Developing a Machine Learning Pipeline for Predicting Neurological Outcomes in Comatose Cardiac Arrest Survivors Using Continuous EEG Data

Quenaz Soares[1,2], Felipe M Dias[1,2], Estela Ribeiro[1], Jose E Krieger[1] and Marco A Gutierrez[1]

[1] Heart Institute, Clinics Hospital, University of Sao Paulo Medical School, Brazil
[2] Polytechnique School, University of Sao Paulo, Brazil

Abstract

As part of the 'Predicting Neurological Recovery from Como After Cardiac Arrest: The George B. Moody PhysioNet Challenge 2023', we propose a two-step approach, which comprises i) a custom convolutional neural network designed to classify Cerebral Performance Categories (CPC) based on 30-second EEG signals from 19 channels, and ii) random forest models that combine the logits obtained in the previous step with the patient metadata to predict the CPC values and the patient outcome. The final predictions for patient CPC and outcome are determined through a time-weighted average of EEG window prediction. Our submission achieved challenge metric scores of 0.351, 0.386, 0.431, and 0.475 for the prediction outcome at 12h, 24h, 48h, and 72h, respectively. These scores placed the AIMED team in the 20^{th} position in the rank of the official phase. Our study highlights the potential of a machine learning pipeline to predict neurological outcomes in comatose cardiac arrest survivors. This is achieved through a straightforward methodology that combines features extracted from EEG signals with patient metadata.

1. Introduction

Cardiac arrest remains a significant global healthcare challenge, with an annual incidence of over six million cases and a survival rate ranging from 1% to 10% [1]. Among individuals who survive the initial resuscitation, neurological injury emerges as the primary cause of mortality [2]. Therefore, it is imperative to prioritize the prediction and management of post-cardiac arrest neurological outcomes. The majority of patients who survive the initial resuscitation are admitted to intensive care units in a comatose state. The prognostic information derived from their condition plays a pivotal role in treatment decisions. Favorable prognoses often lead to continued care, while unfavorable prognoses may result in the consideration of withdrawing life-sustaining interventions [3, 4].

The George B. Moody PhysioNet Challenge 2023 [5, 6] is dedicated to predicting neurological outcomes in comatose cardiac arrest survivors by analyzing clinical time-series recordings, which include electroencephalogram (EEG) and electrocardiogram (ECG) data. The assessment of neurological function in this context is based on the Cerebral Performance Category (CPC) scale, which ranges from 1 to 5 and is divided into two primary categories: "Good outcome," encompassing CPC scores of 1 or 2; and "Poor outcome," comprising CPC scores of 3, 4, or 5.

In this study, we present a two-step methodology for forecasting the outcomes of comatose cardiac arrest survivors. Initially, we use a custom VGG network with EEG signals to predict patient outcomes. Subsequently, the logits from this classification are combined with patient metadata, and a random forest classifier is employed. The final classification results are obtained through a weighted average across individual time windows for each patient.

2. Methods

We have devised a two-step pipeline that employs a custom VGG architecture, adapted from [7], in conjunction with a Random Forest classifier to predict patient outcomes.

Our overall methodology is succinctly presented in Figure 1. It illustrates the random epoch selection process, the extraction of the initial 30-second windows, the training of a custom VGG model for CPC classification, and the subsequent use of CPC classification logits as features in a Random Forest model. For a more in-depth comprehension, additional details of each step of the methodology will be provided in the subsequent sections.

2.1. Step 1: EEG Feature Extraction

Given the substantial size of the raw dataset, we implemented a sampling technique to reduce training and inference times. Specifically, we randomly selected up to 5 EEG epochs for each patient. During training, we further randomized the choice of up to ten 30-second windows.

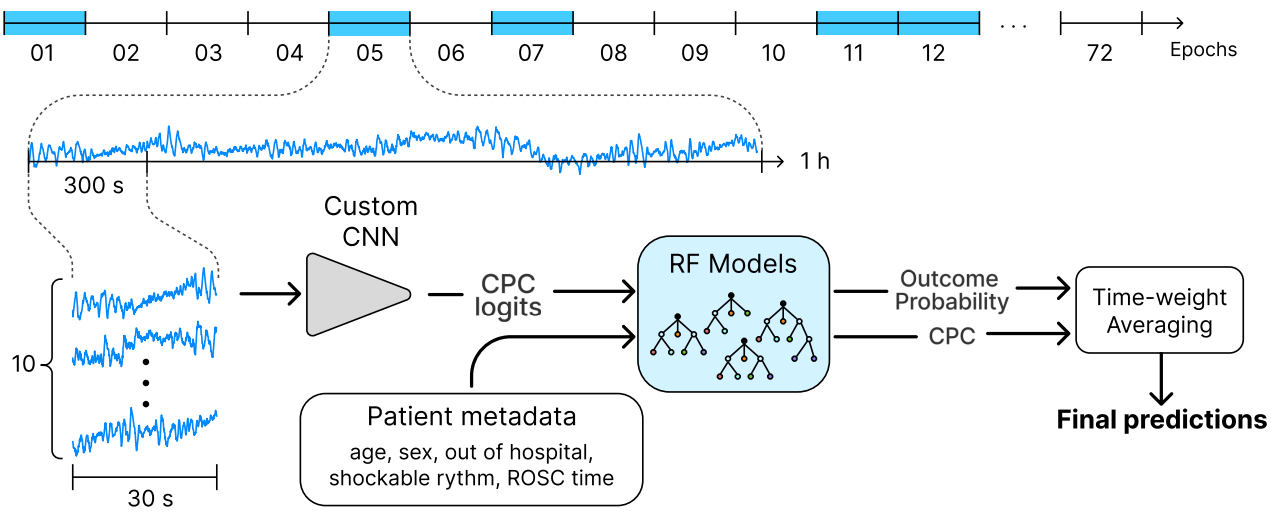

Figure 1: Summary of the proposed methodology for predicting neurological outcomes in comatose cardiac arrest survivor patients.

For inference, we extracted the initial 300-second signals from these epochs and split them into non-overlapping 30-second windows. These signals underwent resampling to 100 Hz and low-pass filtering at 30 Hz using a second-order Butterworth filter. Only the 19 channels consistently present across all EEG signals in the dataset were utilized, namely: C3, C4, Cz, F3, F4, F7, F8, Fp1, Fp2, Fz, O1, O2, P3, P4, Pz, T3, T4, T5, and T6.

Next, we employed a custom VGG network with 13 layers, featuring separable convolutions [7]. The primary objective of this network was to classify the patient's CPC value solely based on 30-second EEG data, without considering patient metadata or temporal information.

The proposed neural network takes a 3,000 x 19 matrix as input, representing a 30-second EEG window with 19 channels. The network consists of 5 convolutional blocks each consisting of a batch normalization layer, two separable convolutional layers, a max pooling layer, and a spatial dropout layer. In the convolutional layers, a filter size of 5 is applied, initiating with 32 filters in the first block and progressively increasing the number of filters by a factor of $\sqrt{2}$ in each subsequent block. After the convolutional blocks, a global max-pooling layer is applied, followed by three fully connected layers comprising 64, 32, and 5 units, respectively. The first two fully connected layers also include dropout layers for regularization. The convolutional layers use leaky ReLU activation functions with an alpha of 0.3, while the fully connected layers employ ReLU activation functions. Figure 2 illustrates the architecture of the network.

To train the neural network, we employed the Adam optimizer with a learning rate of 1E-3, and the training was conducted for 2 epochs using the focal loss [8]. In addition to being trained for CPC value classification, we utilized the logits from the classification as features for our subsequent step.

2.2. Step 2: Random Forest Prediction

As previously mentioned, the feature extraction phase involved extracting the logits derived from the CPC classification using the VGG neural network architecture. In the current stage, we combined the extracted logit features with relevant patient metadata, including age, sex, time from cardiac arrest to return of spontaneous circulation, an indicator for cardiac arrest occurred outside the hospital, and details of targeted temperature management.

Next, we employed two Random Forest (RF) models with ten features as input. The first model was designed for CPC prediction as a regression model, and the second was used to classify patient outcomes. Both RF models were configured with 256 estimators and a constraint of 128 as the maximum allowable number of leaf nodes. In cases where no valid EEG window data were available for a particular patient, RF models were constructed using only the patient metadata as input.

Subsequently, the final prediction for each patient was determined through a time-weighted averaging approach. To compute the time-weighted average of the CPC, each individual CPC predicted obtained from the sample was assigned a $weight$ computed as:

$$weight = \frac{t}{72} + 1 \tag{1}$$

Where t represents the elapsed time, measured in hours, between the return of spontaneous circulation and the sam-

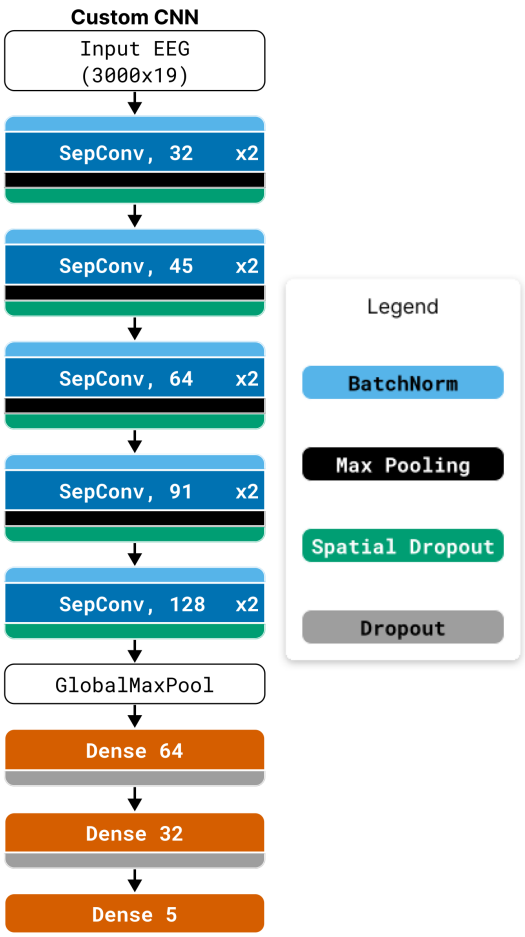

Figure 2: Custom VGG net employed for EEG feature extraction.

ple used for the prediction.

The final time-weighted average for CPC is then calculated as:

$$CPC_{pred} = \frac{\sum_i CPC_{pred_i} \times weight_i}{\sum_i weight_i} \quad (2)$$

The final probability of a poor outcome is calculated in a similar approach to that used for CPC. Subsequently, the final outcome classification is determined by applying a threshold of 0.5 to this probability.

The utilization of the sampling technique in conjunction with the time-weighted approach empowers the proposed pipeline to be adaptable across all the prediction horizons defined within the challenge. Furthermore, the pipeline exhibits the capability to provide predictions for patients with limited EEG data, and even extends its predictive capacity to cases where no EEG data are available.

2.3. Evaluation

The primary objective of the challenge is to assess the effectiveness of the proposed methods in identifying patients with a poor outcome 72 hours after the return of spontaneous circulation.

The challenge metric relies on the true positive rate (TPR) for predicting poor outcomes (CPC of 3, 4, or 5) while ensuring a false positive rate (FPR) of 0.05 or lower at 72 hours following the return of spontaneous circulation. The Challenge Score is computed as the mean TPR calculated for each hospital, defined as follows:

$$TPR = \frac{\sum_{h=1}^{N} TP_{\theta_h}}{\sum_{h=1}^{N} FP_{\theta_h} + \sum_{h=1}^{N} FN_{\theta_h}} \quad (3)$$

Where θ and h represent the decision threshold value and the hospital index, respectively. Additionally, θ_h represents the maximum decision threshold applicable to a specific hospital h.

Apart from the Challenge score metric, we also evaluated the performance of our model using several other metrics, including the Area under the Receiver Operating Characteristic curve (AUC-ROC), the Area under the Precision-Recall curve (AUC-PR), Accuracy (Acc), and F1-score for outcome prediction. In the case of CPC prediction, we provided metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) for assessment.

2.4. Data Split

To comprehensively assess the performance of our methods using the dataset provided by the challenge, we implemented a GroupK-fold validation approach. The groups were defined based on individual hospitals, ensuring that each fold contained data exclusively from a single hospital. Employing this approach allowed us to assess the generalizability and performance of our methods across a wide range of hospital settings.

3. Results

In Table 1, we provide the outcomes derived from our internal k-fold validation approach, as well as those from the train, validation and test sets of the challenge dataset for the 72h time-frame. The challenge score obtained for the test set was employed by the challenge organizers to rank participating teams.

Our model achieved challenge scores of 0.351, 0.386, 0.431, and 0.475 for prediction horizons of 12 hours, 24 hours, 48 hours, and 72 hours, respectively, in the test set during the official phase of the challenge. Notably, our results in the challenge's test set and validation set surpassed those obtained from our internal GroupK-fold cross-validation.

Table 1: Performance metrics for CPC and outcome prediction using our model in interval validation, and the challenge dataset's training, validation and test sets.

Metric	Internal Valid	Train	Valid	Test
AUC-ROC	0.705 ± 0.025	0.957	0.783	0.807
AUC-PRC	0.808 ± 0.079	0.975	0.881	0.882
Acc	0.686 ± 0.046	0.888	0.701	0.768
F1	0.586 ± 0.085	0.877	0.66	0.718
MSE	4.829 ± 1.384	0.854	3.185	3.087
MAE	1.727 ± 0.251	0.684	1.401	1.326
Challenge Score	0.280 ± 0.120	0.851	0.552	0.475

4. Discussion

In this study, we proposed a two-step approach to predict the outcomes for patients in a coma following a cardiac arrest. We observed some discrepancies in our methodology's results when comparing our internal validation set with the official validation and test sets of the CinC Challenge.

The aforementioned disparity can be attributed to our validation approach, which inherently presents a greater challenge. In our internal cross-validation, we ensured that the testing fold of each k-fold iteration contained no patient data from the same hospital as the training fold. We believe that our chosen validation strategy provides a more robust assessment of the generalizability of our method to other hospitals.

Despite achieving a lower challenge score, our proposed model is designed with simplicity in mind, and its generability across hospitals has been consistently evaluated. Nonetheless, we acknowledge that one of the principal challenges of this year's CinC Challenge was handling the substantial volume of provided data. A significant portion of our development effort was directed towards optimizing I/O performance, which limited the time available for refining our model further.

5. Conclusion

Our study demonstrated the potential of a machine learning pipeline to predict neurological outcomes in comatose cardiac arrest survivors by leveraging both continuous EEG data and patient metadata. The proposed model achieved challenge metrics of 0.351, 0.386, 0.431, and 0.475 for the prediction horizons of 12 hours, 24 hours, 48 hours, and 72 hours, respectively. These results positioned us in the 20^{th} place out of the 36 teams during the official phase of the challenge.

Acknowledgments

This research received support from FAPESP (grants n^o 2021/12935-0 and 2014/50889-7), Foxconn Brazil, and Zerbini Foundation as part of the research initiative titled "Machine Learning in Cardiovascular Medicine".

References

[1] Mehra R. Global public health problem of sudden cardiac death. Journal of electrocardiology 2007;40(6):S118–S122.

[2] Ghassemi MM, Amorim E, Alhanai T, Lee JW, Herman ST, Sivaraju A, Gaspard N, Hirsch LJ, Scirica BM, Biswal S, Moura Junior V, Cash SS, Brown EN, Mark RG, Westover MB. Quantitative Electroencephalogram Trends Predict Recovery in Hypoxic-Ischemic Encephalopathy. Critical Care Medicine 2019;47(10):1416–1423.

[3] Nagaraj SB, Tjepkema-Cloostermans MC, Ruijter BJ, Hofmeijer J, van Putten MJ. The revised Cerebral Recovery Index improves predictions of neurological outcome after cardiac arrest. Clinical Neurophysiology 2018; 129(12):2557–2566. ISSN 1388-2457.

[4] Tjepkema-Cloostermans M, da Silva Lourenço C, Ruijter B, Tromp S, Drost G, Kornips FHM, Beishuizen A, Bosch F, Hofmeijer J, van Putten MJAM. Outcome Prediction in Postanoxic Coma With Deep Learning. Critical Care Medicine 2019;47(10):1424–1432.

[5] 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.

[6] Amorim E, Zheng W, Ghassemi M, Aghaeeaval M, Kandhare P, Karukonda V, Lee J, Herman S, Adithya S, Gaspard N, Hofmeijer J, van Putten M, Sameni R, Reyna M, Clifford G, Westover M. The International Cardiac Arrest Research (I-CARE) Consortium Electroencephalography Database. Critical Care Medicine 2023;In press.

[7] Soares QB, Monteiro R, Jatene FB, Gutierrez MA. A Lightweight Unidimensional Deep Learning Model for Atrial Fibrillation Detection. In 2022 Computing in Cardiology (CinC), volume 498. IEEE, 2022; 1–4.

[8] Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision. 2017; 2980–2988.

Address for correspondence:

Quenaz Soares
Heart Institute (InCor) - Av. Dr. Enéas Carvalho de Aguiar, 44 - Cerqueira César, São Paulo - SP, Brazil
quenaz.soares@hc.fm.usp.br