

Predicting Neurological Recovery from Coma with Longitudinal Electroencephalogram Using Deep Neural Networks

Jingsu Kang¹, Hao Wen²

¹Tianjin Medical University, Tianjin, China

²College of Science, China Agricultural University, Beijing, China

Abstract

Aim: This work studies the problem of predicting neurological recovery from coma with longitudinal electroencephalogram (EEG) recordings raised by the George B. Moody PhysioNet Challenge 2023.

Methods: Deep neural network (DNN) models were trained to predict cerebral performance category (CPC) scale (1 to 5) from bipolar EEGs which were rescaled to zero mean and unit variance. The prediction was treated as a 5-class classification task. The models adopted a bottleneck SE-ResNet backbone with long short-term memory (LSTM) and global attention modules on its top.

Via a stratified splitting, 20% of the training data were left out as a validation set for model selection. Recordings were chosen and sliced according to the pre-computed signal quality index for training. The AdamW optimizer and the OneCycle scheduler were used to optimize model weights on the asymmetric loss of the training data. Predictions of multiple EEG recordings from one patient were merged via voting strategies to give a final prediction.

Results: Our team ("Revenger") received a challenge score of 0.554 for the clinical outcome prediction on the hidden test set, ranked 13th out of 36 teams.

Conclusion: Our solution offers a practical way to continuously monitor the brain after cardiac arrest using EEGs, with room and potential for further enhancements.

1. Introduction

Cardiac arrest (CA), especially out-of-hospital cardiac arrest (OHCA), is a major worldwide public health issue with a low survival rate. CA often results in cerebral lesions, a leading cause of early death for patients resuscitated from it [1]. Physicians prognosticate the neurological recovery outcomes for survivors comatose after initial treatments [2], which is a challenging task requiring a minimal false-positive rate of poor outcomes.

With the aim of improving objectivity and accessibility for reliable prognostication after CA, the George B.

Moody PhysioNet Challenge 2023 [3, 4] proposed to develop automated auxiliary prognostic support systems using longitudinal EEG. In this work, a novel approach providing qualitative neurological interpretations and predictions is presented.

2. Methods

2.1. Dataset and SQI-based Data Selection

The public dataset for the Challenge is the International Cardiac Arrest REsearch consortium (I-CARE) database [5]. It contains mainly EEG recordings from comatose patients with cardiac arrest which were collected up to 72 hours from their return of spontaneous circulation (ROSC). Although this dataset contains physiological signals other than EEG, we base this study only on the EEG data for the following 2 reasons:

- The amount of EEG data in the public dataset is already enormous, exceeding 30,000 hours in length.
- Previous literature [6] has illustrated the feasibility of using EEG as the sole physiological signal source on the problem this work considers.

As the computation resources and execution time are limited under the circumstances of the Challenge, further offline data selection was conducted according to the signal quality index (SQI) of the EEGs. The SQI was computed for every 5-minute epoch as its proportion of "normal" 10s segments. A 10s segment is regarded as "normal" if no artefact is detected from it, including abnormal values and patterns in the waveform and in the spectrum¹ At most one epoch was chosen (requiring $SQI \geq 0.95$) from each EEG recording as a representative of it for developing our neural network (NN) models (models will be discussed in Section 2.3). The total number of 5-minute epochs selected was 23350.

¹Refer to https://github.com/DeepPSP/cinc2023/blob/master/utils/artifact_pipeline for more details.

2.2. Preprocessing Criteria

In this subsection, we describe our preprocessing pipeline for model development and making inferences.

The EEG recordings provided by the I-CARE dataset have varying numbers (19 - 21) of channels but share a common set of 19 channels. The voltage values of each EEG signal are relative to an “unknown” reference potential. Hence in order to align these voltage values, the varying-dimensional EEG signals were transformed into the bipolar format with the following 18 channels

$$\begin{aligned} &Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, \\ &T4-T6, T6-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, \\ &Fp2-F4, F4-C4, C4-P4, P4-O2, Fz-Cz, Cz-Pz \end{aligned}$$

Bipolar values were obtained by subtracting the latter channel from the former channel of the above 18 pairs.

Bipolar EEGs were further standardized with the following 3 sequential operations:

- resample to 100 Hz using polyphase filtering;
- Butterworth bandpass filter of order 4 and cutoff frequencies 0.5 - 30 Hz;
- rescale to zero mean and unit variance (also called z-score normalization).

It has to be emphasized that the z-score normalization is a crucial step, without which the model performance degraded significantly as observed in offline experiments which is demonstrated in Figure 1.

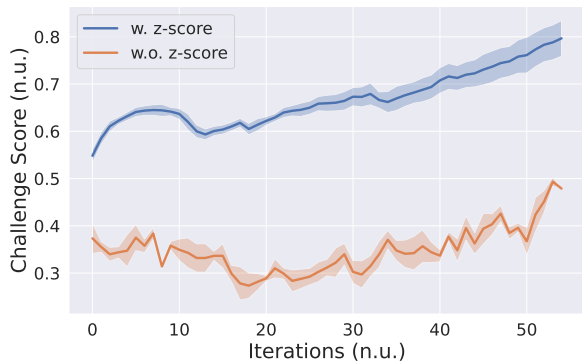


Figure 1: Comparison between mean curves of challenge scores from multiple experiments on the training set with (w.) and without (w.o.) z-score normalization. Shaded areas are error bounds (standard error of the mean).

2.3. Neural Network Model Selection

Considering that there’s a clear and widely accepted (also adopted in the Challenge) mapping from CPC scores to clinical outcomes as follows:

$$\begin{aligned} \text{“Good outcome”} & \text{ for CPC score } \in \{1, 2\}; \\ \text{“Poor outcome”} & \text{ for CPC score } \in \{3, 4, 5\}, \end{aligned} \quad (1)$$

and that the CPC scores are discrete scalars, we regard the problem as a 5-class classification problem, i.e. predicting the 5 discrete CPC scores.

A Time-incremental Convolutional Recurrent Neural Network (TiCRNN) [7] was adopted as the CPC score classifier. Models with CRNN architecture, especially those with SE-ResNet [8] CNN backbones, have proven effective in various physiological signal processing tasks [7, 9]. The building block of the SE-ResNet backbone used in this work consists of 3 sequential convolutional blocks of bottleneck shape followed by an SE (squeeze-and-excitation) block with an extra shortcut connecting the input and output as sketched in Figure 2. The whole architecture of the TiCRNN model used in this work is depicted in Figure 3. It is a sequential model with 1 stem convolutional block which takes preprocessed EEG waveforms as input, 4 SE-Bottleneck blocks, 2 bidirectional LSTM blocks, 1 SE global attention module, 1 adaptive average pooling layer which outputs the feature vectors, and a multi-layer perceptron which takes in the feature vectors and outputs the probability vectors for the 5 CPC scores.

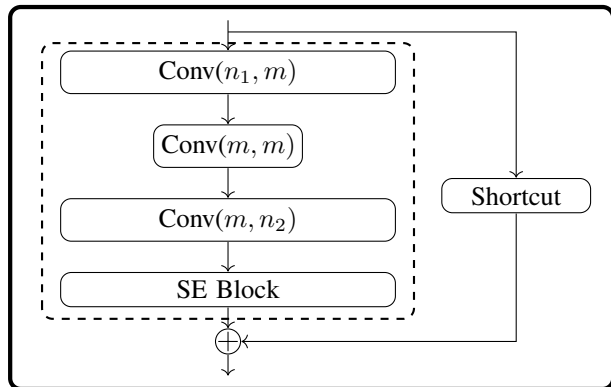


Figure 2: The structure of an SE-Bottleneck block. The mainstream in the dashed box consists of 3 convolutional blocks (actually compositions of convolution, batch normalization and activation) followed by an SE block. The channels n_1, n_2 are typically several times of m , hence giving the name “bottleneck”. The shortcut is typically convolutions of kernel size 1, whose stride and input/output channels match the mainstream.

Probability vectors from multiple EEGs of one patient are averaged and re-normalized via softmax to get the final CPC score predictions. The binary clinical outcomes (good, poor) are obtained by applying the mapping (1).

2.4. Training Strategies

As can be inferred from Figure 4a, We have a highly imbalanced distribution of the learning objective (the CPC score) in the I-CARE dataset, which is also divergent

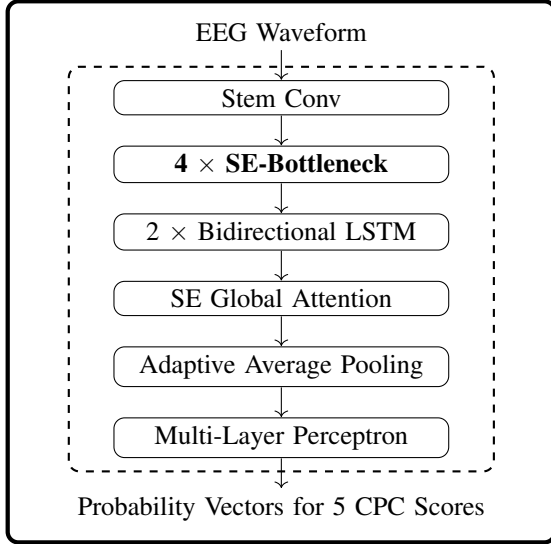
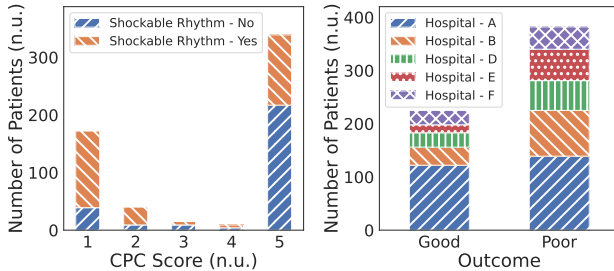


Figure 3: Architecture of the TiCRNN model. Considering that the preprocessed input EEG waveforms have a sampling frequency of 100 Hz, convolutions, except the shortcuts in SE-Bottleneck blocks, have a kernel size of 3. The total number of trainable parameters is 92M.

across different hospitals as illustrated in Figure 4b. Therefore, the asymmetric loss [10], which is the state-of-the-art loss function for multi-label (and also for multi-class) classification problems, was chosen as the minimizing objective. The self-adaptive optimizer AdamW was used in conjunction with the OneCycle scheduler to incrementally update the NN model weights towards their optimal.



(a) Dist. against shockable rhythm. (b) Dist. against hospitals.

Figure 4: Distributions (Dist.) of the ‘‘CPC Score’’ and ‘‘Outcome’’ against 2 typical categorical metadata variables. Missing values for each variable were not counted.

The monitor for model selection is the Challenge score on a left-out validation set which was randomly selected and consisted of 20% of the public training data. The Challenge score is defined as the largest TPR (true positive rate) such that FPR (false positive rate) ≤ 0.05 for the poor clinical outcome prediction. This agrees with the demand for

methods with minimal false-positive rate of poor outcomes as introduced in Section 1.

A key tension for the Challenge objective lies in the existence of the large data and the limited computational resources (RAM and CPU time). As a consequence, further trade-offs were made: we randomly sliced a 3-minute (180 in the number of sample points) piece from each of the selected 5-minute epochs and loaded all the data pieces into memory every 5 training iterations. The rest primary training hyperparameters are collected in Table 1. It should be noted that the maximum number of training iterations was set to 55 since almost all optimal models were obtained before 30 iterations in our offline experiments using much larger total iteration numbers (e.g. 100 iterations).

batch size	max # iterations	early stop patience	learning rate	
			initial	max
32	55	25	2.5e-3	8e-3

Table 1: Primary hyperparameters for NN model training. To avoid overfitting on the training data, early stop callbacks were set which were triggered if the best Challenge score on the left-out validation set stayed unchanged for 25 global iterations. The learning rates were set for the optimizer and its scheduler.

3. Results

The primary Challenge Score, the largest TPR such that $FPR \leq 0.05$ evaluated on the hidden test set up to 72 hours from ROSC, of our team ‘‘Revenger’’ was 0.554, ranked 13th out of 36 teams. This score and auxiliary scores from the truncated subsets of 48h / 24h / 12h from ROSC and the training and hidden validation sets are gathered in Table 2.

Despite the Challenge Score, our solution achieved remarkable results on MAE of CPC score predictions, which is the learning target of our solution. CPC MAE was 1.014 on the hidden test set, ranked 4th out of 36 teams. Corresponding CPC MAE results were 0.957, 1.111, and 1.347 on the truncated subsets of 48h / 24h / 12h from ROSC.²

4. Discussion and Conclusions

As the results presented in Section 3 indicate, the TiCRNN model proposed in this work provides an effective solution to the problem of predicting the level of neurological recovery for comatose patients after cardiac arrest raised by the Challenge. It is relatively simple and lightweight, but still able to attain a TPR as high as 0.554 with FPR, which is vital for the patients in this problem,

²Refer to <https://github.com/DeepPSP/cinc2023/blob/master/images/Revenger-team-digest.pdf> for a full summary table of all metrics and rankings of our team.

		72h after ROSC		48h after ROSC		24h after ROSC		12h after ROSC	
		Score	Rank	Score	Rank	Score	Rank	Score	Rank
Challenge Score	Test	0.554	13 / 36	0.54	14 / 36	0.257	29 / 36	0.252	16 / 36
	Validation	0.701	3 / 36	0.746	1 / 36	0.403	17 / 36	0.328	13 / 36
	Training	0.649	27 / 36	0.702	21 / 36	0.673	12 / 36	0.793	3 / 36

Table 2: Challenge scores and rankings evaluated on the Training and the hidden Test/Validation sets, and on their truncated subsets of 48h / 24h / 12h after ROSC.

suppressed to a very low level (≤ 0.05) for poor clinical outcome prediction. Another highlight of our solution is that we obtained fairly good results on the CPC score predictions. This is not surprising since the CPC score is our direct learning objective. However, treating it as a classification problem ignores the ordinal relations of the CPC scores, which is a potential drawback of this treatment.

As introduced in Section 2.1 and 2.4, we made trade-offs for the limited computation resources by dropping a large proportion of the EEG data. The data used for training the NN models merely constitutes 6 % of the total EEG data. Our team had planned to make full use of the data to train a larger model that learns latent representations from EEGs via unsupervised contrastive learning. However, due to the constraints of time and computation resources, we finally decided to stick to the simple TiCRNN model. Architecture design and unsupervised training mechanisms for large EEG models are left as future research directions.

The computation of SQI for EEGs as described in Section 2.1 is time-consuming. This prohibited us from adding a similar selection procedure in the pipeline of model evaluation on the hidden data and is highly probable to have a negative influence on our overall performance, especially for EEGs which are heavily contaminated with artifacts. Developing a faster and end-to-end SQI computation method would also be a meaningful research problem.

Despite the large EEG data amount, another nice feature of the I-CARE database is that it provides simultaneous physiological signals of other types, including electrocardiogram (ECG) signals, along with the EEGs. This makes it possible to explore and develop multi-modal solutions to the Challenge problem, which is also a topic worth researching but not yet studied in this work.

Acknowledgments

The authors thank Professor Deren Han from Beihang University and Professor Wenjian Yu from BNRist, Tsinghua University for helping accomplish this work.

References

[1] Benghanem S, Pruvost-Robieux E, Bouchereau E, Gavaret M, Cariou A. Prognostication After Cardiac Arrest: How

EEG and Evoked Potentials May Improve the Challenge. *Annals of Intensive Care* 12 2022;12(1).

[2] Cronberg T, Greer DM, Lilja G, Moulaert V, Swindell P, Rossetti AO. Brain Injury After Cardiac Arrest: From Prognostication of Comatose Patients to Rehabilitation. *The Lancet Neurology* 7 2020;19(7):611–622.

[3] 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.

[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] Amorim E, Zheng WL, Ghassemi MM, Aghaeeval M, Kandhare P, Karukonda V, et al. The International Cardiac Arrest Research Consortium Electroencephalography Database. *Critical Care Medicine* 10 2023;(Online ahead of print).

[6] Zheng WL, Amorim E, Jing J, Ge W, Hong S, Wu O, et al. Predicting Neurological Outcome in Comatose Patients After Cardiac Arrest with Multiscale Deep Neural Networks. *Resuscitation* 12 2021;169:86–94.

[7] Kang J, Wen H. A Study on Several Critical Problems on Arrhythmia Detection using Varying-Dimensional Electrocardiography. *Physiological Measurement* 6 2022; 43(6):064007.

[8] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2020;42(8):2011–2023.

[9] WEN H, Kang J. Searching for Effective Neural Network Architectures for Heart Murmur Detection from Phonocardiogram. In *2022 Computing in Cardiology (CinC)*. IEEE, 9 2022; 1–4.

[10] Ridnik T, Ben-Baruch E, Zamir N, Noy A, Friedman I, Protter M, et al. Asymmetric Loss for Multi-Label Classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 10 2021; 82–91.

Address for correspondence:

Hao Wen
No. 17, Qinghua East Road, Haidian District, Beijing, China
wenh06@cau.edu.cn, wenh06@gmail.com