Predicting Neurological Outcomes for Cardiac Arrest Patients from Longitudinal EEG Based on Short-Time Fourier Transform and 3-D Deep Residual Network

Pan Xia[1,2], Dongfang Zhao[1,2], Yicheng Yao[1,2], Zhongrui Bai[2], Yizi Shao[1,2], Saihu Lu[1,2], Fanglin Geng[1,2], Yusi Zhu[3], Peng Wang[2], Lidong Du[1,2,4]

[1]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China
[2]State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China
[3]School of Physics and Electronic Information, Yunnan Normal University, Kunming, China
[4]Personalized Management of Chronic Respiratory Disease, Chinese Academy of Medical Sciences, China

Abstract

The George B. Moody PhysioNet Challenge 2023 focused on predicting neurological recovery from coma after cardiac arrest by longitudinal electroencephalogram (EEG). Our team, MetaHeart_YNNU, proposed a novel approach to predict good and poor patient outcomes after cardiac arrest by combing short-time Fourier transform (STFT) and 3-dimensional (3-D) residual neural network. Firstly, a 90-second signal segment was obtained from the 8-channel EEG recording of each hour by truncating the first 15 seconds every 10 minutes, and then these segments from each hour were transformed into the time-frequency spectrograms by STFT to obtain a richer representation of neurological activity. Secondly, a modified 3-D residual neural network was designed to extract the complex physiological patterns of neurological activity from hour, time-frequency dimensions, respectively. Thirdly, a joint loss combing binary focal loss and recall loss was designed in our approach. The binary focal loss alleviated the class imbalance problem by paying more attention to hard-to-distinguish samples. The recall loss was adopted to reduce the probability that patients with the true good label were misclassified into the poor category by maximizing the recall rate of the good category. Finally, our proposed model received a Challenge score of 0.56 (ranked 12th out of 36 teams) on the hidden test set.

1. Introduction

More than 6 million cardiac arrests happen every year worldwide, which may cause the severe brain injury for patients surviving initial resuscitation and seriously threaten people's lives [1]. Electroencephalogram (EEG) can reflect the recovery of brain consciousness in cardiac arrest patients and assist neurophysiologists to give an objective prognosis [2-3]. The George B. Moody PhysioNet Challenge 2023 is devoted to predict neurological recovery for cardiac arrest patients who present to the hospital in a coma by using basic clinical information and EEG and ECG recordings [4-5]. The participants were asked to build an algorithm that can identify whether the patient outcome is poor or good, as well as the level of neurological recovery for cardiac arrest patients. In this work, we proposed a novel approach to achieve the target by combing short-time Fourier transform (STFT) and 3-dimensional (3-D) residual neural network. The EEG data truncated from each hour was transformed into a time-frequency spectrogram by using STFT. The 72-hour time-frequency spectrograms were stacked together to form a 3-D tensor involved hour, time-frequency dimensions. A modified 3-D residual network was built to extract the complex physiological patterns of neurological activity. A joint loss combing binary focal loss and recall loss was designed to optimize our model. We introduced an ensemble prediction strategy to improve the robustness of model predictions in model prediction process. Our final selected entry was firstly 5-fold cross-validated on the public training set and then was scored on the hidden validation set by the Challenge organizers. Ultimately, our final selected entry was scored as well as ranked on the hidden test set with the Challenge metric.

2. Methods

2.1. Datasets and Preprocessing

The Challenge data originates from the International Cardiac Arrest EEG Consortium (ICARE), which was collected from 1,020 cardiac arrest patients from seven hospital in Europe and the U.S. [6]. The Challenge data is divided into the public training set, the hidden validation set and the hidden test set in proportions of 60%, 10%, and 30%, respectively [4]. Each patient in the Challenge data has multi-modal physiological signals including EEG, ECG and/or other clinical time series data. The data was recorded from several hours after the cardiac arrest occurred, and the length of data duration could vary from several hours to days.

In this work, we only used EEG to predict neurological recovery in cardiac arrest patients. We applied the following data pre-processing procedures. Firstly, to reduce memory usage and speed up the inference process, we chose 8 typical channels covering most brain areas from the 19-channel EEGs, namely Fp1, Fp2, F3, F4, C3, C4, P3, P4. The selected 8-channel EEGs were resampled to 100 Hz, and then we applied a bandpass filter with a cutoff frequency of 0.7~30Hz to the EEG signal to eliminate baseline drift and power frequency noise. Secondly, to reduce the memory usage and speed up the calculation, we selected 90-second signal segment from the 8-channel EEG recording of each hour by truncating the first 15 seconds every 10 minutes. These segments from each hour were transformed into the time-frequency spectrograms of the shape (65, 283) by using short-time Fourier transform to obtain a richer representation of neurological activity. Where 65 is the size in the frequency dimension and 283 is the size in the time dimension. And then, Z-score normalization was applied to normalize each spectrum. Thirdly, the 72-hour time-frequency spectrograms were stacked together to form a 3-D tensor with shape of (72, 65, 283) involved hour, time-frequency dimensions, respectively.

2.2. Model Architecture

Since the 72-hour EEG data is massive, in order to more effectively capture the continuous changes of neurological activity patterns from the cardiac arrest patients, we modeled this problem as a 3-D time series variation, that is, hour, time-frequency dimensions. Given that deep residual networks have effective feature extraction capabilities [7-8]. A 16-layer 3-D residual neural network was built in our approach. The overall structure of the model was shown in Figure 1. The input shape of main network is (8, 72, 65, 283), which denotes a sample consists of 72 groups (i.e. 72 hours) of time-frequency spectrograms with shape of (65, 283) from 8 channels. The input data was first fed into a convolution layer with convolution kernel size (5, 5, 7) and using stride 2. 6 residual blocks were stacked to form the backbone of network. After the global average pooling (GAP) operation, the length of feature vectors was converted to 2 (number of categories) by using a linear operation. Then, the classifier logit outputs were forwarded into the loss function. *ReLU* is adopted to all layers except for the output layer, and the output layer uses *Softmax* because the neurological outcome prediction is a mutually exclusive binary classification task in this year's Challenge. *Softmax* can map the model output to a vector whose probability values add up to 1.

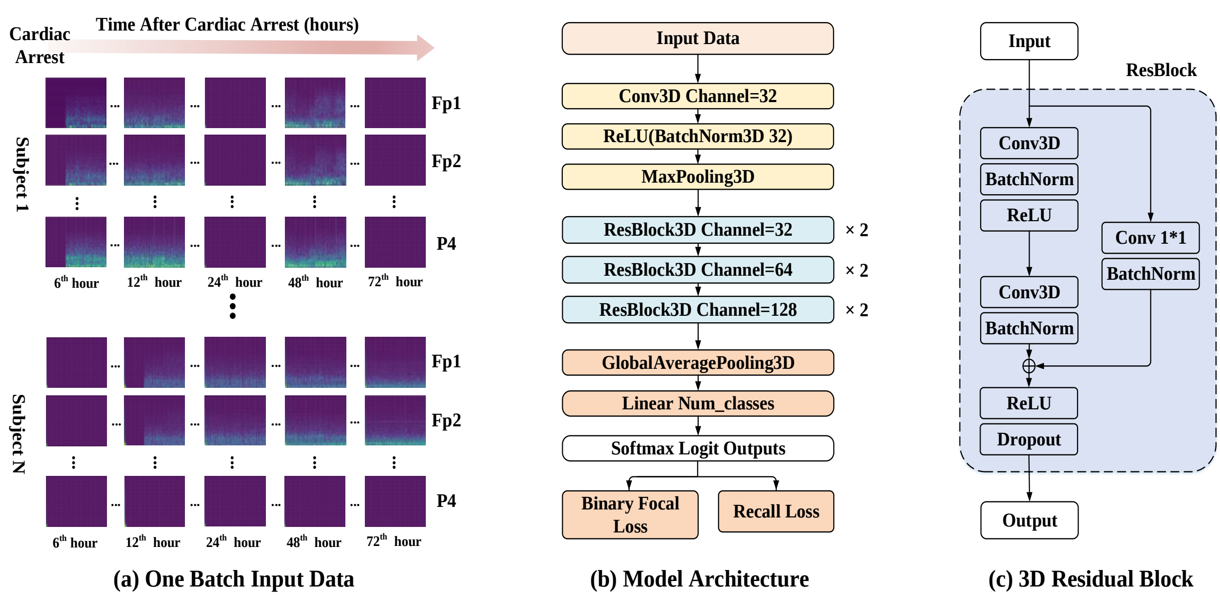

Figure 1. The architecture diagram of our proposed approach. (a) One batch input data for the 3-D residual neural network. (b) The model architecture of 3-D deep residual neural network. (c) 3-D residual block.

2.3. Loss Functions

We adopted the joint supervision of binary focal loss and recall loss to train the proposed model. The total loss formulation is given in Eq. 1.

$$L = l_1 + \lambda * l_2 \tag{1}$$

where the λ denotes a hyper parameter for balancing the two loss functions. In this work, λ is set to 0.5.

Binary Focal Loss

We counted the number of samples for poor and good categories in the public training set, there was a slight class imbalance in the Challenge data, which could domain the optimization process, and easily classified samples comprise the majority of the loss and dominate the gradients while under-emphasizing gradients from hard classified samples during training.

Inspired by Lin et al [9], we introduced the binary focal loss for our proposed model training to alleviate the above problems, and which is defined as Eq. 2.

$$l_1 = -\frac{1}{m}\sum_{i=1}^{m} \alpha * (1 - p_t)^{\gamma} * log(p_t) \tag{2}$$

where

$$p_t = \begin{cases} p, & if\ y = 1 \\ 1 - p, & otherwise \end{cases} \tag{3}$$

where p denotes the model's corresponding prediction probability for the true label $y = 1$. The weighting factor α and the tunable focusing parameter γ are set to 1 and 2, respectively. The total average is considered as the focal loss term.

Recall Loss

In this year's Challenge, the scoring metric is calculated as the true positive rate (TPR) for predicting a poor outcome given a false positive rate (FPR) of less than or equal to 0.05 at 72 hours after return of spontaneous circulation, which is more inclined to make the recall rate high enough for good category. The scoring metric does not expect the patients with true good labels to be mistakenly classified as poor, so that the patients with true good labels lose more treatment resources. In order to make the model maintain the above prejudice, the recall loss is introduced to optimize the model. We prefer to maximize the recall rate of the good category to reduce the probability that patients with the true good labels were misclassified into the poor category, therefore, the recall loss is defined as Eq. 4:

$$l_2 = 1 - mean(recall_{good}) \tag{4}$$

$$where \quad recall_{good} = \frac{TN}{TN + FP} \tag{5}$$

Where TN denotes the number of patients with true good labels who are predicted as good. Where FP denotes the number of patients with true good labels who are predicted as poor.

As above, the binary focal loss and recall loss were adopted to simultaneously optimize our proposed model, which allows the model to perform binary classification while ensuring a higher recall rate for the good category.

2.4. Model Training

Our proposed model is trained 25 epochs with a batch size of 12 using an NVIDIA GeForce RTX 3090. *Adam* optimizer with an initial learning rate of 0.0005 was applied for model optimization. The multiple step learning rate scheduler was adopted to dynamically adjust the learning rate in the training process, and the strategy of reducing the learning rate with a ratio of 0.5 during training was adopted to speed up model convergence. Model training is stopped when the model's score on the validation set does not improve after 13 epochs. The optimal hyper-parameter of our network (convolution kernel size, dropout rate, number of residual blocks, etc.) and the parameters of loss function (loss balancing factor λ, weighting factor α, and tunable focusing parameter γ) are selected according to the model's 5-fold cross validation performance on the public training set.

2.5. Model Prediction

Considering that the data continuously monitored for 72 hours is massive, under the conditions of limited computing and memory resources, it is impossible to traverse all the data to make the final prediction. However, it is very necessary to use as much data as possible to support model predictions, which can improve the robustness of model prediction results. To address this problem, we use an ensemble learning strategy to vote on multiple prediction results. Specifically, we extracted 5 groups of data in each hour by truncating a 15-second signal segment every 10 minutes to implement model prediction integration. For the prediction of a certain class such as poor, only when the number of prediction results is greater than or equal to 3, the class is finally predicted as poor, otherwise good. The average of the 5 groups of outcome probabilities is taken as the final predicted outcome probability output of the model.

3. Results

We evaluated our proposed algorithms through 5-fold cross-validation on the public training set with the official Challenge metric. The Challenge scores on both the public training set, hidden validation set, and hidden test set that our final selected entry obtained were shown in Table 1.

Training	Validation	Test	Ranking
0.54±0.08	0.58	0.56	12/36

Table 1. True positive rate at a false positive rate of 0.05 (the official Challenge score) for our final selected entry

(team MetaHeart_YNNU), including the ranking of our team on the hidden test set. We used 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set.

4. Discussion and Conclusions

The 72-hour longitudinal EEG recordings provide an opportunity to objectively evaluate the neurological recovery of cardiac arrest patients. Considering the non-stationary characteristics of EEG signals, we used short-time Fourier transform technology to represent the changes of neurological activity patterns. We considered the changes of neurological activity patterns in the several hours and days after cardiac arrest to be critical information for identifying neurological recovery. Therefore, we adopted 3-D convolution to capture the patterns features switching in hour-dimensional. Although we believed that our approach was relatively reasonable, the performance of the approach on the public training set and the hidden validation set is not significant, which makes us wonder whether it is because our approach failed to capture the more general features from 72-hour longitudinal EEGs. In addition, severe data missing and irregular external defibrillation operations may also be the reasons for the poor performance of our model.

In this paper, we proposed a novel approach to predict good and poor patient outcomes after cardiac arrest by combing short-time Fourier transform (STFT) and 3-D residual neural network. The longitudinal EEG recordings were transformed into time-frequency spectrums to obtain a richer representation of neurological activity. By jointly optimizing the proposed model using binary focal loss and recall loss, our model could perform binary classification while ensuring a higher recall rate for the good category. In addition, an ensemble prediction strategy was adopted to improve the robustness of model predictions by voting on five prediction results in model prediction process. Our proposed models were firstly evaluated on the public training set, we achieved 5-fold cross-validation score of 0.54 with the Challenge evaluation metric. Finally, our classifier received a Challenge score of 0.56 (ranked 12th out of 36 teams) on the hidden test set.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62371441), National Natural Science Foundation of China (No. 62331025), National Natural Science Foundation of China (No. 62071451), and CAMS Innovation Fund for Medical Sciences (2019-I2M-5-019).

References

[1] Medicine I of. Strategies to Improve Cardiac Arrest Survival: A Time to Act [Internet]. Washington, DC: The National Academies Press; 2015. [Cited 2021 Dec 5]. Available from: https://www.nap.edu/catalog/21723/strategies-to-improve-cardiac-arrest-survival-a-time-to-act

[2] Hofmeijer J, Beernink TMJ, Bosch FH, et al. Early EEG contributes to multimodal outcome prediction of postanoxic coma. Neurology 2015;85(2):137–143.

[3] Tjepkema-Cloostermans MC, da Silva Lourenço C, Ruijter BJ, et al. Outcome Prediction in Postanoxic Coma With Deep Learning. Crit. Care Med. 2019;47(10):1424–1432.

[4] Reyna MA, Amorim E, Sameni R, Weigle J, Elola A, Bahrami Rad A, et al. Predicting neurological recovery from coma after cardiac arrest: The George B. Moody PhysioNet Challenge 2023. Computing in Cardiology 2023;50:1–4.

[5] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215–e220.

[6] Amorim E, Zheng WL, Ghassemi MM, Aghaeeaval M, Kandhare P, Karukonda V, Lee JW, Herman ST, Adithya S, Gaspard N, Hofmeijer J, van Putten MJAM, Sameni R, Reyna MA, Clifford GD, Westover MB. The International Cardiac Arrest Research (I-CARE) Consortium Electroencephalography Database. Critical Care Medicine 2023 (in press); doi:10.1097/CCM.0000000000006074.

[7] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2016; 770-778.

[8] Andreotti F, Carr O, Pimentel M, et al. Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG. Computing in Cardiology 2017;1–4.

[9] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision (ICCV). 2017; 2980-2988.

Address for correspondence:

Yusi Zhu
School of Physics and Electronic Information, Yunnan Normal University, No. 1, Yuhua, Chenggong District, Kunming, China.
zhuyusi16@mails.ucas.edu.cn.

Peng Wang and Lidong Du
State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences,
No. 19, North Fourth Ring West Road, Zhongguancun, Haidian District, Beijing, China.
wangpeng01@aircas.ac.cn; lddu@mail.ie.ac.cn.