Autoencoder Artefact Removal for Brain Signals and Impact on Classification Performance

Mengyao Li[1], Le Xing[1], Alexander J. Casson[1]

[1]The University of Manchester, Manchester, UK

Abstract

As part of the George B. Moody PhysioNet Challenge 2023, we developed a computational approach that uses 6 channels of electroencephalograms (EEGs) to predict neurological recovery outcomes of patients following cardiac arrest. Our team, UoM_EEE, developed a 2-Dimensional Convolutional Neural Network, using the Short-Time Fourier Transform to obtain an image representation of the EEG. It uses an optimised Binary Focal Cross-entropy loss function for balancing weights of two-outcome classes. As standard EEG analysis pipelines using Independent Component Analysis (ICA) to remove artefacts are not suitable due to the limited channel count, we hypothesized that an autoencoder machine learning approach may allow a channel count independent artefact removal, and potentially an improved true positive rate, while naturally complementing machine learning based classification used for the main Challenge problem. A 5-run class-stratified nested holdout was performed, with Area under the Receiver Operating Characteristic Curve, AUC, as metric for model selection. Our model received a Challenge score of 0.188 (ranked 34th out of 36 teams) on the hidden test set, 0.388 on the hidden validation set, and 0.67 averaged across 5-trial cross-validation on the public training data.

1. Introduction

Cardiac arrest (CA) is a significant global public health concern and responsible for a substantial portion of fatalities [1]. Clinicians are usually asked to predict the probabilities of patients recovering to consciousness. To this aim, the 2023 George B. Moody PhysioNet Challenge [1, 2] was organized to encourage teams to design an automated analysis and interpretation system for neurological recovery prediction of patients following CA, based on provided electroencephalography (EEG) and other clinical information [3]. To participate in it, our team submitted an entry, in Python, combining Short-Time Fourier Transform (STFT) and Convolutional Neural Network (CNN) for enhanced representation, feature extraction and classification from EEG.

To improve performance we explored the hypothesis that artefact removal would be critical. EEG signals are small and very sensitive to noise and interference. In standard analysis pipelines Independent Component Analysis (ICA) is used to remove artefacts. However the numerical condition means ICA requires a large number of channels for good performance, and so is not best suited for EEG data in the Challenge, nor for use in the Intensive Care Unit (ICU) where a reduced number of channels can help with setup time and ease of use. We hypothesized that an autoencoder machine learning approach may allow a channel count independent artefact removal approach, giving cleaned EEG, and potentially an improved true positive rate, while naturally complementing machine learning based classification used for the main Challenge problem.

2. Methods

2.1. Data preprocessing

We made use of only the EEG data. Recordings collected beyond 72 hours after Return of Spontaneous Circulation (ROSC) were discarded, followed by selecting the latest record, rather than the first one verified informative by Team EEGnition, for each patient. Similar to Team AIrhythm, six EEG channel pairs, F3-T3, T3-P3, F3-P3, F4-T4, T4-P4, F4-P4 were selected, those verified effective for patients post CA [4]. Data was prepossessed (provided by the Challenge) with a 0.1–30 Hz 4$^{\text{th}}$ order bandpass Butterworth filter and resampling to 128 and 125 Hz for even and odd original sampling rates respectively. Data was scaled to the interval of $[-1, 1]$ with the Max-Min rule.

2.2. Short-Time Fourier Transform

To give a 2-Dimensional (2D) format for the deep learning input, the STFT was applied to the EEG, according to [5]. The STFT was applied to each time-series bipolar EEG channel utilizing the `stft` function of the `signal` module in Python. The settings were kept as the default, including Hann window, 512 samples within one segment, and half-overlapping sliding, except the sampling rates which should be consistent with their respective resampling rates. Considering that the lengths of the EEG recordings for each patient are different, the size of obtained STFT images were resized to 128×128 with linear interpolation through the `cv2.resize()` function of the OpenCV library cv2.

2.3. Convolutional neural network

A 2D CNN was employed to estimate whether coma patients recovered following CA with given EEG. The EEG data, after the aforementioned procedures, were transformed into a 4-Dimensional tensor. All EEGs were treated as images with the last 3 dimensions of inputs corresponding to width, height and depth of images, which could be combined with hand-crafted features used by Team TUD_EEG.

The designed CNN comprised of 10 stages to automatically learn multilevel features. Each of the first 6 stages contained a 2D convolutional layer with identical kernel size of 3×3, rectified linear unit (ReLU) activation function, and 'lecun uniform' kernel initialization which was different from Team ComaToss by whom pre-trained weights were used, followed by a batch normalization layer and a MaxPooling layer with the pooling size of 2×2. The number of filters in these 6 layers was increased as the depth of the CNN structure increased, being 16, 32, 64, 128, 256 and 512 respectively. The 7th stage was a flatten layer which reshaped the feature maps obtained after the first 6 stages into a column-like data series. Subsequently, two dense layers were added as fully-connected layers with 1024 neurons, a linear activation function and 'lecun uniform' kernel initialization. To avoid overfitting a dropout layer was implemented at the end of the first dense layer with a dropout rate of 75%. The last stage also contained a dense layer with 2 neurons indicating the categories in recovery outcome, softMax activation function and 'lecun uniform' kernel initialization.

For the optimization algorithm, an Adaptive Moment Estimation estimator with learning rate of 0.001 was utilised to iteratively update neural weights for outcome prediction on EEG data. In the model training process, the batch size and the number of epochs were 32 and 150 respectively. To reduce overfitting an early stopping function was applied with the number of early-stopping epochs being 10 and the criterion that the validation loss starts to rise.

Binary focal cross-entropy [6] was used as loss function, wherein the alpha factor, α_t, was 0.5 for equal weight-balancing factors for good (Class 0, negative) and poor (Class 1, positive) outcomes; the gamma factor, γ, was 2 as default to focus more on hard-to-classify examples. The Area under the Receiver Operating Characteristic Curve, AUC, was utilised to evaluate the model performance for outcome prediction.

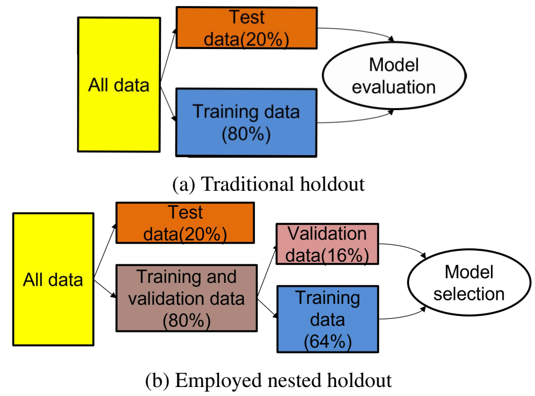


Figure 1: Schematic diagrams of cross validation methods.

2.4. Nested holdout based model validation

As demonstrated in [7], nested techniques could provide better model performance in terms of robustness and accuracy when working with small datasets. We used nested holdout using the `train_test_split` function in the `sklearn.model_selection` library with class-based stratification to the EEG as shown in Figure 1. In our case, the validation process was performed to select the optimal model from 5 runs, followed by testing with the optimal model on internally hidden data. We randomly split 10 patients into a hold-out test set, and used the rest for 5-run cross validation and final model selection. This procedure was repeated for 5 trials to obtain our averaged cross-validation score and standard deviation.

2.5. Autoencoder-based artefact removal

To investigate the impact of autoencoder-based EEG artefact removal on the classification performance, we implemented our previous 1-Dimensional deep convolutional autoencoder neural network, reported in [8], as an optional pre-processing stage. The autoencoder network is shown in Figure 2 and was developed for removing ocular, muscular, and motion artefacts from single-channel EEG signals, and reconstructing the originally clean EEG without introducing distortions, in a segment-by-segment processing manner. To accommodate to the sampling frequency and recording duration of the EEG data provided in the Challenge, the original autoencoder was retrained to process the EEG segments with a 5-second duration. The retrained model was validated to have a similar performance as reported in [8].

We thus trained and tested our CNN network twice,

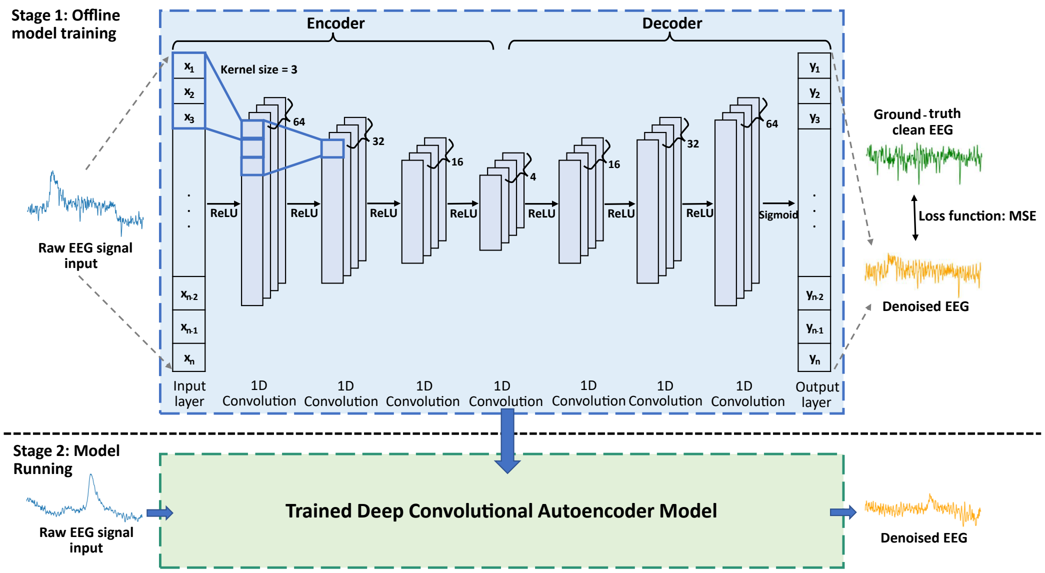


Figure 2: Deep autoencoder architecture for clean EEG reconstruction and EEG artefact removal, reported in [8]. Stage 1 is the offline model training, and Stage 2 is the model running online. Reprinted under CC BY 4.0 license.

once using the autoencoder-cleaned EEG as the input to the STFT stage, and once using the raw EEG as the input. The official Challenge score reported is for the raw EEG input network. We also compared the performance in the cases where normalisation was conducted on the whole EEG recording or on every segment separately.

3. Results

Figure 3 exemplifies comparisons between time domain and time-frequency domain diagrams for EEG channel F4-T4 from Patients 0675 and 0888, representing good and poor outcomes. The enhanced data representation of the STFT method is demonstrated with time and frequency information, which provided easier-to-classify features for outcome prediction.

Table 1 summarizes the Challenge scores on the training, validation and test sets. Table 2 illustrates the impact of the proposed autoencoder based artefact removal on Challenge performance.

4. Discussion and Conclusions

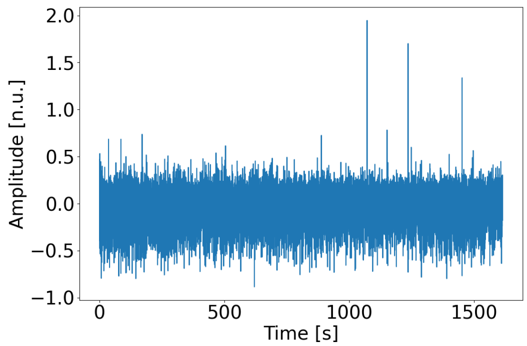
Table 1 and Figure 3 verifies the feasibility and effectiveness of the proposed CNN model combining STFT method. As mentioned before, hand-crafted features could be included and combined with a pre-trained model for future work. The decrease in Challenge scores using autoencoder for artefact removal in Table 2 implies a negative effect of the proposed autoencoder model, especially when segmenting was implemented before normalisation,

Training	Validation	Test	Ranking
0.207	0.388	0.188	34/36

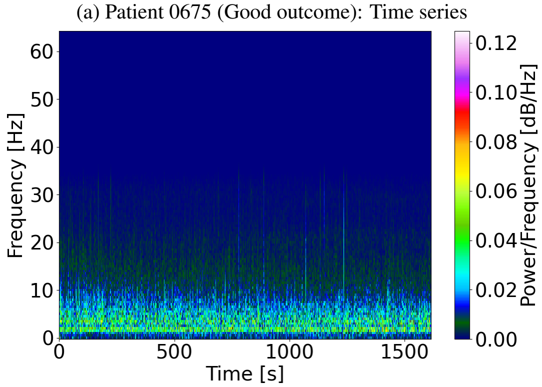
Table 1: True positive rate (TPR) at a False positive rate (FPR) of 0.05 (the official Challenge score) for our final selected entry (team UoM_EEE), including the ranking of our team on the hidden test set. We ran 5-run cross validation on the public training set with 10 patients held out for 5 trials, and calculated mean and standard deviation with score of 0.67 ± 0.14, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set.

i.e. the normalisation was conducted segment-by-segment as would be required for real-time (rather than offline) use. As stated in [8], one of the explanation would be that the normalisation would scale the EEG to [0,1], leading to a sudden increase or decrease of EEG amplitude when concatenating the segments. It could have an adverse effect on STFT-based image representation and therefore deteriorate model performance.

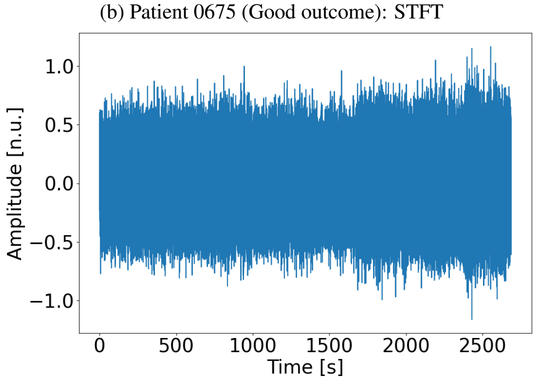
Our hypothesis was therefore not supported. It could be that the ICU EEG data is relatively artefact free, with patients being stationary, or that the autoencoder based artefact removal is too aggressive and removes useful signal information. While the performance metrics for signal reconstruction in [8] are promising, this is the first time we have applied the approach for cleaning *real* signals where no ground truth comparison trace is available. EEG signals are small and very sensitive to noise and interference, and while further work on low-channel count artefact removal



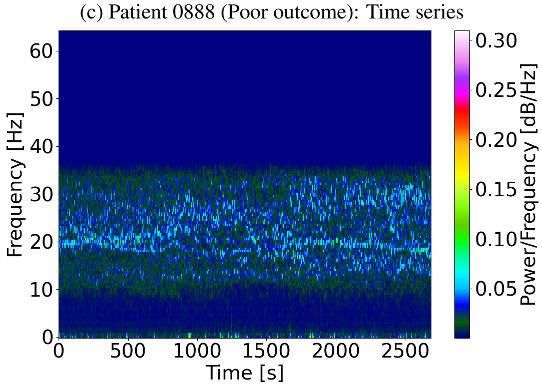
(a) Patient 0675 (Good outcome): Time series



(b) Patient 0675 (Good outcome): STFT



(c) Patient 0888 (Poor outcome): Time series



(d) Patient 0888 (Poor outcome): STFT

Figure 3: Comparisons between time series and STFT diagrams of F4-T4 EEG channel pair, wherein STFT indicates Short-Time Fourier Transform, n.u. means no units.

Experimental settings	Challenge score
No artefact removal	0.39
Artefact removal (whole signal normalised at same time)	0.21
Artefact removal (normalization applied segment by segment)	0

Table 2: Challenge scores on hidden validation set using different experimental settings. The last two rows represent results after autoencoder-based artefact removal, wherein normalisation was conducted on the whole EEG recording or every segment respectively.

algorithms is required, we would still posit that dedicated artefact removal stages will be beneficial for practical deployments into ICUs.

References

[1] 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.

[2] 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.

[3] 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);URL `https://doi.org/10.1097/CCM.0000000000006074`.

[4] Backman S, Cronberg T, Rosén I, Westhall E. Reduced EEG montage has a high accuracy in the post cardiac arrest setting. Clinical Neurophysiology 2020;131(9):2216–2223.

[5] Wang Q, Wang X. EMG-based hand gesture recognition by deep time-frequency learning for assisted living & rehabilitation. In IEEE UEMCON. 2020; 0558–0561.

[6] Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In IEEE ICCV. 2017; 2980–2988.

[7] Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PloS One 2019;14(11):e0224365.

[8] Xing L, Casson A. Deep autoencoder for real-time single-channel EEG cleaning and its smartphone implementation using TensorFlow Lite with hardware/software acceleration. TechRxiv Preprint 2023;URL `https://doi.org/10.36227/techrxiv.23933058.v1`.

Address for correspondence:

Alexander J. Casson
Royce Hub Building, Oxford Road, Manchester, M13 9PL, UK
alex.casson@manchester.ac.uk