally. In this subsection, we propose an adaptive loss cost weight update mechanism, which dynamically adjusts the loss weights of different tasks to make the model focus tasks with large differences in sample cases.

$$loss_k = w_k \cdot mask\ loss \tag{6}$$

The loss function is shown as equation 6. w_k is the loss weight corresponding to the k-th task in the multi-task learning mechanism. It is an adaptive weight coefficient.

The initial loss weights is set as equation 7. N is the total number of local samples. N_k is the number of local samples in the k-th category. w_L is the threshold of loss weights, which ensures that some symptoms do not cause the initial weight assigned to be too large because the sample ratio is too low. w_k^0 is the initial weight of the k-th task:

$$w_k^0 = \min\left(\log_{10}\left(\frac{N}{N_t}\right), w_L\right) \tag{7}$$

The loss function can be adjusted adaptively according to different distribution of samples in multiple nodes. When the number of samples for this task is large, the model will tend to neglect it in the iterative process, so the weight coefficient is set to be small. When the number of samples for this task is small, it's easier to identify inaccurately in the iterative process of the model. Therefore, Therefore, increasing the loss weight of the task in the loss function. The model increases the penalty for misjudgment of the task, which makes the model more bias towards tasks with fewer samples.

In the subsequent iterative training, the loss weight of the task will be adaptively fine-tuned according to the loss changes of different tasks. The dynamic loss weight of the multi-task learning mechanism is dynamically adjusted. The details are as follows. $loss_k^i$ is the loss of the i-th iteration of the k-th task, and $loss_k^{i-1}$ is the loss of the (i-1)th iteration of the k-th task. $diff_k^i$ is the difference of the k-th task in loss change between the adjacent iteration, which indicates the training difficulty of the current task, as shown in equation 8.

$$diff_k^i = \mathrm{loss}_k^i - \mathrm{loss}_k^{i-1} \tag{8}$$

As shown in equation 9, d_k^i is the adjustment rate of the dynamic loss weight w_k. It is the ratio of the average difference of the loss change of all tasks to the difference of the loss change of the k-th task. When the change of the loss weight of the k-th task becomes smaller than that of other tasks, the multi-task model will appropriately increases the loss weight of the task.

$$d_k^i = \frac{mean\left(\sum_{k=1}^{T} diff_k^i\right) - diff_k^i}{diff_k^i} \tag{9}$$

As shown in equation 10, w_k^i is the dynamic loss weight of the ith iteration, and w_k^{i-1} is the loss weight of the i-1th iteration. *step* is the maximum value of the adjustment step of the regulation coefficient d_k. The task does not benefit from increasing the loss weight significantly because some tasks may have completed convergence earlier in the model iteration. For this reason, it is necessary to constrain the change span of the loss weight of different tasks and prevent the imbalance of the dynamic loss weight of the multi-task learning mechanism.

$$w_k^i = \begin{cases} w_k^{i-1} + min(step, d_k), & d_k > 0 \\ w_k^{i-1} + max(-step, d_k), & d_k \le 0 \end{cases} \tag{10}$$

In summary, the adaptive cost-sensitive modeling loss based on the mixed loss can be expressed as equation 11.

$$Loss = \sum_{k=1}^{T} w_k \cdot \mathrm{loss}_k \tag{11}$$

$loss_k$ represents the loss function of the kth task. T represents the total number of tasks for multi-task classification. k represents the k-th task. w_k is the dynamic loss weight. Through the above loss function, the secondary learning of the misclassified samples of the global model can be realized by combining with the imbalance between multiple tasks for the personalized update of the local model.

IV. EXPERIMENTS

A. Datasets

We collect two clinical datasets from primary medical institutions for evaluation. One comes from a basic children's hospital in Zhejiang province, China, and the other one comes from a basic general hospital in Zhejiang province, China. There are 12 leads of ECG data including 6 leads of the limbs and 6 leads of the chest. Basic unit of ECG is splitted with 10 seconds, and the sampling frequency of the equipment is 500Hz. The sample size of cases of basic general hospitals is 2156 , and the sample size of cases of children's hospitals is 2594. The specific case distribution of data is given in Table 1.

B. Experimental environment

Experiments are conducted on Jiutian deep learning platform provided by China Mobile CMCC. It provides GPU cloud services. CPU adopts E5-2640V4. The memory is 128GB DRAM, 250GB SDD. Graphics card adopts TeslaV100. Standard library CUDA10 is pre-built. The 1.12.0 version of tensorflow and 2.2.4 version of keras are pre-installed.

C. Evaluation criteria

We use stardand accuracy, precision, recall and F1 to evaluate the performance of the proposed method.

$$\begin{aligned} \text{Accuracy } &= = \frac{TP+TN}{TP+TN+FP+FN} \\ \text{Precision } &= \frac{TP}{TP+FP} \\ \text{Recall } &= \frac{TP}{TP+FN} \\ \text{F1 } &= 2 * \frac{\text{Precision*Recall}}{\text{Precision } + \text{ Recall}} \end{aligned} \tag{12}$$

As shown in equation 12, TP indicates the number of positive samples correctly classified. TN indicates the number of negative samples correctly classified. FP indicates the number of negative samples that are misclassified as positive samples. FN indicates the number of positive samples that are misclassified as negative samples.

D. Analysis

The global model will be evaluated on the basic general hospitals with a data set with size 2156. The model will be evaluated in the primary children's hospital data with 2594 samples. The evaluation result of the global model in basic general hospitals and children's hospitals is shown in Table 2.

TABLE I. CASES DISTRIBUTION IN GENERAL HOSPITALS AND CHILDREN'S HOSPITALS

Case	Junior General Hospitals	Junior Children's Hospitals
Poor r wave progression	8	0
ST-T variant	71	4
ST segment variant	115	16
T wave variant	365	50
RAD	20	55
LAD	52	14
Sinus bradycardia	173	14
Sinus tachycardia	115	402
Sinus rhythm	1764	1557
Sinus arrhythmia	72	603
Atrial fibrillation	39	0
Atrioventricular block	39	3
Atrial escape rhythm	1	7
Atrial ectopic beats	57	16
Counterclockwise rotation	15	3
Paroxysmal vertricular tachycardia	11	0
Ventricular premature beats	40	5
Bifascicular block	1	1
Clockwise rotation	3	3
Myocardial infarction	5	0
Ventricular preexcitation	4	5
Abnormal q wave	12	9
Right bundlebranch block	67	12
Normal ecg	1208	1425
Limb leads low voltage	16	0
Left anterior hemiblock	15	0
Left bundle branch block	4	0
Left ventricular hypertrophy	5	1
High voltage of left ventricular	116	24
Limb lead reversal	1	7

TABLE II. THE PERFORMANCE OF THE GLOBAL MODEL IN DIFFERENT MEDICAL INSTITUTIONS

mode	best acc	best recall	best precision	best F1
Primary General Hospital	78.14%	80.36%	76.18%	78.22%
Children's Hospital（before update）	54.32%	41.58%	61.29%	49.54%
Children's Hospital（after update）	80.35%	75%	75.98%	75.50%

The distribution of samples in basic general hospitals are close to that of training data in global model. Its accuracy rate, recall rate, precision rate and F1 can reach 78.14%, 80.36%, 76.18% and 78.22% respectively. The distribution characteristics of cases of children's hospitals are quite different to the distribution of global model, and its accuracy, recall, precision, and F1 can only reach 54.32%, 41.58%, 61.29% and 49.54%. This shows the global model cannot adapt to the special medical environment like children's hospitals directly. The classification performance of the model is significantly reduced. By means of the personalized update of the classification model based on mixed loss. The accuracy rate, recall rate, precision rate, and F1 of children's hospitals rare 80.35%, 75%, 75.98% and 75.5%. The performance is close to the level of basic general hospitals.

E. *Comparative experiment*

We conducted a comparative experiment of two different methods SMOTE and ANASYN, on the primary children's hospital data set to verify the effectiveness of the multi-task learning mechanism. Typical focal loss is used as loss function,. All methods use fixed task weights, and the specific experimental results are given in table 3.

TABLE III. COMPARISON OF MODEL EVALUATION UNDER DIFFERENT PROCESSING MODES

Processing	Avg acc	Avg recall	Avg precision	Avg F1
SMOTE	72.23%	68.21%	70.45%	68.11 %
ANASYN	74.06%	69.55%	71.12%	69.23 %
focal loss + fixed weight	80.54%	71.55%	77.57%	74.44 %
mask loss + dynamic weight	80.35%	75%	75.98%	75.50 %

As shown in Table 3, the proposed loss function method is generally better than the data augment method. The new sample data synthesized by the method of up and down may not actually meet the medical characteristics. The case distribution of each medical node actually reflects the medical health characteristics of the population in a region, so we can not simply make adjustments at the data level. Correspondingly, the focal loss function reduces the weight of a large number of simple negative samples, so that the model can deal with difficult and misclassified samples. The mask loss function sets dynamic thresholds and dynamic task weights, so that the model has different emphasis samples in different training stages, and its final result is improved compared to focal loss.

F. *Ablation experiment*

We conducted ablation experiments of a variety of personalized modes on primary children's hospital dataset to verify the effectiveness of the proposed method. They include: binary crossentropy + fixed weight, focal loss + fixed weight, mask loss + fixed weight, mask loss + dynamic weight. The specific verification method is that: the 2594 samples from the Children’s Hospital are cross-validated with 5 folds. The average of evaluation scores are obtained. The evaluation performance under different modes are shown in Table 4.

The accuracy rate, recall rate, precision rate and F1 of the model can be restored to a better level via personalized updating. And all of them are the optimal values under various models through mask loss and dynamic weight. Compared with the standard two-class cross-entropy, focal loss can focus on samples that are difficult to classify. The adaptive sample mask weight update is dynamically masked based on the focal loss. The model will pay more attention to samples separated difficultly while reducing the impact of samples separated simply on the model. Adaptive loss cost weight update adopts dynamic weight adjustment. The model does not focus on the loss of the classification task of rare cases in children, and pay more attention to classification tasks that are significantly different from the case distribution of the global model. The model performance has been significantly improved.

TABLE IV. COMPARISON OF MODEL EVALUATION UNDER DIFFERENT LOSS MODES

Personalized update mode	Avg acc	Avg recall	Avg precision	Avg F1
Binary crossentropy + fixed weight	80.54 %	71.12%	77.83%	74.32 %
focal loss + fixed weight	80.54 %	71.55%	77.57%	74.44 %
mask loss + fixed weight	79.96 %	75.43%	74.79%	75.11 %
mask loss + dynamic weight	80.35 %	75%	75.98%	75.50 %

V. CONCLUSIONS

We proposes a personalized update method for cardiovascular disease classification. The measure of weight update by adaptive sample masking makes the model focus on samples that the global model cannot accurately predict, so that the model can converge more quickly on local medical data. Through the weight update via the adaptive loss cost, the model is trained with tasks that have major differences between the global and local tasks, so that the model can obtain a better classification effect in the local medical institution. The model is validated on the dataset of primary children's hospitals, which is better than the traditional update method of the model. It has achieved a great improvement in the final classification effect, and its evaluation result is close to the classification level of the model on the data of basic general hospitals.

REFERENCES

[1] Kong K Y, Ng C Y, Ong K. Web-based monitoring of real-time ECG data[C]. Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163) Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163), 2000. IEEE, 2000: 189-192.

[2] Nikolaev N, Gotchev A, Egiazarian K, et al. Suppression of electromyogram interference on the electrocardiogram by transform domain denoising[J]. Medical & Biological Engineering & Computing. 2001, 39(6): 649-655.

[3] Singh P, Pradhan G, S. S. Denoising of ECG signal by non-local estimation of approximation coefficients in DWT[J]. Biocybernetics and Biomedical Engineering. 2017, 37(3): 599-610.

[4] Martinez J P, Almeida R, Olmos S, et al. A Wavelet-Based ECG Delineator: Evaluation on Standard Databases[J]. IEEE Transactions on Biomedical Engineering. 2004, 51(4): 570-581.

[5] Lemm S, Blankertz B, Curio G, et al. Spatio-Spectral Filters for Improving the Classification of Single Trial EEG[J]. IEEE Transactions on Biomedical Engineering. 2005, 52(9): 1541-1548.

[6] Muirhead R J, Puff R D. A Bayesian classification of heart rate variability data[J]. Physica A: Statistical Mechanics and its Applications. 2004, 336(3-4): 503-513.

[7] Christov I, Jekova I, Bortolan G. Premature ventricular contraction classification by theK th nearest-neighbours rule[J]. Physiological Measurement. 2005, 26(1): 123-130.

[8] Jekova I, Bortolan G, Christov I. Pattern recognition and optimal parameter selection in premature ventricular contraction classification[C]. 2004. IEEE, 2004: 357-360.

[9] Osowski S, Linh T H. ECG beat recognition using fuzzy hybrid neural network[J]. IEEE Trans Biomed Eng. 2001, 48(11): 1265-1271.

[10] Guler I, Ubeyli E. ECG beat classifier designed by combined neural network model[J]. Pattern Recognition. 2005, 38(2): 199-208.

[11] Melgani F, Bazi Y. Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization[J]. IEEE Transactions on Information Technology in Biomedicine. 2008, 12(5): 667-677.

[12] Soman S, Jayadeva. High performance EEG signal classification using classifiability and the Twin SVM[J]. Applied Soft Computing Journal. 2015, 30: 305-318.

[13] She Q, Ma Y, Meng M, et al. Multiclass Posterior Probability Twin SVM for Motor Imagery EEG Classification[J]. Computational Intelligence and Neuroscience. 2015, 2015: 1-9.

[14] Rahhal M M A, Bazi Y, Alhichri H, et al. Deep learning approach for active classification of electrocardiogram signals[J]. Information Sciences. 2016, 345: 340-354.

[15] Xiong P, Wang H, Liu M, et al. A stacked contractive denoising auto-encoder for ECG signal denoising[J]. Physiological Measurement. 2016, 37(12): 2214-2230.

[16] Luo K, Li J, Wang Z, et al. Patient-Specific Deep Architectural Model for ECG Classification[J]. Journal of Healthcare Engineering. 2017, 2017: 1-13.

[17] Acharya U R, Oh S L, Hagiwara Y, et al. A deep convolutional neural network model to classify heartbeats[J]. Computers in Biology and Medicine. 2017, 89: 389-396.

[18] Shaker A M, Tantawi M, Shedeed H A, et al. Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks[J]. IEEE Access. 2020, 8: 35592-35605.

[19] Xu X, Liu H. ECG Heartbeat Classification Using Convolutional Neural Networks[J]. IEEE Access. 2020, 8: 8614-8619.

[20] Wang G, Yang L, Liu M, et al. ECG signal denoising based on deep factor analysis[J]. Biomedical Signal Processing and Control. 2020, 57: 101824.

[21] Chiang H, Hsieh Y, Fu S, et al. Noise Reduction in ECG Signals Using Fully Convolutional Denoising Autoencoders[J]. IEEE Access. 2019, 7: 60806-60813.

[22] Li R, Yang G, Wang K, et al. Robust ECG biometrics using GNMF and sparse representation[J]. Pattern Recognition Letters. 2020, 129: 70-76.

[23] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: Synthetic Minority Over-sampling Technique[J]. The Journal of artificial intelligence research. 2002, 16: 321-357.

[24] Han H, Wang W, Mao B. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning[M]. Berlin, Heidelberg:Springer Berlin Heidelberg, 878-887.

[25] Barnaba C, Taylor E, Brozik J A. Dissociation Constants of Cytochrome P450 2C9/Cytochrome P450 Reductase Complexes in a Lipid Bilayer Membrane Depend on NADPH: A Single-Protein Tracking Study[J]. Journal of the American Chemical Society. 2017, 139(49): 17923-17934.

[26] Zheng Z, Qu C, Zou X, et al. Fulde-Ferrell Superfluids without Spin Imbalance in Driven Optical Lattices[J]. Physical review letters. 2016, 116(12): 120403.

[27] Prati R C B G E A. Learning with Class Skews and Small Disjuncts[C]. Advances in Artificial Intelligence – SBIA, 2004. 2004.

[28] Yen S, Lee Y. Cluster-based under-sampling approaches for imbalanced data distributions[J]. Expert systems with applications. 2009, 36(3): 5718-5727.

[29] Shalizi C R, Rinaldo A. Consistency under sampling of exponential random graph models[J]. The Annals of statistics. 2013, 41(2): 508-535.

[30] Tian J, Liu Y, Glaser N, et al. Posterior Re-calibration for Imbalanced Datasets[J]. Proceedings of the 34th Conference on Neural Information Processing Systems. 2020.

[31] Lin T, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[C]. 2017. IEEE, 2017: 2999-3007.