Deep-Learning-Assisted Prediction of Neurological Recovery from Coma After Cardiac Arrest

Vasanth Kumar Babu[1], Navneet Roshan[1], Rahul Pandit[1]

[1] Centre for Condensed Matter Theory, Department of Physics, Indian Institute of Science, Bangalore, India

Abstract

We develop a deep-learning-based algorithm to predict the probability of recovery of a comatose patient who has suffered a heart attack by analyzing electroencephalogram (EEG) and electrocardiogram (ECG) data. These have been provided to participants in the George Moody Physionet Challenge (2023); our team name is RPG@IISC. Given EEGs and ECGs, we extract, from hour-long traces for each patient, the burst-suppression (BS) rate, inter-channel EEG correlations, time intervals between successive peaks of ECG, and associated heart variability rate (HVR) metrics. We also use other information provided, e.g., patient age, sex, return of spontaneous circulation (ROSC), in-hospital or out-of-hospital cardiac arrest, presence of a shockable rhythm, and targeted temperature management. With these features, we then use combinations of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to make predictions of (a) the probability of recovery $\mathcal{P}$ and (b) the cerebral performance category (CPC), at hourly scales; we then combine these hourly results to predict final values for $\mathcal{P}$ and CPC. In the official phase, when evaluated at 72 hours after ROSC, the score obtained by our algorithm on the hidden-validation data and hidden-test data is 0.63, and 0.43(ranked 24^{th}), respectively.

1. Introduction

After a heart attack, a significant number of patients become comatose [1]. During this comatose period, an efficient diagnosis of the condition of the patient is crucial for predicting the probability of recovery $\mathcal{P}$ and the optimal treatment. Hence, it is essential to collect patient data continuously for careful analysis. Electroencephalography (EEG) recordings from patients play a vital role here [2–4]. However, they lead to very large data sets, whose complexity and analysis pose considerable challenges for all except a limited number of specialist clinicians. Therefore, automated-data-analysis techniques can help greatly to extract $\mathcal{P}$ from these recordings.

Some recent studies have obtained machine-learning-based predictions of outcomes for post-heart-attack comatose patients [see, e.g., Refs. [5, 6]]. These approaches focus on extracting features from EEG data sets, which are then used to predict neurological recovery. We build on these studies by developing an efficient machine-learning algorithm that extracts features from the EEG, ECG, and other patient data to predict $\mathcal{P}$ and the cerebral performance category (CPC) [7], for a patient who is recovering from a coma after a heart attack, at hourly scales; we then combine these hourly results to predict the final values for $\mathcal{P}$ and CPC. For our analysis, we use the data that have been provided to participants in the 2023 George Moody Physionet Challenge [8, 9]; our team name is RPG@IISC.

2. Methods

The challenge database [10] consists of information on 1020 patients; data for 607 patients have been made available to develop models; the remaining data (hidden) have been used for the validation and testing of these models. To develop our machine learning algorithm, we use some commonly studied metrics that are used to analyze EEG and ECG data. During the unofficial phase, 18-channel EEG data were provided; and in the official phase, 22-channel EEG data [Fp1, Fp2, F7, F8, F3, F4, T3, T4, C3, C4, T5, T6, P3, P4, O1, O2, Fz, Cz, Pz, Fpz, Oz, and F9 channels] and 5-channel ECG data [ECG, ECG1, ECG2, ECGL, and ECGR] were provided. We use the following metrics:

1. Burst-suppression [11, 12] (BS) patterns, with the BS rate [12], in an interval of EEG, quantified as

$$\text{BS rate} = \frac{\text{time EEG spends below a specified threshold}}{\text{total time interval}} \tag{1}$$

2. We quantify inter-channel EEG correlations using Pearson's correlation [13].

3. We measure the EEG background activity (EBA) in our EEG data [14, 15] with power spectral density of α, β, δ,

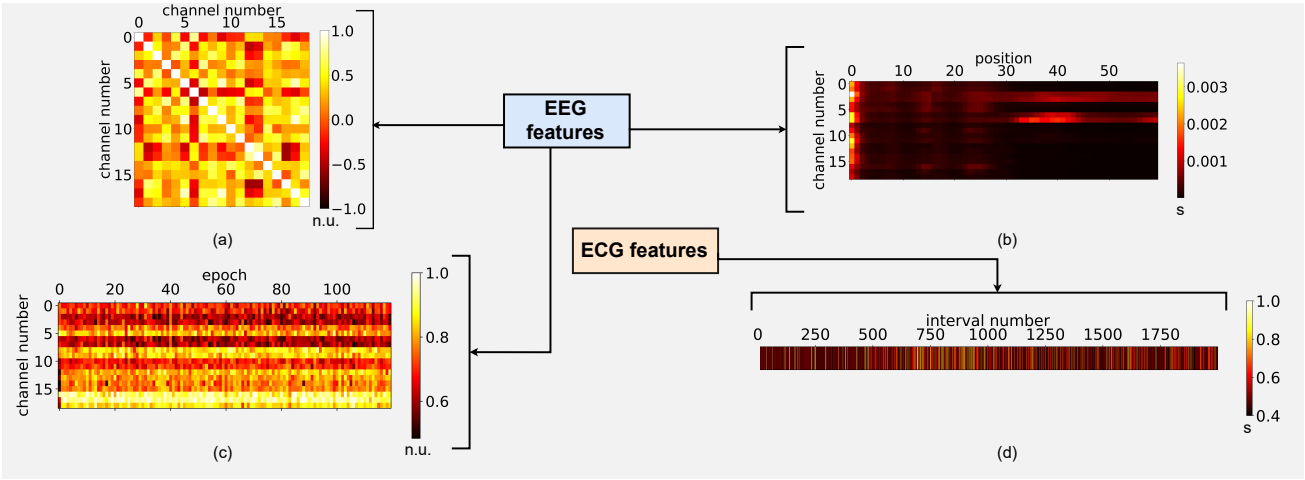

Figure 1. Representative EEG and ECG features: (a) Inter-channel EEG correlations [here a 19×19 matrix, EEG channel numbers $0-18$ label channels Fp1, Fp2, F7, F8, F3, F4, T3, T4, C3, C4, T5, T6, P3, P4, O1, O2, Fz, Cz, Pz arranged in this order]; (b) EEG background activity: the first 60 entries of the power-spectral densities of δ, θ, α, and β waves, concatenated in this order [i.e., for 19 channels we have a 19×60 matrix]; (c) the burst-suppression (BS) rate matrix for 19 channels and the first 120 epochs (a 19×120 matrix); (d) 2000 entries of time intervals (RR-Intervals) between successive peaks in the ECG data, separated by at least 0.4s.

and θ waves with frequency ranges $8-12$ Hz, $12-30$ Hz, $0.5-8$ Hz, and $4-8$ Hz respectively.

4. ECG data: we calculate the time intervals between successive peaks, separated by at least 0.4 s (RR-intervals), and use these to quantify heart rate variability (HRV) [16] metrics, e.g., the mean, standard and root-mean-square deviations of these intervals, and the percentage of adjacent intervals that differ by 50 ms.

5. Additional patient features: age [given as an integer]; sex [100-female, 010-male, and 001-other]; return of spontaneous circulation (ROSC) in minutes; in-hospital or out-of-hospital cardiac arrest; the presence of a shockable rhythm [0 if false and 1 if true]: and targeted temperature management (TTM) [33, 36, or NaN for no TTM]; i.e., a total of 8 entries.

6. Our output data employ two metrics to quantify the recovery: (1) Patient outcome: 0 if the patient recovers favorably (good outcome) and 1 otherwise; (2) cerebral performance category (CPC) [7]: 1 - good recovery; 2 - moderate disability; 3 - severe disability; 4 - unresponsive wakefulness syndrome; 5 - death. [CPCs 1 or 2 - good outcome; and CPC of 3, 4, and 5 - poor outcome].

We use 1D Convolutional neural networks (CNNs) and Long-short-term memory networks (LSTMs) [Fig. 2]. In our CNNs, we first have a convolutional layer with 256 filters, kernel size 2×2, and a max-pool layer with pool size 2×2 and the ReLU (Rectified Linear Unit) activation function. We then add a convolutional layer with 64 filters, with kernel size 2×2 and the ReLU activation function; the flattened output from the previous layer is passed through a dense layer of 64 nodes, followed by another dense layer of 32 nodes, and with 2 nodes in the final layer. We use the mean-squared-error loss function and the Adam Optimizer to update weights. We use variable length, single-layered LSTM with 128 nodes and 2 output nodes, with the mean-squared-error loss function and the Adam Optimizer to update weights. The CNN and LSTM architectures are the same throughout (not optimized) for simplicity. We implement these algorithms with TensorFlow [17].

To train our CNNs for $\mathcal{P}_t$, the probability of recovery, and $\mathcal{C}_t$ (CPC), based on data provided at hour t [with each hour of the patients' data, we assign their output data as training labels], we use EEG features $1-3$ as inputs and combine all the channel features for these inputs. We train separately for the features from each channel of ECG, as the available channel data vary significantly from patient to patient. We train with an array of intervals of successive differences of peaks, separated by at least 0.4 s, from the ECG as inputs to our CNNs. Below, we outline different methods we use in this study to arrive at the final outputs.

- M1: In an hour, we consider the cleanest 5 minutes of EEG data in the unofficial phase of the challenge; this requires 30000×18 matrices [30000 points in the time series recording (100 Hz sampling frequency) for each of the 18 EEG channels]. We reduce the dimensionality of the data by extracting EEG features. To get the BS rate, we consider 30 epochs [30 intervals of 10s with 1000 data points each] to calculate the fraction of the EEG below the 1.56mV threshold [not optimized]. With all the 18 EEG channel data combined, we have a 30×18 representation.

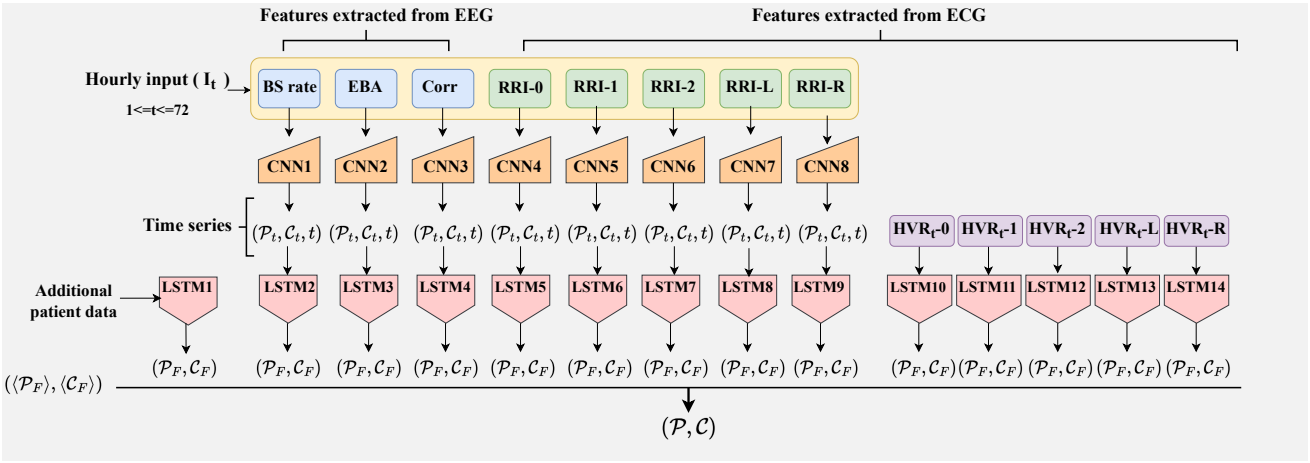

Figure 2. Schematic diagram of our algorithm: The features extracted from EEG data: burst-suppression (BS) rate, background activity (EBA), inter-channel correlation matrix (Corr), and the array of time intervals between two successive peaks in the ECGs (RR-intervals: RRI-0, RRI-1, RRI-2, RRI-L, RRI-R corresponds to five different ECG channels), are fed as inputs into the 1D CNNs (CNN1-CNN8) from the hourly data [I_t stands for input at hour t]; CNN1-CNN8 are trained to predict the probability of recovery $\mathcal{P}_t$ and the CPC, i.e., C_t at hour t. We feed $(\mathcal{P}_t, \mathcal{C}_t, t)$ into LSTM2-LSTM9 and train them to predict the final outcome for each feature-$(\mathcal{P}_F, \mathcal{C}_F)$. LSTM10-LSTM14 are trained to predict $(\mathcal{P}_F, \mathcal{C}_F)$ with heart rate variability metrics [HRV_t-0, HRV_t-1, HRV_t-2, HRV_t-L, HRV_t-R, for 5 ECG channels] as inputs. LSTM1 is trained with additional patient data (age, sex,.. etc.) as inputs. Then we average over all the features (only EEG or both EEG and ECG features): $\langle \mathcal{P}_F \rangle$ and $\langle \mathcal{C}_F \rangle$, to get the final probability of recovery $\mathcal{P}$ and CPC ($\mathcal{C}$)).

For each channel, we concatenate into one vector of length 85 the α, β, δ, and θ power spectral densities (PSDs) calculated via the Welch method; thus, 18 channels yield an 85×18 matrix. Inter-channel correlations yield an 18×18 matrix. We replace all missing data with their average values. We feed $\mathcal{P}_t$, $\mathcal{C}_t$ and t into our LSTM networks and train them to obtain the predictions $\mathcal{P}_F$ and $\mathcal{C}_F$ for each feature. Furthermore, we use LSTMs to predict patient outcomes for the additional patient features. Then we average over all the features [$\langle \mathcal{P}_F \rangle$ and $\langle \mathcal{C}_F \rangle$ in Fig. 2] to get the final probability of recovery $\mathcal{P}$ and CPC ($\mathcal{C}$).

- M2: The input features for M2 are the same as those for M1. We split the training data into two parts and trained two separate sets of CNNs to make the hourly predictions $\mathcal{P}_t$ and $\mathcal{C}_t$. The CNNs are trained on one set; they are then used to predict $\mathcal{P}_t$ and $\mathcal{C}_t$ on the other set. The LSTMs are then trained on these data to predict the final outcome for each feature. We use the same strategy as in M1 to get the final $\mathcal{P}$ and CPC ($\mathcal{C}$)).
- M3: In an hour, we consider EEG inputs up to a maximum size of 600000×19 [600000 points in time series recording (sampling frequency from $250 - 528$ Hz) for each of the 19 EEG channels] as the available data sizes vary in the official phase. The EEG signal is then band-pass filtered; this removes frequencies outside 0.1Hz-30Hz [we apply a notch filter to remove the utility frequency if it lies in the range 0.1Hz-30Hz] and reduced to a maximum size of a 300000×19 matrix [normalized in the range [-1,1], with a re-sampled frequency of 128 Hz]. We use the BS rate for 120 epochs with 0.25 as the threshold [which gave the best results] for each 10s interval; the combined data for all channels yields a 120×19 [Fig. 1(c)] matrix. EBA is quantified as in M1; the input size of the EEG time series is not constant, so the PSD vector has a variable length; we considered the first 60 entries for each channel (not optimized) yielding a 60×19 matrix [Fig. 1(b)]; the inter-channel correlation matrix yields a 19×19 matrix [Fig. 1(a)]. The final $\mathcal{P}$ and CPC are inferred as in M1.
- M4: EEG input features are the same as that of M3; in addition, we consider ECG data from 5 channels. We calculate the first 2000 intervals between peaks [Fig. 1(d)] in the hourly ECG data and then train the CNNs to output $\mathcal{P}_t$ and $\mathcal{C}_t$. We then train LSTMs with $\mathcal{P}_t$, $\mathcal{C}_t$ and t, as inputs, to predict $\mathcal{P}$ and CPC. We also train some LSTMs to predict outcomes with HRV and time t. We then average these predictions as in M1.
- Other methods: We have also looked at methods in which the final predictions are obtained by taking the averages of $\mathcal{P}_t$ and $\mathcal{C}_t$, over both times and features [for both ECG and EEG features]; however, the performances of these methods are inferior to M1 and M3.

3. Results

In the challenge, the algorithms were evaluated with the true positive rate at a false positive rate of 0.05 as the scor-

Method	Training	Validation	test
M3	0.41	0.63	0.43 (rank: 24)
M1	0.45	0.61	-
M2	0.32	0.38	-
M4	0.27	0.28	-

Table 1. Scores [true positive rate at the false positive rate of 0.05] on the training-cross-validation data, hidden-validation and test data for methods M1-M4 when evaluated at the end of 72 hours.

ing metric [9]. When trained on the cleanest 5-minute data set [unofficial phase] and evaluated at 72 hours after ROSC, our method M1 gave a score of 0.45 on the training cross-validation data and a score of 0.61 on the hidden validation data. On replacing M1 with M2, these scores deteriorated to 0.32 and 0.38, respectively, so we did not use method M2 during the official phase. During the official phase, we used method M3, with only EEG data, and M4, with both EEG and ECG data. At the end of the first 72 hours after ROSC, we obtained a score of 0.63, with M3, and 0.28, with M4, on the hidden-validation data [see Table 1]. M3 obtained a score of 0.43 (ranked 24^{th}) on the hidden-test data.

4. Discussion and Conclusions

We have explored our models with only EEG features and with both EEG and ECG features. Our algorithms have performed best with EEG features, and the inclusion of ECG data has not increased their performance; this is consistent with most of the other teams, as the majority of the algorithms presented in the challenge used only EEG data. Our work can be extended by tuning model parameters, optimizing CNN and LSTM architectures, and by including feature cross-correlations.

Acknowledgments

We thank S.S.V.Kolluru and J.K.Alageshan for discussions, the National Supercomputing Mission (NSM) and SERB (India) for support, and SERC (IISc) for computational resources.

References

[1] 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.

[2] Amorim E, et al. Continuous EEG monitoring enhances multimodal outcome prediction in hypoxic–ischemic brain injury. Resuscitation 2016;109:121–126.

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

[4] Ruijter BJ, et al. Early Electroencephalography for Outcome Prediction of Postanoxic Coma: A Prospective Cohort Study. Annals of Neurology 2019;86(2):203–214.

[5] Zheng WL, Amorim E, et al. Predicting neurological outcome in comatose patients after cardiac arrest with multiscale deep neural networks. Resuscitation 2021;169:86–94.

[6] Tjepkema-Cloostermans MC, et al. Outcome Prediction in Postanoxic Coma With Deep learning. Critical Care Medicine 2019;47(10):1424–1432.

[7] Jenett B, Bond M. Assessment of outcome after severe brain damage: a practical scale. Lancet 1975;1:480–484.

[8] 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.

[9] 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.

[10] Amorim E, Zheng WL, Ghassemi MM, Aghaeeaval M, Kandhare P, Karukonda V, et al. The International Cardiac Arrest Research (I-CARE) Consortium Electroencephalography Database. Critical Care Medicine 2023;(in press); doi:10.1097/CCM.0000000000006074.

[11] Hofmeijer J, Tjepkema-Cloostermans MC, van Putten MJ. Burst-suppression with identical bursts: A distinct EEG pattern with poor outcome in postanoxic coma. Clinical neurophysiology 2014;125(5):947–954.

[12] Zhan T. Investigating EEG Burst Suppression for Coma Outcome Prediction. Ph.D. thesis, Massachusetts Institute of Technology, 2018.

[13] Shaw JC. Correlation and coherence analysis of the EEG: A selective tutorial review. International Journal of Psychophysiology 1984;1(3):255–266.

[14] Hermans MC, et al. Quantification of EEG reactivity in comatose patients. Clinical Neurophysiology 2016; 127(1):571–580.

[15] Chen W, Liu G, Su Y, Zhang Y, Lin Y, Jiang M, et al. EEG signal varies with different outcomes in comatose patients: A quantitative method of electroencephalography reactivity. Journal of Neuroscience Methods 2020;342:108812.

[16] Shaffer F, et al. An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health 2017;258.

[17] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A System for Large-Scale Machine Learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). 2016; 265–283.

Address for correspondence:

Vasanth Kumar Babu, Department of Physics, Indian Institute of Science, Bangalore-560012, Karnataka, India. vasanthkumarb42@gmail.com, vasanthb@iisc.ac.in