# Predicting Coma Recovery After Cardiac Arrest With Residual Neural Networks

Kuba Weimann[1], Tim O. F. Conrad[1]

[1] Zuse Institute Berlin, Berlin, Germany

## Abstract

***Aims***: *Interpretation of continuous EEG is a demanding task that requires the expertise of trained neurologists. However, these experts are not always available in many medical centers. As part of the 2023 George B. Moody PhysioNet Challenge, we developed a deep learning based method for analyzing EEG data of comatose patients and predicting prognosis following cardiac arrest.*

***Methods***: *Our approach is a two-step pipeline that consists of a prediction model and a decision-making strategy. The prediction model is a residual neural network (ResNet-18) that extracts features and makes a prediction based on a short 5-minute EEG recording. In the second step, a majority vote over multiple predictions made for several EEG recordings of a patient determines the final prognosis.*

***Results***: *Based on 10-fold cross-validation on the training set, we achieved a true positive rate (TPR) of 0.41 for predicting poor outcome while keeping the false positive rate below 0.05 at 72 hours after recovery of spontaneous circulation. On the official challenge leaderboard, our team ZIB_Visual scored 0.426 TPR.*

***Conclusion***: *Our approach, while simple to implement and execute, faced overfitting challenges during the official competition phase. In this paper, we discuss our implementation and potential improvements to address these issues.*

## 1. Introduction

Analyzing continuous electroencephalograms (EEG) is important for predicting the outcome of cardiac arrest, where severe brain injury is a leading cause of death. However, manual EEG interpretation is resource-intensive and often relies on specialized neurologists, limiting accessibility. These limitations, coupled with the impact of false positives, where patients initially predicted to fare poorly ultimately recover well, emphasize the need for improved automation techniques to efficiently assess patient outcomes following cardiac arrest.

Automated EEG analysis has the potential to improve accuracy and expand access where experts are scarce. The International Cardiac Arrest REsearch consortium (I-CARE) [1] provides a dataset from multiple hospitals for the purpose of developing new analysis approaches. The George B. Moody PhysioNet Challenge 2023 [2, 3] offers an opportunity to leverage this data, advancing coma prognostication in cardiac arrest with more than 1,000 subjects and 50,000 hours of EEG monitoring data.

Following the recent trends in automatic EEG analysis [4,5], we use deep convolutional networks to predict prognosis following cardiac arrest. Our approach is a two-step pipeline that consists of a residual network [6] that predicts outcome based on a short EEG recording followed by a majority vote over several predictions made for a patient that determines the final prognosis. Applied to the official challenge data, our approach reached a 0.426 true positive rate for predicting poor outcome, while keeping the false positive rate below 0.05 at 72 hours after recovery of spontaneous circulation. We discuss the implementation of our model and its shortcomings.

## 2. Training data

The challenge data was collected from seven academic hospitals and contains records of 607 comatose patients who suffered cardiac arrest. The records span from hours to several days and include continuous EEG, ECG and other signals.

Each record file contains up to an hour of signal data. Different channels are available for different patients, and channels are sometimes noisy or disconnected. The channels are organized in four groups: EEG (up to 21 channels), ECG (up to 5 channels), reference (up to 6 channels) and up to 10 other channels.

The available clinical meta-data was anonymized and contains information about the patients, including age, sex, and information about the cardiac arrest. Clinical outcome was determined using the Cerebral Performance Category (CPC) scale. The CPC scores were grouped into two categories: *good outcome* and *poor outcome*. The objective of the challenge is to predict the outcome category.

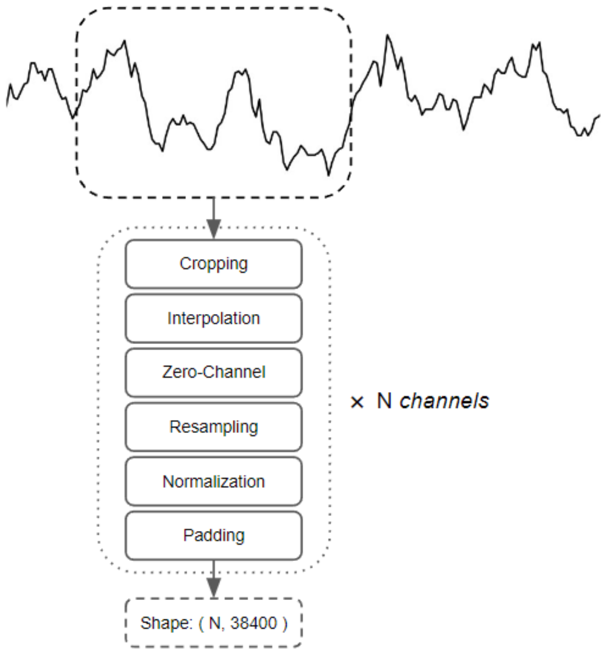


Figure 1. Preprocessing steps taken to obtain input data for the prediction model. See text for more details.

## 2.1. Preprocessing

We preprocess every record file to level the differences in signal data and improve the convergence of our models. We performed the following preprocessing steps (see Figure 1):

1. Cropping—we randomly crop a 5-minute window from all available channels.
2. Interpolation—we linearly interpolate NaN values in noisy channels.
3. Zero-channel detection—we detect empty channels and zero them out.
4. Resampling—we resample every channel at 128Hz to keep equal sampling rate across records.
5. Normalization—we normalize every channel using the mean and standard deviation computed over the window.
6. Padding—we pad every channel with zeros to keep equal shape among input data.

After preprocessing, every crop is exactly 5 minutes long and always contains the same number of channels, even if the channels are not available for that hospital (or patient). After preprocessing, the crops are evaluated and invalid crops (e.g., due to all channels being empty) are removed.

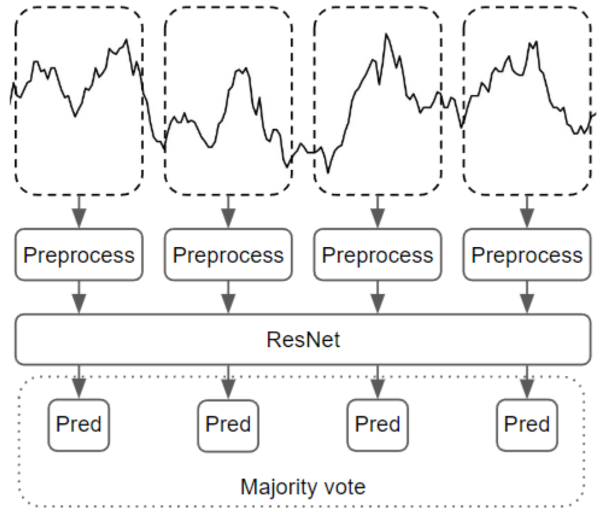


Figure 2. Our two-step prediction pipeline: First, we crop the signal data of a patient and preprocess the crops. Next, we feed the crops to the ResNet-18, which predicts outcomes. Finally, we predict a single outcome for the patient by taking the majority vote from all crop predictions.

# 3. Method

Our approach is based on a deep residual network to predict prognosis following a cardiac arrest. We developed a two-step pipeline that consists of a prediction model and a decision-making strategy. The prediction model encodes a short crop of the signal and makes predictions. The decision-making strategy is a simple majority vote over all predictions made for the many crops of the signal (see Figure 2).

## 3.1. Prediction model

We use a ResNet-18 architecture [6] with the full preactivation design [7] and group normalization [8] (see Figure 3 for an overview of our residual block). We use 1-dimensional convolutional layers with larger kernel sizes (7, 5, 5, 3 in each stage), which we found to perform better on longer sequences. Figure 4 illustrates our model.

The prediction model has two classification heads (linear output layers), to predict (1) the outcome and (2) the CPC score.

## 3.2. Training

We train our prediction model for 5000 steps using the ADAM optimizer [9] with default parameters and a batch size of 64. We use a cosine learning rate schedule [10] without restarts over 5000 steps with a linear warm-up period of 100 steps. We clip gradients whose norm exceeds

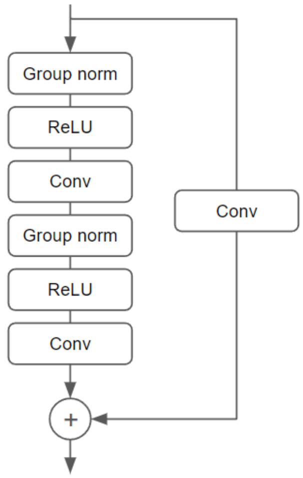


Figure 3. Full pre-activation residual block with group normalization. The shortcut path contains an additional convolutional layer in blocks that downscale the signal or increase the number of filters.

1.

We train our model using a cross entropy loss computed only on the outcome variable. The CPC output layer is trained separately so that the weights of the ResNet are not updated with the gradients from the CPC loss.

During training, we collect batches by sampling crops from records of all patients so that each batch contains diverse training data, reducing overfitting. We use only the 19 EEG channels that are available for all patients.

## 3.3. Majority vote

During evaluation, we predict a single outcome for the patient by taking the majority vote from all crop predictions. The crops stretch across every record without padding or overlapping (see Figure 2).

We implement the majority vote as a mean over logits followed by the sigmoid function (or softmax in case of CPC score). We use logits from the last $K$ model checkpoints, which are saved every 500 training steps. We empirically found that $K = 2$ gives best results. We use the same approach to predict the CPC score.

## 3.4. Predictions without data

In some cases, there may be no patient data available to make a prediction, in particular, during the first twelve hours after the cardiac arrest. For these cases, we trained an XGBoost [11] model with default parameters on the clinical information about the patient and the arrest.

We fill missing values using the most frequent value in the case of categorical variables, and with the mean value in the case of continuous variables.

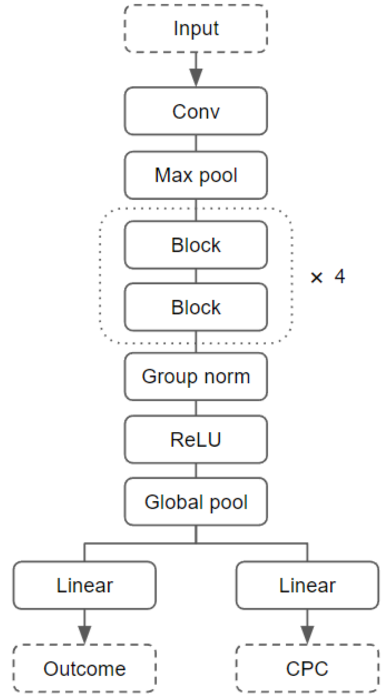


Figure 4. Our prediction model that follows the architecture of ResNet-18.

## 4. Results

We ran our approach on an internal train-validation split and later submitted it for official validation. In both cases, we found a significant drop in performance compared to training on data from the first (unofficial) phase of the challenge, which contained up to one crop per hour from the first 72 hours after the arrest.

## 4.1. Challenge metric

The challenge metric is the true positive rate (TPR) for predicting poor outcomes while keeping the false positive rate (FPR) below 0.05 at 72 hours after recovery of spontaneous circulation. This metric was chosen because in clinical practice, avoiding false positives that could lead to premature withdrawal of life support is critical, and 72 hours was selected to prevent premature predictions.

## 4.2. Internal validation split

We evaluated the predictive capabilities of our approach using 10-fold cross-validation on the training set. We focused on the TPR for predicting poor outcomes at 72 hours, as that is the metric used to rank submissions to the challenge. Our approach achieved an average TPR of 0.41. We found that using all (42) channels as input results in worse performance.

## 4.3. Official scores

In the official scores for the challenge, our approach achieved a TPR of 0.426 at 72 hours on the test set, 0.522 on the validation set, and 1.00 on the training set.

# 5. Discussion

The challenge featured very long continuous signals, spanning up to several days, which required some sort of compression or cropping to fit it into a deep learning model. We chose a simple scheme of cropping the signal into 5 minute long windows, which forced the model to focus only on the short-term patterns in the signal. In earlier experiments, we also trained a recurrent model on the embedding of crops. However, we abandoned that idea because this model significantly overfitted to the training data.

Overfitting in general was a major challenge during the development of the presented approach. We believe that training the network to predict final prognosis given a very short signal forced it to learn unique features that identify a patient, which could be alleviated with more diverse data, i.e., more patients, or better regularization. In hindsight, our approach may have needed regularization techniques such as dropout [12] or DropBlock [13] to generalize better.

# 6. Conclusion

In this paper, we presented our two-step approach to predicting coma recovery after cardiac arrest, which was the objective of the 2023 George B. Moody PhysioNet Challenge. The approach consists of a ResNet that predicts clinical outcome based on a short EEG recording, and a majority vote over several such predictions to determine the final prognosis. On the official test set, our team ZIB_Visual achieved a challenge score of 0.426. Our approach is simple to implement and execute. However, we suspect that it might be prone to overfitting.

# Acknowledgments

This work was supported by the German Ministry for Education and Research (BMBF) within the Berlin Institute for the Foundations of Learning and Data - BIFOLD (project grants 01IS18025A and 01IS18037I) and the Forschungscampus MODAL (project grant 3FO18501).

# References

[1] Amorim E, Zheng WL, Ghassemi MM, Aghaeeaval M, Kandhare P, Karukonda V, Lee JW, Herman ST, Adithya S, N. G, Hofmeijer J, van Putten MJAM, Sameni R, Reyna MA, Clifford GD, Westover MB. The International Cardiac Arrest Research (I-CARE) Consortium Electroencephalography Database. Critical Care Medicine 2023; doi:10.1097/CCM.0000000000006074.

[2] Reyna MA, Amorim E, Sameni R, Weigle J, Elola A, Rad AB, Seyedi S, Kwon H, Zheng WL, Ghassemi MM, van Putten MJAM, 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 LAN, 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 2023;101(23):e215–e220.

[4] Schirrmeister RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W, Ball T. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping 2017;38(11):5391–5420.

[5] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of Neural Engineering 2018;15(5):056013.

[6] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016; 770–778.

[7] He K, Zhang X, Ren S, Sun J. Identity Mappings in Deep Residual Networks. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer, 2016; 630–645.

[8] Wu Y, He K. Group Normalization. In Proceedings of the European Conference on Computer Vision (ECCV). 2018; 3–19.

[9] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv14126980 2014;.

[10] Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts. In International Conference on Learning Representations. 2016; .

[11] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; 785–794.

[12] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 2014;15(1):1929–1958.

[13] Ghiasi G, Lin TY, Le QV. DropBlock: A regularization method for convolutional networks. Advances in Neural Information Processing Systems 2018;31.

Address for correspondence:

Kuba Weimann
Zuse Institute Berlin, Takustraße 7, 14195 Berlin, Germany
kuba.weimann@zib.de