Fusion of Features with Neural Networks for Prediction of Secondary Neurological Outcome After Cardiac Arrest

Philip Hempel[1], Philip Zaschke[1], Miriam Goldammer[2,*], Nicolai Spicher[1,*]

[1] Department of Medical Informatics, University Medical Center Göttingen, Georg August University of Göttingen, Germany
[2] Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany
[*] shared senior authorship

Abstract

As contribution to the 2023 George B. Moody challenge, we – team "BrAInstorm" – aimed for fusing semantic features based on medical knowledge with an end-to-end residual neural network to predict the secondary neurological outcome after successful resuscitation. More precisely, we fused numerical (e.g. age) and categorical (e.g. gender) information as well as features extracted from biosignals: We extracted absolute and relative power bands, coupling, and coherence from standard electroencephalography (EEG) frequency bands. To investigate the interplay between heart and brain, we computed deceleration capacity (DC) from electrocardiograms (ECGs). In contrast to these semantic features, we adapted a residual neural network based on agnostic features which are derived from the training data. The network architecture was originally developed for classification of ECGs and was adjusted to the challenge EEG data. The best metric scores were reached using only the neural network, demonstrating the complexity of outcome prediction and effectiveness of end-to-end methods. We received a challenge score of 0.57 ± 0.15 *during* 5*-fold cross validation on training data and* 0.448 *on the hidden validation data. On the hidden test data we received a final score of* 0.68 *(rank* 8 *of* 36*).*

1. Introduction

Cardiac arrests are defined by a sudden stop of the heart's mechanical activity. A primary good outcome is reached if the heart starts beating again whereas the secondary neurological outcome is assessed a few months later. In the period of unconsciousness during the first hours after the event, the treating clinicians estimates neurological outcome prospectively. The assessment has two aims: (1) mitigation of the financial burden on healthcare system, and (2) facilitation of informed decision making for patients and relatives.

To contribute scientific evidence for this complex problem, we used the rich dataset [1] of this year's George B Moody challenge [2, 3] for automatic prediction of binary outcome (good vs. poor). An additional task is to predict values following the Cerebral Performance Category (CPC) scale which is ordinal and ranging from 1 (good neurological function and independent for activities of daily living) to 5 (dead).

2. Methods

As shown in Tab. 1, three different types of data are processed, namely numerical, categorical, and timeseries data. Fig. 1 depicts the proposed pipeline and how the input data is combined for outcome prediction. The final output of the network is either good (CPC: 1, 2) or poor outcome (CPC: 3, 4, 5).

In our pipeline, the provided data is separated with respect to their structure to serve as input layer: Timeseries data (EEG, ECG) is processed by semantic feature extraction and an end-to-end neural network (1D-ResNet) with the extracted features both serving as input for an long short-term memory (LSTM) network. This timeseries information is subsequently combined with numerical and categorical data in a dense layer for outcome prediction. Our processing pipeline, developed in Python, is freely available.

Table 1. Data processed within the proposed pipeline.

Timeseries	Numerical	Categorical
1D-ResNet	DC, BBDC, DCsgn	Gender
EEG features	Age	Hospital
	ROSC	OHCA
		Shockable Rhythm
		TTM

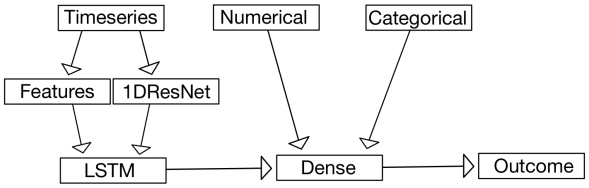

Figure 1. Proposed pipeline combining timeseries, numerical and categorical data.

The majority of the numerical and categorical data is provided within the dataset of the challenge and are therefore loaded directly from files. For the provided biosignals, we provide a preprocessing module described in section 2.1. We describe the modules for EEG feature extraction, the residual network, and DC computed from ECG in sections 2.2, 2.3, 2.4, respectively.

2.1. EEG Preprocessing

Preprocessing includes several typical steps, namely channels reorganization, bandpassfiltering, removal of non-physiological values, downsampling, bipolar re-referencing and scaling.

To maintain a consistent channel order across different recordings, the EEG channels are reordered according to a predefined list of 19 channels, excluding 'F9', 'Fpz', and 'Oz' as they were not consistently present in all recordings. Subsequently, a bandpass filter restricts the signal between 0.5 and $45\,$Hz using a Butterworth filter design. In addition, non-physiological values ± 200 mV are set to zero to avoid noisy EEG recordings disturbing the subsequent scaling. Since the provided EEGs have different sampling frequencies, they are downsampled to $100\,$Hz. For that task, *scipy*'s *resample_poly* and *resample* functions are used.

A bipolar re-referencing is applied to denoise the raw 19 EEG channels by subtracting them from nasion to inion, resulting in 18 channels. This sagittal montage aims to minimize EEG differences due to lateralization in contrast to a coronal montage [4].

For signal scaling, we aimed to extract the most recent $60\,$min EEG recording from each patient. However, due to memory constraints, only data from 300 patients could be used to apply *sklearn*'s *robust scaler* to remove the median of all EEG recordings and to scale the data according to the quantile range. Both, raw EEG and the bipolar montage are separately scaled for further processing.

2.2. Residual Neural Network

Our residual neural network is based on an open source 1D-ResNet architecture adjusted for ECG classification [5]. We adopted and hyperparameter-tuned this network using *RandomSearch* and *tensorflow* library.

We adapt the 1D-ResNet to the EEG by changing the input layer from 12 to 18 channels and using 4 x $40\,$s segments of the bipolar EEG montage as input. The segments are chosen based on two heuristically-defined criteria: (1) a variance closest to the median variance of the entire EEG data, and (2) the lowest kurtosis among segments meeting the first condition.

For standalone evaluation of 1D-ResNet only, the prediction outcome is determined primarily based on a median threshold above 0.6 for poor outcome. Additionally, if extreme values below 0.1 or above 0.95 are in the prediction array, the outcome is considered as good or poor, respectively, and the median threshold is not considered. This approach gives more weight to extreme predictions, potentially recognizing more certain outcomes over the median-based determination.

2.3. Semantic EEG Features

EEG features proposed by neurosurgeons to investigate the changes in a dying human brain [6] are implemented as semantic features for outcome prediction. Frequency bands of interest are adapted from sleep research and defined as delta ($0.5 - 4\,$Hz) for deep sleep, theta ($4 - 8\,$Hz) for drowsiness, alpha ($8 - 12\,$Hz) for relaxed states, beta ($12 - 30\,$Hz) for alertness, and gamma ($30 - 45\,$Hz) for cognitive processing. All present data is segmented in non-overlapping segments of $300\,$s duration to calculate absolute and relative power bands, phase-amplitude coupling and coherence for all segments and frequency bands. In case of missing data, the results are zeropadded to obtain arrays of the same length for each recording.

Each power band represents the signal power within a certain frequency band. Absolute power refers to the total power within that band and relative power is determined by calculating the proportion of the absolute power within that band to the total power across all frequency bands, thus providing insights into the dominant frequency bands [7]. Phase-amplitude coupling (PAC) represents the functional integration between distinct brain regions. Coherence measures the linear correlation between signals measured at two EEG electrodes in the frequency domains to assess their functional connectivity.

Power bands are calculated on the bipolar montage using the *welch* function provided by *scipy*. PAC and coherence are calculated using the raw signal analogously to the sagital electrode setting for re-referencing of the bipolar montage. The *hilbert* function provided by *scipy* is used for 5 different PAC variants including delta-theta, theta-alpha, alpha-beta, beta-gamma and alpha-gamma. The coherence is determined using the function *coherence* provided by *scipy*.

2.4. Deceleration Capacity

DC is a measure for the ability of the autonomous nervous system to control the cardiac frequency and load [8]. It is derived from the heart rate variability and was shown to correlate with mortality in different intensive care settings [8–10]. Therefore, it is used to include the neurocardiac linkage into our pipeline. We implement three different methods to calculate DC using the ECG lead with the highest number of detected inter-beat intervals (IBIs).

For calculating IBIs, the whole ECG is first filtered using the function *ecg_clean* provided by *neurokit2* [11], followed by R-peak detection using *neurokit2*'s *ecg_peaks* function. From the R-peaks, heart rate (HR) is estimated by counting the beats-per-minute (bpm). Using a lower threshold of 50 bpm, signal segments with an acceptable signal quality are detected. If HR is below 50 bpm, the ECG processing is performed again with recordings of 1 hour length instead of taking the entire length. Subsequently, IBIs are calculated using *numpy*'s *diff* function.

Next to the original DC algorithm [8], two approaches addressing noisy data [12] are added to the pipeline, namely beat to beat deceleration capacity (BBDC) [12] and deceleration capacity sign (DCsgn) [12]. They are adopted due to the rather low signal-to-noise ratio of the challenge ECGs.

The key idea of DC is to detect all IBIs in which the heart is in the process of deceleration, removing all IBIs which are not in a physiological range (more than $5\,\%$ difference w.r.t the previous IBIs). Subsequently, potentially overlapping segments are defined for all IBIs meeting this criteria. The original DC approach calculates segments of length 4 and filters only the first 2 for physiological boundaries, whereas BBDC does this for all 4 and DCsgn takes only segments of length 2. Finally, the different DC values can be calculated as the average of the slopes of the regression lines fitted to heart rate deceleration sequences, defined by the segments, representing the heart's ability to slow down over time.

3. Results

For the prediction of secondary neurological outcome after cardiac arrest, we – team BrAInstorm – aimed for a fusion of semantic and handcrafted ECG and EEG features with a 1D-ResNet neural network. Our challenge score using both training and validation data for the individual approaches can be seen in Tab. 2.

Our first approach ('1D-ResNet only') did not make use of fusion but only applied the end-to-end classifier. Our second approach ('Pipeline w/o') is based on the pipeline depicted in Fig. 1 but without the 1D-ResNet. We used 5-fold cross validation for evaluation on the public training set. Unfortunately, due to issues with memory during training and validation, we were unable to score the full combined pipeline shown in Fig. 1. Hence, the final challenge score and the rank for the test data are present for the 1D-ResNet only. However, our validation results clearly show that the 1D-ResNet outperforms the pipeline based on only semantic features and categorical and numerical values.

Table 2. Results and ranks for our two approaches: True positive rate at a false positive rate of 0.05 (the official challenge score).

Approach	Training	Val.	Test	Rank
1D-ResNet only	0.57 ± 0.15	0.45	0.68	$8/36$
Pipeline w/o NN	0.24 ± 0.12	0.22		

Additionally, we compared the performance of different DC methods, since it is a promising feature for mortality prediction in similar settings, *e.g.* after myocardial infarction [8], mitral valve repair [9], or acute ischemic stroke [10]. However, the resulting differences between good vs. poor outcome were too small and therefore of limited use in our experiments. We also compared the resulting values of the semantic EEG features such as power bands, coherence, and coupling and observed that there were visible difference but only for a rather small amount of patients.

4. Discussion and Conclusion

Our results indicate that the adoption of a 1D-ResNet for ECG classification for secondary outcome prediction based on EEG signals in an end-to-end approach worked well. We did not perform an in-depth optimization of hyperparameters, just some small adjustments and best practices.

Manually analyzing the results for the EEG features showed that the "classic" semantic EEG features show a solid basis for prediction of secondary neurological outcome after cardiac arrest. This could be similar to clinical practice where certain EEG patterns like low voltages or continuous signals, which appear in a minority of patients, can also be used for predictions with high confidence [13].

Interestingly, despite their promising results in myocardial infarction [8], mitral-valve repair [9] or acute ischemic stroke [10], the ECG-derived DCs showed no potential for statistical differentiation between patient groups. A possible explanation could be that the thorax compressions during resuscitation elicit in intense traumatic contusion of the myocardium, which could have a large impact on DC.

Moreover, we did not observe a drop in model performance when applied to the test data, which contains data from a hospital that was not present in the training data. This highlights the ability of the 1D-ResNet to learn ag-

nostic features which generalize well and minimize overfitting. However, classic approaches still have the advantage of providing transparent medical insights in contrast to the "black-box" predictions by neural networks. These can be explained [14], but only to a smaller extend, therefore limiting their translation to clinical practice.

Summing up, an end-to-end-classifier (1D-ResNet) which was adapted from an open-source neural network classifier [5] showed the best results for our team, and outperformed semantic features. This is in line with several current recent publications reporting on the excellent performance of neural networks in different medical fields. Our results demonstrate the effectiveness of the 1D-ResNet and its capability to generalize even to another biosignal.

Code Availability

The code contributed to the challenge is available at: `https://gitlab.gwdg.de/medinfpub/biosignal-processing-group/georgebmoodychallenge2023_brainstrom.git`

Acknowledgements

We would like to thank the Center for High Performance Computing at the Georg August University of Göttingen and the Technische Universität Dresden for the generous provision of computing resources.

Also, we would like to to thank the jury of the George B. Moody PhysioNet challenge 2023 for appointing us Best PhysioNet challenge Team Name: BrAInstorm.

References

[1] Amorim E, Zheng WL, Ghassemi MM, Aghaeeaval M, Kandhare P, Karukonda V, Lee JW, Herman ST, Sivaraju A, Gaspard N, Hofmeijer J, Van Putten MJAM, Sameni R, Reyna MA, Clifford GD, Westover MB. The International Cardiac Arrest Research Consortium Electroencephalography Database. Critical Care Medicine October 2023 (in press);ISSN 0090-3493.

[2] Reyna MA, Amorim E, Sameni R, Weigle J, Elola A, Bahrami Rad A, Seyedi S, Kwon H, Zheng WL, Ghassemi M, van Putten MJ, Hofmeijer J, Gaspard N, Sivaraju A, Herman S, Lee JW, Westover MB, Clifford GD. Predicting neurological recovery from coma after cardiac arrest: The George B. Moody PhysioNet Challenge 2023. Computing in Cardiology 2023;50:1–4.

[3] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[4] Beniczky S, Schomer DL. Electroencephalography: basic biophysical and technological aspects important for clinical applications. Epileptic Disorders December 2020; 22(6):697–715. ISSN 1294-9361, 1950-6945.

[5] Ribeiro AH, Ribeiro MH, Paixão GM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MP, Andersson CR, Macfarlane PW, Meira Jr W, et al. Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature communications 2020;11(1):1760.

[6] Vicente R, Rizzuto M, Sarica C, Yamamoto K, Sadr M, Khajuria T, Fatehi M, Moien-Afshari F, Haw CS, Llinas RR, et al. Enhanced interplay of neuronal coherence and coupling in the dying human brain. Frontiers in aging neuroscience 2022;14:80.

[7] Duan W, Chen X, Wang YJ, Zhao W, Yuan H, Lei X. Reproducibility of power spectrum, functional connectivity and network construction in resting-state EEG. Journal of Neuroscience Methods January 2021;348:108985. ISSN 01650270.

[8] Bauer A, Kantelhardt JW, Barthel P, Schneider R, Mäkikallio T, Ulm K, Hnatkova K, Schömig A, Huikuri H, Bunde A, et al. Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study. The lancet 2006;367(9523):1674–1681.

[9] Mizera L, Rath D, Schreieck J, Seizer P, Gawaz MP, Duckheim M, Eick C, Müller KA. Deceleration capacity of heart rate predicts 1-year mortality in patients undergoing transcatheter edge-to-edge mitral valve repair. Clinical Cardiology 2023;46(5):529–534.

[10] Duckheim M, Gaebler M, Mizera L, Schreieck J, Poli S, Ziemann U, Gawaz M, Meyer-Zuern CS, Eick C. Deceleration capacity for rapid risk stratification in patients suffering from acute ischemic stroke: a prospective exploratory pilot study. Medicine 2021;100(13).

[11] Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SA. Neurokit2: A python toolbox for neurophysiological signal processing. Behavior research methods 2021;1–8.

[12] Arsenos P, Manis G. Deceleration capacity of heart rate: Two new methods of computation. Biomedical Signal Processing and Control 2014;14:158–163.

[13] 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.

[14] Bender T, Beinecke JM, Krefting D, Müller C, Dathe H, Seidler T, Spicher N, Hauschild AC. Analysis of a deep learning model for 12-lead ecg classification reveals learned features similar to diagnostic criteria. IEEE Journal of Biomedical and Health Informatics 2023;1–12.

Address for correspondence:

Philip Hempel
Robert-Koch-Straße 40, 37075 Göttingen, Germany
philip.hempel@med.uni-goettingen.de