MMCTNet: Multi-Modal Conv-Transformer Network for Predicting Good and Poor Outcomes in Cardiac Arrest Patients

Xiuli Bi[1] Shizhan Tang[1], Zonglin Yang[1] Xin Deng[1] Bin Xiao[1] Pietro Liò[2]

[1] Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2] University of Cambridge, U.K.

Abstract

Electroencephalography (EEG) has been demonstrated to be a valuable tool for predicting neurological outcomes after cardiac arrest. However its complexity limits timely interpretation. As part of George B. Moody Physionet Challenge 2023, our team (CQUPT_FP_mana) proposes a Multi-Modal Conv-Transformer network to accurately and timely assess probability of coma recovery with complex EEG. We select the five-minute EEGs nearest to the six moments of 12h, 24h, 36h, 48h, 60h, and 72h, respectively, resample them to 100Hz, filter them using a 5-order Butterworth bandpass filter, and slice the EEGs into 10-second slices. The one-dimensional representation of the EEG and the time-frequency spectrograms after a short-time Fourier transform are then used as inputs to the network structure. Our network consists of two convolutional branches, a transformer encoder, and a classification header. In the end, our network scored 0.48 points in the tournament, placing us 19th out of all challenged teams.

1. Introduction

More than 6 million cardiac arrests (CA) occur worldwide each year, with survival rates ranging from 1% to 10% depending on geographic location [1, 2], and for the medically advanced U.S., cardiac arrest is the third leading cause of death, with more than 356,000 out-of-hospital cardiac arrests (OHCA) annually [3]. Most patients who survive to hospitalization are comatose due to hypoxic brain injury, and severe brain injury is the most common cause of death in surviving patients [4]. In the first few days after cardiac arrest, physicians are asked to provide a prognosis as to whether the patient will regain consciousness. Different outcomes may result in continued care or withdrawal of life support until death. However, false positives occur: poor prognostic outcomes, but actual patients recovered better. False prognosis raises concerns. Therefore, an early and accurate prognosis is essential for clinical decision-making as well as timely intervention, and several guidelines on CA prognosis have been proposed in recent decades [5, 6].

The purpose of brain testing and EEG is to eliminate the subjectivity of neurological prognosis after cardiac arrest, and clinical neurophysiologists have recognized a number of patterns of brain activity that help to predict the prognosis of cardiac arrest, including the presences of reduced voltage, burst suppression (alternating periods of high and low voltage), seizures, and a variety of seizure-like patterns [7]. With continued EEG monitoring, the evolution of EEG patterns can provide additional information on prognosis [8, 9]. However, continued qualitative EEG interpretation is laborious and expensive and requires specialized training and review by experienced senior neurologists, resources that are not available in most areas. Therefore, computer-assisted physician prediction of a patient's ability to regain consciousness after cardiac resuscitation is a way to improve prediction accuracy and reduce overhead. Automated analysis of continuous EEG data has been shown to improve prediction accuracy and increase the chances of brain detection that is not accessible to experts [9]. In this paper, we propose a deep learning framework based on the EEG data [10] provided by the International Cardiac Arrest REsearch consortium (I-CARE) to the George B. Moody PhysioNet Challenge [11] to joint learning and prediction in the time domain and time-frequency domain to determine whether a patient can regain consciousness after cardiac resuscitation.

2. Methods

2.1. Dataset

The PhysioNet Challenge database from seven hospitals included 1,000 subjects who collectively received over 50,000 hours of EEG monitoring. Each patient had EEG data of different lengths of time. The recordings typically begin several hours after the arrest and have brief interruptions while in the ICU, so gaps in the data may be present, and the text data recorded the patient's age, gender, OHCA, Shockable Rhythm, as well as outcomes and Cerebral Performance Category (CPC) and so on. There are other auxiliary signals such as electrocatdiogram (ECG), electromyography (EMG) and so on in the Challenge file package. In this paper we use only EEG data.

2.2. Preprocessing

For different patients have different time series of EEG signals, we finally selected signals from 6-time points, i.e., starting from the 12th hour, with each time point 12 hours apart. If the time point we need is not present in the data then we found the nearest one directly from around that time point. If the time points are duplicated then one is kept. We intercepted the first 5 minutes of the EEG at each time point as data for this time point and resampled to 100 Hz [12]. We filtered by a 5th-order Butterworth band-pass filter.EEGs were re-referenced to 18 bipolar channels (Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4,T4-T6, T6-O2, Fp1-F3, F3-C3, C3-P3, P3 -O1, Fp2-F4, F4-C4.C4-P4, P4-O2, Fz-Cz, Cz-Pz). Choosing a multi-polar EEG can bring us more space-related information.

2.3. Feature Extraction

Because the network structure constructed in this paper is a multi-modal two-branch structure, the features of the input network are divided into two parts. First, we used a sliding window of length 10 seconds to segment a 5-minute long signal as the first part of the input $X_1 \in \mathbb{R}^{18\times1000}$. The window slides without overlap. Then we considered using the STFT as time-frequency domain transform method. We took the time-frequency spectrogram of X_1 after STFT as the second division input $X_2 \in \mathbb{R}^{18\times129\times9}$. It is worth noting that the both inputs to networks possess temporal information, so there is no need to consider timing alignment.

2.4. Neural Network Structure

In recent years, multi-modal has already achieved good results in visual recognition tasks [13]. Multi-modal inputs can bring more auxiliary information to the network model to help the model improve the classification performance. Therefore, we used the multi-modal multi-branch network structure to determine the likelihood of a patient being awake through EEG and its transformations. The framework of our network model inspired by [12] and is shown in Figure 1. Our network model is divided into three parts: convolution, self-attention, and classification head.

In the first part, the two convolutional branches learn to extract different features from different inputs, and for different modalities, we consider different sizes of convolutional kernels. For 18 channels of 1D EEG, the convolution kernel for our first convolution was set to 1x25, in order to extract features for each channel. And the second convolution used a convolution kernel of 18x1 in order to obtain information about the relationship between the channels. This setup allows us to more fully obtain the information of the 1D signal. Furthermore, in the second branch, to better utilize the corresponding 18 time-frequency spectrograms obtained through the short-time Fourier transform we used the results of the ResNet18 [14] classical network as a feature extractor. Finally, we concatenate the features of the two branches. In the second part, it is the standard self-attention part of the Transformer that extracts the global correlation of local temporal features, or even the global correlation between multiple modalities. The last part uses a simple classifier consisting of fully connected layers to achieve classification of EEGs.

2.5. Voting mechanisms

The voting mechanism, based on the principle of majority rule, can improve the accuracy and generalization of our model predictions. We voted for the final prediction for each patient based on the classification results of a total of 180 EEG signal slices from 6 time points.

3. Results

3.1. Settings and Metrics

Our work is based on python 3.8 and the pytorch framework. MMCTNet was trained for 100 epoch on a single GPU (NVIDIA RTX 3090 Ti with 24GB memory) with mini-batch size 64. We chose Adam optimizer and set the learning rate to 0.0002. For the training sample imbalance problem, we set different weights for the cross-entropy loss of each sample. For this Challenge, the scoring metric is the true positive rate (TPR) for predicting a poor outcome (CPC of 3, 4, or 5) given a false positive rate (FPR) of less than or equal to 0.05 at 72 hours after return of spontaneous circulation [15].

3.2. Challenge Score

In this section, the results of the official phases of our method on official unpublished validation and hidden test datasets will be shown. In the local unofficial phase, we divided the downloaded official dataset into a training set, validation set, and test set by 8:1:1, and the same for the official phase. As shown in Table 1. We show in the table the local scores as well as Area Under the Curve-Receiver Operating Characteristic (AUC-ROC), F1-Score (F1), and Mean Absolute Error (MAE) in both phases.

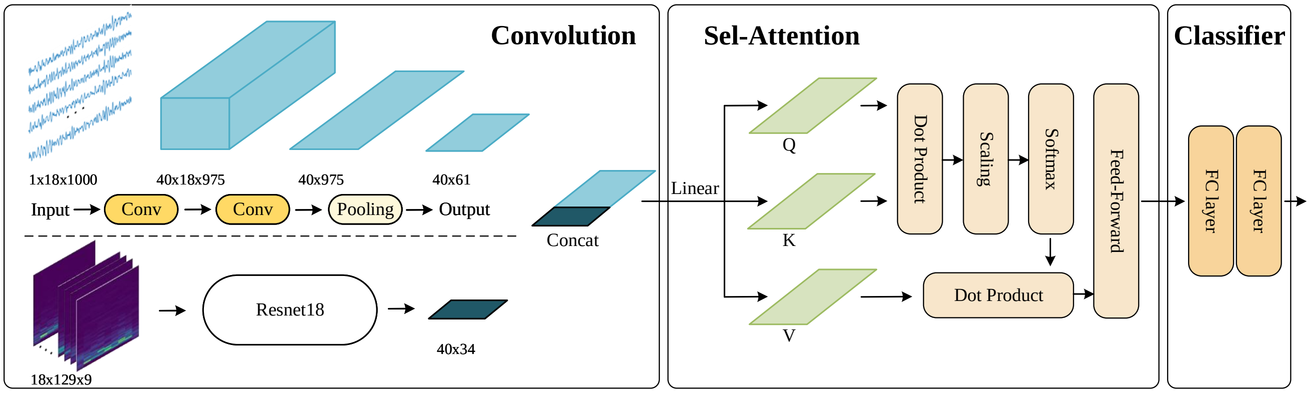

Figure 1. MMCTNet complete structure and intra-structure details.

Table 1. Local scores for both phases and other metrics results.

Phase	Score	ROC	F1	MAE
Unofficial	0.59	0.78	0.75	1.32
Official	0.56	0.76	0.65	1.59

In Table 2, we show the prediction scores of MMCT-Net at different phases, and different time points. From the table, we can find the predictive power of our method for both EEGs collected early in the patient's coma and EEGs collected subsequently. Our model is able to capture changes in EEGs.

Table 2. Local final score results at four time points.

Phase	12h	24h	48h	72h
Unofficial	0.26	0.40	0.49	0.59
Official	0.11	0.26	0.30	0.56

Above are the results of our model on the local training set. Most important is our score on the hidden test set in the competition. As shown in Table 3, we show the final results of our model on the training set, validation set, and test set after the official phase.

Table 3. PhysioNet Challenge 2023 scores for our team.

Set	training	validation	hidden test
Score	1.000	0.567	0.480

From the metrics we can understand that there is overfitting in our model, and the subsequent resolution of the overfitting problem to improve the generalization of the model is worth being investigated.

3.3. Ablation

Our model consists of multiple modules and multiple branches, and in order to further explore whether each module and input is effective in improving the prediction performance, we conducted a simple ablation experiment on the whole model.The results of the experiment are shown in Table 4.

Table 4. Impact on prediction after ablation of different modules.

	our	base	spac w/o	signal w/o	encoder w/o
param	12.33M	-	0.78M	11.67M	12.21M
Score	**0.59**	0.37	0.56	0.19	0.07
ROC	**0.78**	0.70	0.73	0.72	0.75
F1	**0.75**	0.59	0.72	0.68	0.70
MAE	**1.32**	1.58	1.31	1.36	1.38

We directly used the dataset downloaded from the unofficial stage for our experiments, with the same parameter settings and dataset processing as before. The five columns in the table show, respectively, the results for the complete structure of MMCTNet, the results for the sample official decision tree model, and the results for MMCTNet without spectrograms as inputs, without one-dimensional signals as inputs, and without transform encoder structure. Without the corresponding modal input we deleted the corresponding structural part. From the above results, we can conclude that each module in MMCTNet can contribute to the prediction performance improvement, especially the encoder part, which is particularly significant in terms of score. This global long-range dependency is very effective for feature extraction of signals possessing temporal features. The time-frequency spectrogram as a multi-modal auxiliary input has an improvement on all types of metrics.

4. Discussion and Conclusion

In this paper, we propose a multi-branch multi-modal network structure to predict whether a patient will regain consciousness or not. Multi-modal can assist each other to provide more adequate feature information, and the network structure of the combination of convolution and transformer has been proved to be both locally and globally oriented in its feature extraction. We also consider the use of pre-trained models to improve the performance of our model. However, our structure has modifications to the classical model in the implementation details, so we cannot directly use the model parameters that are publicly available on the web, so we used a large heart sound dataset in hand to train the model for migration learning, but the results are not ideal at the moment. We believe that the model has not reached its best performance, and thus, how to further improve the model performance is a direction that can be explored in the future. In the end, our proposed structure scored 0.48 in the challenge, placing it 19th out of all the teams that made it to the official stage.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62172067 and Grant 61976031,in part by the Natural Science Foundation of Chongqing for Distinguished Young Scholars under Grant CSTB2022NSCQ-JQX0001.

References

[1] Graham R, McCoy MA, Schultz AM. Strategies to improve cardiac arrest survival: a time to act 2015;.

[2] Mehra R. Global public health problem of sudden cardiac death. Journal of Electrocardiology 2007;40(6, Supplement 1):S118–S122. ISSN 0022-0736. URL `https://www.sciencedirect.com/science/article/pii/S002207360700667X`.

[3] Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Das SR, et al. Heart disease and stroke statistics—2019 update: a report from the american heart association. Circulation 2019;139(10):e56–e528.

[4] Rundgren M, Westhall E, Cronberg T, Rosen I, Friberg H. Continuous amplitude-integrated electroencephalogram predicts outcome in hypothermia-treated cardiac arrest patients. Critical care medicine 2010;38(9):1838–1844.

[5] Callaway CW, Donnino MW, Fink EL, Geocadin RG, Golan E, Kern KB, Leary M, Meurer WJ, Peberdy MA, Thompson TM, et al. Part 8: post–cardiac arrest care: 2015 american heart association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2015;132(18_suppl_2):S465–S482.

[6] Sandroni C, Cariou A, Cavallaro F, Cronberg T, Friberg H, Hoedemaekers C, Horn J, Nolan JP, Rossetti AO, Soar J. Prognostication in comatose survivors of cardiac arrest: an advisory statement from the european resuscitation council and the european society of intensive care medicine. Intensive care medicine 2014;40:1816–1831.

[7] Hirsch LJ, Fong MW, Leitinger M, LaRoche SM, Beniczky S, Abend NS, Lee JW, Wusthoff CJ, Hahn CD, Westover MB, et al. American clinical neurophysiology society's standardized critical care eeg terminology: 2021 version. Journal of clinical neurophysiology official publication of the American Electroencephalographic Society 2021;38(1):1.

[8] Amorim E, Rittenberger JC, Zheng JJ, Westover MB, Baldwin ME, Callaway CW, Popescu A, et al. Continuous eeg monitoring enhances multimodal outcome prediction in hypoxic–ischemic brain injury. Resuscitation 2016; 109:121–126.

[9] Hofmeijer J, Beernink TM, Bosch FH, Beishuizen A, Tjepkema-Cloostermans MC, van Putten MJ. Early eeg contributes to multimodal outcome prediction of postanoxic coma. Neurology 2015;85(2):137–143.

[10] Amorim E, Zheng W, Ghassemi M, Aghaeeaval M, Kandhare P, Karukonda V, Lee J, Herman S, Sivaraju A, Gaspard N, et al. The international cardiac arrest research (i-care) consortium electroencephalography database. medrxiv 2023. URL httpswww me drxiv orgcontentearly20230828202 ;3(28.23294672).

[11] Goldberger AL Amaral LA GLHJIPMRMJMGPCSHPP, PhysioNet. Physionet: components of a new research resource for complex physiologic signals. Circulation 2000; 101(23):e215–e220.

[12] Zheng WL, Amorim E, Jing J, Wu O, Ghassemi M, Lee JW, Sivaraju A, Pang T, Herman ST, Gaspard N, et al. Predicting neurological outcome from electroencephalogram dynamics in comatose patients after cardiac arrest with deep learning. IEEE transactions on biomedical engineering 2021; 69(5):1813–1825.

[13] Wang Y, Peng J, Zhang J, Yi R, Wang Y, Wang C. Multimodal industrial anomaly detection via hybrid fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023; 8032–8041.

[14] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016; 770–778.

[15] Reyna MA, Amorim E, Sameni R, Weigle J, Elola A, Bahrami Rad A, Seyedi S, Kwon H, Zheng WL, Ghassemi M, et al. Predicting neurological recovery from coma after cardiac arrest: The george b. moody physionet challenge 2023. Computing in Cardiology 2023;50:1–4.

Address for correspondence:

Xin Deng

No.2 Chongwen Road, Nanan District, Chongqing, 400065, China.

dengxin@cqupt.edu.cn