HyperEnsemble Learning from Multimodal Biosignals to Robustly Predict Functional Outcome after Cardiac Arrest

Morteza Zabihi[1], Alireza Chaman Zar[1,2], Pulkit Grover[2], Eric S. Rosenthal[1]

[1]Massachusetts General Hospital, Boston, MA, USA
[2]Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

Aligned with the George B. Moody PhysioNet Challenge 2023, we aim to create an open-source algorithm to predict coma recovery 3-6 months following return of spontaneous circulation after cardiac arrest. Our algorithm employs multimodal biosignals from a dataset spanning five U.S. and European hospitals. The preprocessing pipeline enhances signal quality through bandpass and notch filtering, resampling, and z-score normalization. EEG channels are organized into 21 bipolar montages and standardized ECG lead selection. We extract 362 EEG expert-based features, encompassing temporal, spectral, time-frequency, and connectivity metrics. ECG features include 16 heart rate variability indices, features indicating atrial tachyarrhythmia, and shockable rhythms. Our classification approach involves initial positional encoding, followed by an ensemble of CatBoost combined with stacked ensembles with diverse base learners in the first layer. To improve robustness and reduce bias, our ensemble strategy employs various cross-validation committees, including confounder-isolating cross-validation. Our proposed predictive framework demonstrates an average challenge score mechanism of 0.651 ± 0.077 through 5-fold cross-validation. Our "AIrhythm" scored 0.792 on the hidden test set and officially ranked 1st in the PhysioNet Challenge 2023.

1. Introduction

Predicting functional outcomes of comatose cardiac arrest patients demands timely, precise, and robust approaches that inform clinical decision-making [1]. Traditional prognostic approaches, like visual inspection of electroencephalography (EEG) by neurophysiologists, are limited by subjectivity, risking bias and lack of reproducibility [2]. Machine learning (ML) for EEG analysis [3], [4] is increasingly utilized, and the opportunity exists for simultaneously using ECG cardiac telemetry and EEG neurotelemetry to predict functional outcomes [5].

We sought to utilize longitudinal EEG and ECG recordings available in cardiac arrest patients during the initial 72 hours following return of spontaneous circulation (ROSC) to predict a Cerebral Performance Category (CPC) scale dichotomous outcome: "good" (CPC 1-2) or "poor" (CPC 3-5). We introduce an ML framework (Figure 1) for predicting functional outcomes that integrates multimodal features, introducing some novel handcrafted EEG features, positional encoding via gradient boosting trees, confound-isolating cross-validation (CV) strategies, and a combination of stacking and boosting ensembles (HyperEnsemble).

2. Materials and methods

The training dataset included 607 adult cardiac arrest patients in coma, monitored continuously for hours to days with EEG and ECG, as previously described [6]–[8].

2.1. Preprocessing

EEG and ECG signals were bandpassed (0.1-45 Hz), resampled at 125 or 128 Hz, z-score normalized, and segmented into non-overlapping 3-minute segments. Given ECG lead inconsistencies, we integrated the first lead into our framework. We created a range of effective EEG channel montages through empirical validation: F3-C3, C3-P3, F4-C4, C4-P4, Fz-Cz, P3-O1, T5-O1, P4-O2, T6-O2, Fp1-Fp2, T3-T4, T5-T6, Fp1-T3, Fp2-T4, T4-O1, T3-O2, Fp1-Cz, Fp2-Cz, O1-O2, P3-P4, and Cz-Pz.

2.2. EEG quantitative measures

Single-lead measures: We extracted 43 features from 6 EEG channels, C3-P3, C4-P4, Fp1-Fp2, Fz-Cz, T3-T4, and Cz-Pz, including various temporal (Hjorth Activity, Mobility, and Complexity) and spectral (power below 1Hz [9], power ratios of 4-12/12-30Hz, 4-12/8-35Hz, and the alpha-to-delta power ratio features. Mean and variance were also computed for signal components across varying scales through wavelet decomposition, employing Daubechies 4 wavelets with level 5. We also extracted the interquartile range of the first difference of zero-crossing

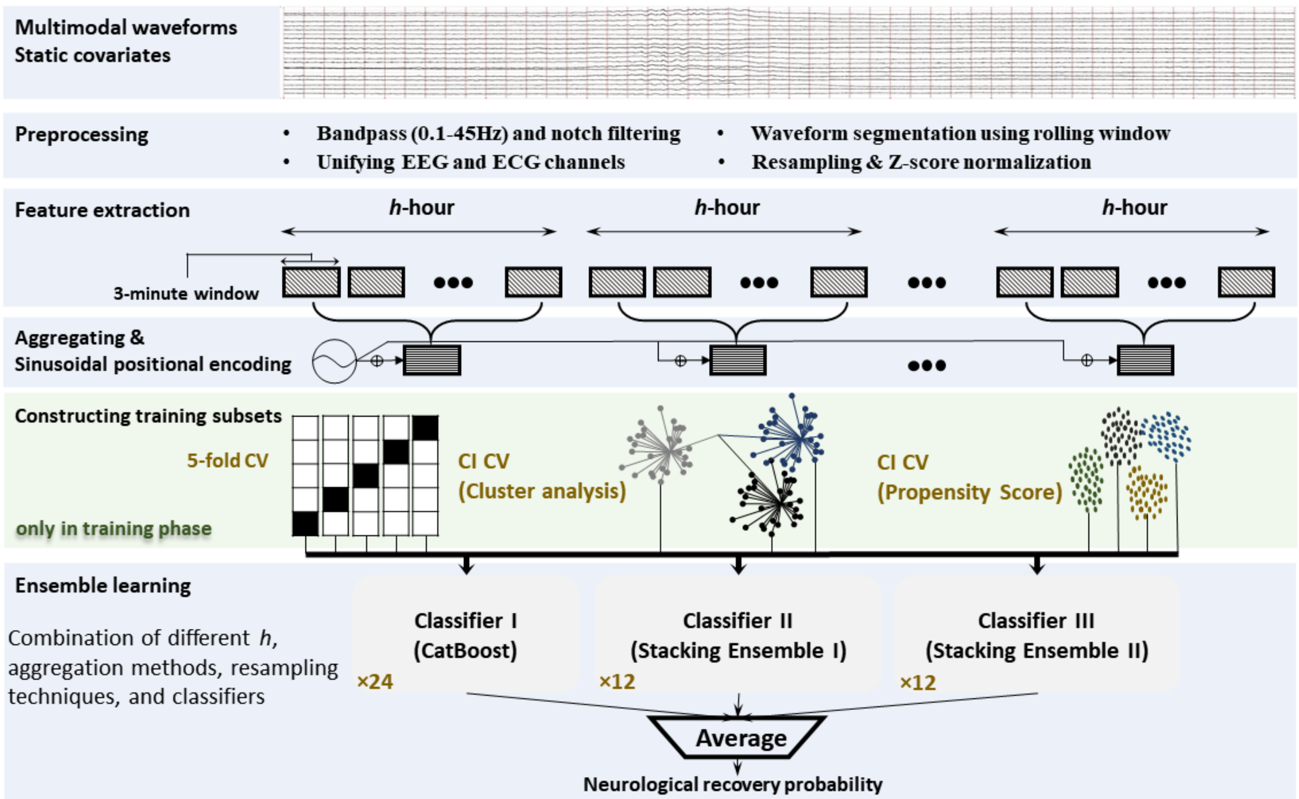

Figure 1. In the proposed predictive framework, **preprocessing (1)** and **feature extraction (2)** from 3-minute windows were followed by **aggregation and sinusoidal positional encoding (3)**, wherein quantiles of features were computed at every *h*-hour intervals and aggregated windows were assigned a sinusoidal pattern. **Training set construction (4)** involved conventional and confound-isolating (CI) cross-validation. Finally, **HyperEnsemble learning** (5) leveraged diverse classifiers from data grouped using varying quantiles and *h* values to generate an averaged final prediction probability.

positions in the raw and the 0.5-4Hz band-passed segments. Additionally, spectral features, such as the frequency corresponding to the maximal magnitude and the variance of the spectrogram, were extracted. Furthermore, the mean and standard deviation of the instantaneous frequency were extracted, providing insights into rhythmic patterns. The framework also incorporates the coefficient of Burg's linear regression method to model the EEG segment as an autoregressive process to afford a window into EEG temporal interdependencies. Additionally, the spectral edge frequencies corresponding to diverse cumulative power levels (50, 70, 80, 90, and 95%) were computed. We also extracted seasonal autocorrelation at different lags (0.3, 0.8, and 1.5 seconds).

Connectivity measures: For all 21 EEG channels, we derived eigenvalues from the correlation matrix, the EEG covariance matrix, and the autocorrelation matrix evaluated at lags of 2, 8, and 32 seconds. Furthermore, we explored the Phase Locking Value (PLV) to quantify the phase synchrony exhibited between pairs of six channels mentioned above within 0.5-4 and 16-25 Hz. Additionally, we computed relative alpha rhythm power and spectral slope for electrode pairings of (Fp1-Fp2, O1-O2) and (Fp1-T3, Fp2-T4) to examine both frontal-occipital and right-left hemisphere interactions.

Frontal EEG measures: Given the role of frontal channels in consciousness studies [10], we extracted three features from Fp1-T3, Fp2-T4, Fp1-Cz, and Fp2-Cz, including 1) the logarithm of the ratio between the average power observed within the frequency ranges of 30-47 and 11-20 Hz; 2) the logarithm of the ratio between the spectral energy within 30-42.5 and 6-12 Hz; and 3) the computed slope of the power spectral density on a logarithmic scale.

Temporal dynamics: To model the temporal information, two discrete features were calculated at Fz-Cz: 1) the signature [11] of the reconstructed phase space of the EEG signal, utilized to capture interactions of the highly oscillatory components and 2) Normalized Compression Distance [12] using the Gzip compression to model the temporal complexity between the consecutive rolling windows. To our knowledge, these two features have not been utilized previously for EEG analysis.

2.3. ECG quantitative measures

Heart Rate Variability (HRV) measures: The first 10 features are HRV measures derived from R-R intervals. Two features gauge the occurrences of abrupt R-R interval changes exceeding a 50-millisecond threshold, and the percentage of these among all R-R intervals. Subsequently, we computed SD1, SD2, the ratio SD1/SD2 and SD2/SD1, capturing both short- and long-term HRV [13]. Further, we extracted the cardiac sympathetic index by relating the dimensions of an ellipse defined by SD1 and SD2 and the Cardiac Vagal Index by logarithmically transforming the product of SD1 and SD2. A modified variation was utilized by computing the square of SD1 divided by SD2. Two additional features included the squared differences between R-R intervals and the ratio of the standard deviation to the mean of R-R differences [14].

Atrial tachyarrhythmia measures: We also extracted three features for identifying atrial tachyarrhythmia: AFEvidence, ATEvidence, and OrgIndex [15].

Shockable measures: We extracted three features inspired to identifying shockable rhythms [16]: 1) the difference between the maximum and minimum values of the band passed (6.5-30 Hz) ECG segment; 2) the proportion of time for which the first difference squared ECG segment remains below a predefined threshold; and 3) a measure in which the cumulative sum of the normalized power spectrum was first calculated, after which bandwidth was computed by discerning frequencies at which a given proportion of power is enclosed.

2.4. Static covariates

Four static variables were used as predictors and carried forward across rolling windows, including age, ROSC, out-of-hospital cardiac arrest (OHCA), and shockable rhythm. Sex and hospitals were considered confounders in this study.

2.5. Classification framework

Aggregating and positional encoding: For every *h*-hour interval (*h*=5, 6, or 12), the 3-minute sequential segments were combined using quantiles (0.88 or 0.89) derived from individual feature values. Following the aggregation, sinusoidal positional encoding [17] was employed and added to feature values to imbue the feature vectors with the relative temporal order.

Constructing training subsets: To build diverse training sets for our proposed ensemble learning approach, we utilized three distinct CV strategies: 1) Five-fold CV generating 5 replicated datasets from the initial training data formed by withholding a random 20% subset to produce 5 intersecting training subsets; 2) Confounder-isolating [18] CV (cluster analysis) utilizing k-means clustering to group training data based on potential confounding variables, namely hospital type and sex, which led to three overlapping training subsets from the original dataset, where each iteration removed a distinct cluster; and (3) Confounder-isolating CV (propensity score), a variant employing logistic regression with predictors like hospital types and sex to predict outcomes then stratifying the population into four non-overlapping quantiles of the predicted probabilities and subsequently generating four intersecting training subsets by excluding a distinct group each time. By constructing these training subsets, we aimed to enhance robustness and reduce bias.

HyperEnsemble learning: Random undersampling was used to mitigate class imbalance within each subset, with each resulting subset inputted into three ensemble classifiers, whose probabilities were averaged to calculate final prediction probabilities. The first ensemble classifier employed was CatBoost, an ML method based on gradient boosting over decision trees [19]. The second ensemble was a stacking ensemble, wherein the base learners were a multilayer perceptron (MLP) comprising three hidden layers of 20 neurons, a support vector machine with a linear kernel, a CatBoost, extremely randomized trees, and a linear discriminant analysis classifier. The meta-learner was an MLP with two hidden layers, each with five neurons. The third classifier was another stacking ensemble with three CatBoost classifiers as base learners and logistic regression as the meta-learner. Apart from the base learners themselves, a pivotal distinction between this stacking ensemble and the previous ensemble resides in the nature of input data being generated by employing distinct aggregating time resolutions, i.e., *h*=5, 6, and 12 hours.

3. Results

Performance of the proposed framework for two variations of CV is shown in Table 1. CV_I tested patientwise 5-fold CV. CV_{II}, tested CV on data isolated from one hospital and 20% of the remaining patients.

Table 1. Results of 5-fold cross-validation. The *challenge score i*s the true positive rate at a false positive rate of 0.05.

Metrics	CV_I	CV_{II}	Test Set
Challenge score	0.691±0.083	0.651±0.077	0.792
Sensitivity	0.80±0.07	0.78±0.06	*NR*
Specificity	0.80±0.08	0.79±0.05	*NR*
AUROC	0.89±0.02	0.87±0.02	0.916
AUPRC	0.93±0.01	0.91±0.02	0.957
F1	0.80±0.04	0.79±0.03	0.791

AUPRC, area under the precision-recall curve; *AUROC*, area under the receiver operating characteristic curve. *NR*, not reported by the Physionet Challenge 2023.

4. Discussion and conclusion

We developed an open-source framework for predicting the outcomes of comatose cardiac arrest patients. Our approach included established and novel EEG and ECG features with positional encoding using non-deep learning models. By adapting ensemble learning and implementing diverse CV committees, we increased the robustness of our framework. The proposed solution focuses on a restricted set of EEG channels and features due to the computational constraints set by the PhysioNet challenge; a broader range of EEG channels and additional features may yield improved performance. Future directions include evaluating the response of clinicians and surrogates to simulated predictions and silent real-time evaluation.

References

[1] E. Westhall *et al.*, "Standardized EEG interpretation accurately predicts prognosis after cardiac arrest," *Neurology*, vol. 86, no. 16, pp. 1482–1490, Apr. 2016, doi: 10.1212/WNL.0000000000002462.

[2] A. O. Rossetti, A. A. Rabinstein, and M. Oddo, "Neurological prognostication of outcome in patients in coma after cardiac arrest," *Lancet Neurol*, vol. 15, no. 6, pp. 597–609, May 2016, doi: 10.1016/S1474-4422(16)00015-6.

[3] M. C. Tjepkema-Cloostermans *et al.*, "Outcome Prediction in Postanoxic Coma With Deep Learning*," *Crit Care Med*, vol. 47, no. 10, pp. 1424–1432, Oct. 2019, doi: 10.1097/CCM.0000000000003854.

[4] W.-L. Zheng *et al.*, "Predicting neurological outcome in comatose patients after cardiac arrest with multiscale deep neural networks," *Resuscitation*, vol. 169, pp. 86–94, Dec. 2021, doi: 10.1016/j.resuscitation.2021.10.034.

[5] S. Bin Kwon, M. Megjhani, D. Nametz, S. Agarwal, and S. Park, "Heart rate and heart rate variability as a prognosticating feature for functional outcome after cardiac arrest: A scoping review," *Resusc Plus*, vol. 15, p. 100450, Sep. 2023, doi: 10.1016/j.resplu.2023.100450.

[6] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.

[7] Amorim E, et al., The International Cardiac Arrest Research (I-CARE) Consortium Electroencephalography Database. Critical Care Medicine 2023 (in press); doi:10.1097/CCM.0000000000006074.

[8] M. A. Reyna *et al.*, "Predicting Neurological Recovery from Coma After Cardiac Arrest: The George B. Moody PhysioNet Challenge 2023," in *Computing in Cardiology*, 2023.

[9] L. J. Hirsch *et al.*, "American Clinical Neurophysiology Society's Standardized Critical Care EEG Terminology: 2021 Version," *Journal of Clinical Neurophysiology*, vol. 38, no. 1, pp. 1–29, Jan. 2021, doi: 10.1097/WNP.0000000000000806.

[10] U. León-Domínguez and J. León-Carrión, "Prefrontal neural dynamics in consciousness," *Neuropsychologia*, vol. 131, pp. 25–41, Aug. 2019, doi: 10.1016/j.neuropsychologia.2019.05.018.

[11] T. Lyons, "Rough paths, Signatures and the modelling of functions on streams," May 2014.

[12] Z. Jiang, M. Yang, M. Tsirlin, R. Tang, Y. Dai, and J. Lin, "'Low-Resource' Text Classification: A Parameter-Free Classification Method with Compressors," in *Findings of the Association for Computational Linguistics: ACL 2023*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 6810–6828. doi: 10.18653/v1/2023.findings-acl.426.

[13] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Front Public Health*, vol. 5, Sep. 2017, doi: 10.3389/fpubh.2017.00258.

[14] M. Zabihi, A. Bahrami Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection of Atrial Fibrillation in ECG Hand-held Devices Using a Random Forest Classifier," Sep. 2017. doi: 10.22489/CinC.2017.069-336.

[15] S. Sarkar, D. Ritscher, and R. Mehra, "A Detector for a Chronic Implantable Atrial Tachyarrhythmia Monitor," *IEEE Trans Biomed Eng*, vol. 55, no. 3, pp. 1219–1224, Mar. 2008, doi: 10.1109/TBME.2007.903707.

[16] U. Irusta, J. Ruiz, E. Aramendi, S. Ruiz de Gauna, U. Ayala, and E. Alonso, "A high-temporal resolution algorithm to discriminate shockable from nonshockable rhythms in adults and children," *Resuscitation*, vol. 83, no. 9, pp. 1090–1097, Sep. 2012, doi: 10.1016/j.resuscitation.2012.01.032.

[17] A. Vaswani *et al.*, "Attention is All You Need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA, 2017, pp. 6000–6010.

[18] D. Chyzhyk, G. Varoquaux, M. Milham, and B. Thirion, "How to remove or control confounds in predictive models, with applications to brain biomarkers," *Gigascience*, vol. 11, Mar. 2022, doi: 10.1093/gigascience/giac014.

[19] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," Jun. 2017.

Address for correspondence:

Morteza Zabihi.
Massachusetts General Hospital, Boston, MA
mzabihi@mgh.harvard.edu