MemoryInception: Predicting Neurological Recovery from EEG Using Recurrent Inceptions

Bjørn-Jostein Singstad[1,2,3], Jesper Ravn[1] and Arian Ranjbar[1]

[1] Akershus University Hospital, Medical Technology and E-health, Lørenskog, Norway
[2] Vestfold Hospital Trust, Department of research and innovation, Tønsberg, Norway
[3] University of Oslo, Institute of Clinical Medicine, Oslo, Norway

Abstract

Cardiac arrest (CA) may cause severe brain damage, cognitive impairments and death. Monitoring neurological recovery after hospitalization is critical to provide suitable treatment. In this study, we aim to develop an algorithm to aid in neurological recovery classification. The dataset used for this purpose includes 1020 patients and contains both continuous sensor measurements, taken from 0 to 72 hours after the CA, and structured data. More specifically the dataset is split into a training set (60%), validation set (10%) and undisclosed test set (30%). The developed model uses a one-dimensional convolutional neural network to extract features from 5-minute time series segments, fed into a recurrent neural network, to capture temporal information and provide adaptability in recording length. The output features are then merged with embedded patient metadata in a fully connected layer, for the final classification of neurological outcome. The project is part of the George B. Moody PhysioNet Challenge 2023, where our team (EEG-Attackers) achieved a challenge score and rank on the undisclosed test dataset of 0.16 (30), 0.13 (33), 0.11 (34) and 0.12 (35) using recordings from the first 12, 24, 48 and 72 hours after CA.

1. Introduction

Cardiovascular disease accounts for approximately 17 million deaths annually, and about 40-50% of these are caused by sudden CA [1]. While the survival rate to hospital admission, for patients having out-of-hospital cardiac arrest (OHCA), is approximately 22%, the survival rate 1 year after the hospital stay is around 7,7% [2]. In patients who survive the initial resuscitation, brain injury is the most common cause of death [3], and patterns in the electroencephalogram (EEG) have shown to be a good indicator to prognosticate the outcome for sudden CA patients surviving after hospital admission [4].

The interpretation of EEG currently relies on trained medical experts such as Neurologists or Neurophysiologists. This procedure demands specialized skills, can be time-intensive, and is susceptible to subjective judgments [5]. Moreover, manual reviewing of the high volume of heterogeneous EEG data poses challenges for clinicians in delivering accurate prognostic information [6]. Novel prognostic methods are needed in order to advance this field and Deep learning has shown success in various medical fields as well as computer-based EEG analysis. Tjepkema-Cloostermans et al. 2019 developed a CNN to classify patient outcome from EEGs with an area under the receiver operating characteristic (AUROC) between 0.85-0.90 [7]. Zheng et al. 2021 showed that a multi-scale CNN-RNN achieved an area under the receiver operating characteristic (AUROC) curve in the range of 0.79-0.93 in classifying patient outcome, depending on the recording length used [8]. In this paper, we describe our approach in George B. Moody PhysioNet Challenge 2023 [9], where we use a single-scale CNN-RNN model to prognosticate good versus poor outcomes for patients admitted to the hospital after having a cardiac arrest. In contrast to previous work, we here use electrocardiograms (ECG), oxygen saturation (SPO_2) and electromyography (EMG) data in addition to EEG.

2. Method

2.1. Data

In this study we used a dataset provided by the International Cardiac Arrest REsearch consortium (I-CARE) [10, 11], containing EEG, ECG, SPO2, EMG measurements, as well as age, gender, return of spontaneous circulation (ROSC) in minutes, OHCA (yes/no), Shockable Rhythm (yes/no), targeted temperature management (TTM), patient outcome (good/poor), Cerebral Performance Category (CPC) (1-5) of 1020 patients admitted to an intensive care unit. The recordings were sampled at 500Hz and ranged from 0 to 72 hours after sudden CA, depending on

 ISSN: 2325-887X DOI: 10.22489/CinC.2023.133

when the patient arrived at the hospital after the event of sudden CA. The dataset was split into 60% training, 10% validation, and 30% test, where only the training set was shared publicly, while the validation and test set were withheld by the organizers.

2.2. Preprocessing

To handle the large data size we downsampled the sample frequency of the recordings from 500Hz to 100Hz and selected only 5 minutes of recording from each hour the patient was monitored. In case the patient was disconnected from the monitoring system, a vector of 5 minutes $\times$ 60 seconds $\times$ 100Hz of zeros was added.

2.3. Model

The model used in this study consists of three parts. The first part was a CNN, used to extract spatial features from the signals. The second part was a recurrent neural network (RNN) used to extract temporal information from the CNN features. Thirdly, a dense neural network (DNN), taking tabular data such as age, gender, etc as input, was merged with the final layer of the RNN. The CNN was trained separately from the RNN and the DNN, as shown in Figure 1, but after training the CNN was merged with the RNN, forming a single model which we have named MemoryInception.

Feature extractor

The feature extractor was built using a one-dimentional Convolutional Neural Network, based on the InceptionTime architecture [12]. Specifically, the network consists of two Inception blocks, where a block is built from three modules with a residual connection, each with four convolutional filters in turn. Features were extracted from 5-minute recordings from EEG, ECG, EMG and SPO_2 (52 signal channels).

Recurrent neural network

The recurrent neural network (RNN) was built using two long short-term memory layers (128 and 72 input neurons), taking h feature vectors of length 128, where h =number of hours after the CA episode that was used in the training scheme (either 12, 24, 48, 72).

Dense neural network

The dense neural network consists of a input layer of 6 neurons, equal to the number of variables; age, gender, OHCA, ROSC, shockable rhythm, and TTM, as inputs.

2.4. Training

Feature extractor

During training of the feature extractor, the overall patient outcome was used as the target for the supervised learning. Sigmoid activation was used to classify poor outcome $= 1$ and good outcome $= 0$. Furthermore, binary cross-entropy (BCE) was used as the loss function and ADAM as the optimizer. The model was trained for 7 epochs using single 5-minute sequences of EEG, ECG, EMG and SPO_2, from each hour the patient was monitored, as input. The start of the 5-minute recording where picked randomly if the recording where >5 minutes. In cases where the recording was <5 minutes, a tail of zeros equal to $5 - l$, where $l =$ length of the current signal, was added to the end of the signal. After training the CNN, the last layer was removed, leaving a layer of 128 neurons as the last layer. The CNN was then used to extract features from all available patient recordings in the training data and stored them in patient-wise feature stacks.

Recurrent and dense neural network

The RNN was trained based on the feature stacks from the CNN, where each stack was $h\times$ 128. The final layer of the RNN was merged with the final layer of the DNN and the two models were trained simultaneously. The model was trained for 50 epochs using BCE loss and ADAM optimization with patient outcome as the target.

Merged model

After training the CNN and the combined RNN and DNN network, they were merged by wrapping the CNN in a time-distributed layer and stacking it on top of the first LSTM layer in the RNN model. The resulting model, MemoryInception, was then used in the inference phase[1].

2.5. Inference

The MemoryInception model was queried to provide a patient-wise neurological outcome prediction. During inference, one 5-minute sequence per hour of available recordings h was selected from each patient. h was truncated to 12, 24 and 48 hours in addition to the full duration of 72 hours, when running the algorithm on the organizer side. In addition to signals, patient's age and gender, as well as information about OHCA, ROSC, shockable rhythm, and TTM were used as input per patient. In case of too short or missing recordings, these were either zero-padded or replaced by zero-value arrays.

2.6. Evaluation

The performance metric, used to evaluate the model in the challenge, was the true positive rate at a fixed false positive rate of 0.05; referred to as the challenge score. The high specificity is chosen to mitigate risk of stopping treatment for patients who have a chance to recover from CA.

[1] All code discussed in this paper are available here: https://github.com/CardiOUS/PhysioNetChallenge2021-CNN

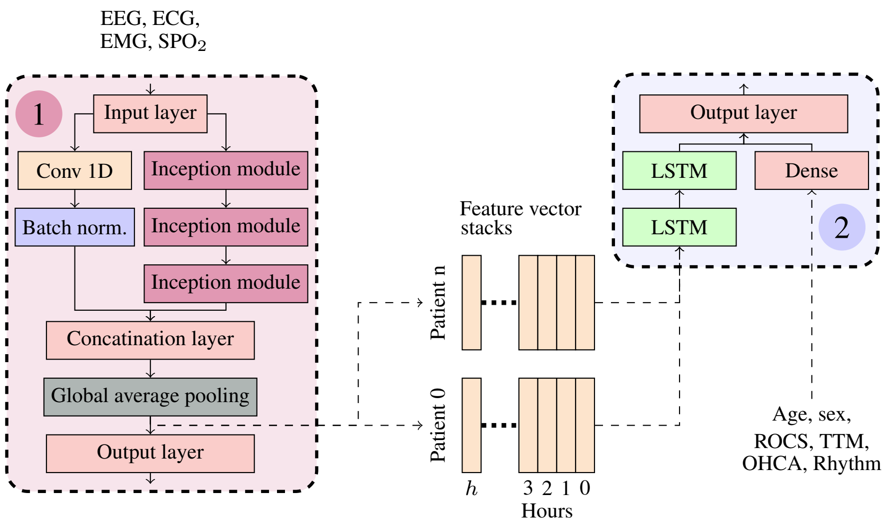

Figure 1. An overview of how the models were trained. The left side shows the InceptionTime model (1), which was first trained to predict patient outcomes from single 5-minute recordings. The trained InceptionTime model was then used to extract features from all available 5-minutes sequences. Furthermore, the feature vectors were stacked together based on patients and sorted from 0 to h after cardiac arrest. The feature vectors were then used to train the second model, which combines a recurrent and a dense neural network. The recurrent part of the model takes h feature vectors per patient and the dense neural network takes age, sex, time to return of spontaneous circulation (ROSC), targeted temperature management (TTM), Out of hospital cardiac arrest (OHCA), Shockable Rhythm. The features of the recurrent part of the network and the dense neural network were merged in the last layer and finally used to classify patient outcomes.

3. Results

The results of employing the proposed model to the undisclosed validation and test set during the official phase are shown in Table 3, as well as the results on the validation set during the unofficial phase. The model was also validated on a hold-out subset (20%) of the training data in the official phase and in the unofficial phase 5-fold cross-validation. The partitioning in both the official and unofficial phases was done at patient level, ensuring that no patient data was present in both the training and validation data.

The training curves in figure 2 show the model performance for each epoch of training in terms of BCE loss and AUROC score on both the training and validation data during training of the CNN and the merged DNN and RNN model.

4. Discussion and Conclusion

In contrast to previous studies [8], and the results obtained in the unofficial part of the challenge, our CNN-RNN model, MemoryInception, did not accurately predict neurological outcomes in patients with CA. A potential explanation might be that the data in the unofficial part of the challenges had better signal quality, than in the official phase.

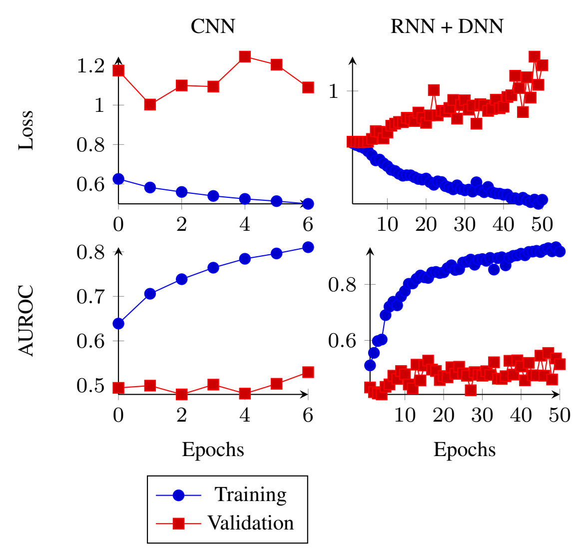

Figure 2. Training curves from validation on a subset of the training set, showing AUROC and loss during training of the convolutional neural network and the combined dense neural network and recurrent neural network.

Furthermore, the training curves during internal validation on a subset of the training set indicate that the model

Table 1. Challenge score achieved in the unofficial and official phase of the challenge. The results from each phase are divided into results from a subset of the training data (Training), the validation scores (Validation) and the test scores (Test)

Phase	Unofficial phase		Official phase		
Dataset	Training	Validation	Training	Validation	Test
0-12 hours	-	0.18	-	0.15	0.16
0-24 hours	-	0.57	-	0.22	0.11
0-48 hours	-	0.61	-	0.10	0.13
0-72 hours	0.45 ± 0.21	0.66	0.08	0.13	0.12

overfitted to the training data. In future research using the MemoryInception architecture, it would be beneficial to develop an algorithm to select the 5 minutes with the best signal quality from each hour of recording. Furthermore, training both the feature extractor and RNN jointly could potentially mitigate overfitting. However, this approach would require either a substantial amount of RAM or a customized data loader and training scheme.

Moreover, substituting absent recordings with zero-value vectors might have introduced inaccuracies in differentiating between genuine patient signals and those from individuals with significant brain damage. Future advancements should focus on identifying signal characteristics associated with patients exhibiting poor outcomes versus good outcomes to effectively replace missing values without introducing any biases.

In conclusion, our MemoryInception model showcases some potential in predicting neurological outcomes for CA patients by utilizing a combination of EEG, ECG, and patient-specific data. However, further research is required to address the problem of overfitting observed during training, and the handling of missing data; to increase performance and clinical applicability.

Funding

This work was supported by the Norwegian Health Association (Grant number #29038 - Artificial intelligence-enabled ECG)

References

[1] Mehra R. Global public health problem of sudden cardiac death. Journal of Electrocardiology November 2007;40(6, Supplement 1):S118–S122, DOI: 10.1016/j.jelectrocard.2007.06.023.

[2] Yan S, Gan Y, Jiang N, Wang R, Chen Y, Luo Z, et al. The global survival rate among adult out-of-hospital cardiac arrest patients who received cardiopulmonary resuscitation: a systematic review and meta-analysis. Critical Care February 2020;24:61, DOI: 10.1186/s13054-020-2773-2.

[3] Rundgren M, Westhall E, Cronberg T, Rosén I, Friberg H. Continuous amplitude-integrated electroencephalogram predicts outcome in hypothermia-treated cardiac arrest patients. Critical Care Medicine September 2010;38(9):1838–1844, DOI: 10.1097/CCM.0b013e3181eaa1e7.

[4] Amorim E, Rittenberger JC, Zheng JJ, Westover MB, Baldwin ME, Callaway CW, et al. Continuous EEG monitoring enhances multimodal outcome prediction in hypoxic-ischemic brain injury. Resuscitation December 2016; 109:121–126, DOI: 10.1016/j.resuscitation.2016.08.012.

[5] Zubler F, Tzovara A. Deep learning for EEG-based prognostication after cardiac arrest: from current research to future clinical applications. Frontiers in Neurology July 2023; 14:1183810, DOI: 10.3389/fneur.2023.1183810.

[6] Zheng WL, Amorim E, Jing J, Wu O, Ghassemi M, Lee JW, et al. Predicting Neurological Outcome from Electroencephalogram Dynamics in Comatose Patients after Cardiac Arrest with Deep Learning. IEEE transactions on bio medical engineering May 2022;69(5):1813–1825, DOI: 10.1109/TBME.2021.3139007.

[7] Tjepkema-Cloostermans MC, da Silva Lourenço C, Ruijter BJ, Tromp SC, Drost G, Kornips FHM, et al. Outcome Prediction in Postanoxic Coma With Deep Learning*. Critical Care Medicine October 2019;47(10):1424, DOI: 10.1097/CCM.0000000000003854.

[8] Zheng WL, Amorim E, Jing J, Ge W, Hong S, Wu O, et al. Predicting neurological outcome in comatose patients after cardiac arrest with multiscale deep neural networks. Resuscitation December 2021;169:86–94, DOI: 10.1016/j.resuscitation.2021.10.034.

[9] Reyna M, Amorim E, Sameni S, 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 W, Lee J, Herman S, Ghassemi M, Sivaraju A, et al. I-CARE: International Cardiac Arrest REsearch consortium Database. URL https://physionet.org/content/i-care/2.0/.

[11] Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation June 2000;101(23), DOI: https://doi.org/10.1161/01.CIR.101.23.e215.

[12] Ismail Fawaz et al. H. InceptionTime: Finding AlexNet for Time Series Classification. Data Mining and Knowledge Discovery November 2020;34(6):1936–1962, DOI: https://doi.org/10.1007/s10618-020-00710-y.

Address for correspondence:

Bjørn-Jostein Singstad
Sykehusveien 25, 1478 Nordbyhagen, Norway
b.j.singstad@fys.uio.no