Predicting Cardiac Arrest Recovery with Shallow and Deep Learning Models

Ekenedirichukwu Obianom[1], Marko Mäkynen[2], Noor Qaqos[2], Shamsu Idris Abdullahi[2], Fernando S Schlindwein[2,3], G André Ng[1,3], Xin Li[2,3]

[1]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK
[2]School of Engineering, University of Leicester, Leicester, UK
[3]National Institute for Health Research Leicester Cardiovascular Biomedical Research Centre, UK

Abstract

Most resuscitated cardiac arrest patients are comatose and often die due to severe brain injury. With the uncertainty of which patient will survive or not, it is important for the right prognosis to be given. This would help decide which patient intensive care should be focused on. Machine learning, which is a revolutionary computer program that can work without explicit instructions, could be used to study neurological patterns, and make better prognosis. As part of the Predicting Neurological Recovery from Coma After Cardiac Arrest: The George B. Moody Physionet Challenge 2023, our team, Leicester Fox, focused on the comparison of the effectiveness of shallow and deep learning machine learning models in giving the right prognosis on chances of survival of cardiac arrest comatose patients. Features extracted from the electroencephalography (EEG) of 607 patients are used for this analysis. Three groups of features (18 features) were extracted and use for training. The official result was a challenge score of 51% for the shallow model. Locally, we had an accuracy of 76% and 65% for shallow and deep learning models respectively. In conclusion, when dealing with smaller number of patients and using features for analysis, shallow classifiers would usually give a better result.

1. Introduction

Most cardiac arrest patients who survive resuscitation are comatose and often die due to severe brain injury [1]. Physicians are often asked to give a prognosis on the recovery of these patients. The prognosis could be good or poor resulting in continued care of the patient, as poor prognosis usually leads to removal of life support of the patient, respectively.

False positives (where poor prognosis is given but the patient still recovers) are not rare and poses an issue to the medical sector. It is paramount that false positives are reduced to the barest minimum so that patients who would truly survive would not be removed from life support. To eliminate the human subjectivity of prognosis, a method comprised of an automated system needs to be built.

Years of research has presented patterns in brain signals which have proven useful in prognosis [5]. These patterns, coupled with clinical data and patient outcomes can be used to design a machine learning (ML) model to give little to no false positives with a high degree of accuracy. ML comes in handy in systems where humans have been unable to recognize substantial patterns that are enough to make a practical system [6]. In the design of a system capable of proffering a prognosis, multiple models need to be tested to be able to conclude on a reliable model.

This work focuses on the design, modelling, and analysis of two ML algorithms to proffer reliable prognosis of the eventual recovery of comatose patients resulting from cardiac arrest. It was proposed by The George B. Moody Physionet Challenge 2023 [3] with data from the International Cardiac Arrest Research consortium (I-CARE) [2].

2. Methodology

2.1. Dataset

I-CARE is a database of comatose patients with, at most, 72 hours of 18-channel electroencephalogram (EEG) recordings, ECG recordings, clinical data, and recovery status of each patient [2, 7]. The database includes seven hospitals from the United States and Europe. The database consists of 1020 adult patients, but this study was done on only approximately 60% (which was what was made available at the time of this study).

 ISSN: 2325-887X DOI: 10.22489/CinC.2023.394

2.2. Shallow Classifier

Three groups of features (patient information, complexity, and category features) were extracted from the dataset.

1. A set of eight patient-related features were gathered for each individual, with these features encompassing data such as age and gender at admission, a hospital identification code, the context of the cardiac event (whether it occurred outside or within the hospital), the specific cardiac rhythm observed during resuscitation (including shockable rhythms like ventricular fibrillation or ventricular tachycardia, as well as non-shockable rhythms like asystole and pulseless electrical activity), and the duration between the cardiac arrest and the restoration of spontaneous circulation (ROSC).
2. The category features and complexity features are described in table 1. Category features quantify the degree of the brain states, while the complexity features quantify the degree of irregularity, randomness and the chaotic in the EEG signals. These two feature classes likely to carry some prognostic significant to classify the EEG signals for predicting the cardiac arrest recovery after the coma.

Table 1. Details and description of category and complexity features.

Groups	Features (parameters)	Descriptions
category	Delta PSD (δ band Power) [8]	0.5-4Hz power spectral range
	Theta PSD (ϴ band power) [8]	4-7Hz power spectral range
	Alpha PSD (α band power) [8]	8-12Hz power spectral range
	Beta PSD (β band power) [8]	13-30Hz power spectral range
	Median Frequency [9]	The median spectral frequency of a signal
complexity	Hjorth parameters (Mobility) [10]	Mean frequency
	Hjorth parameters (Complexity) [10]	Variation of frequency
	Shannon entropy [11]	Additive measure of signal stochasticity
	Higuchi fractal dimension [12]	Investigate the brain responses for the important audio information in patients' brain injury.

$$Mobility = \sqrt{\frac{var\left(x'(t)\right)}{var\left(x(t)\right)}} \quad (1)$$

$$Complexity = \frac{mobility\ (x'(t))}{mobility\ (x(t))} \quad (2)$$

$$Entropy = -\sum_i p_i \log p_i \quad (3)$$

$$PSD = |X(f)|^2 \quad (4)$$

where: $var(x(t))$ and $x'(t)$ are the variance and the first derivative for the input signal $x(t)$, respectively. p_i is the probability that the system is in the $i\ th$ state, $|X(f)|$ is the magnitude for the frequency (f).

Patient information features were extracted from the information file for each patient. Category and complexity features were extracted from 72hrs EEG recording files for a signal duration for the first 5 minutes from each channel. The mean process was done for all channels in the 72 hrs recording file. TreeBagger classifier was used to train and test the model. Five-fold cross validation technique was applied to measure the performance of classifier. 80% of the dataset was used as train set and 20% as a test set.

2.3. Deep Learning Model

The deep neural network (DNN) model was trained using the same features as the shallow model. Typically, the DNN is trained using raw data with model containing multiple layers to optimize the performance [13]. However, training the model usually requires a large amount of data. When a large dataset is used, a deeper model is mainly needed for optimal results. Thus, the data acquisition can be demanding, and computational requirements increase. Furthermore, the data might require pre-processing concerning the perturbation (denoising), which can increase even more the computational requirements. Features can potentially add robustness and provide the use of a compressed deep learning model. We used convolutional neural network (CNN) combined with attention mechanism.

The CNN forms an automated feature in the training process [13]. Each layer uses filters to construct feature maps. And the attention mechanism makes the model focus more on a specific part of the data [14]. The model is presented in Figure 1. Hyperparameters were chosen for the DNN by trying various setups and choosing the best configuration. However, the data amount was considered, and therefore, the compressed model was used. The same preprocessing and features (input) was used as with the shallow model.

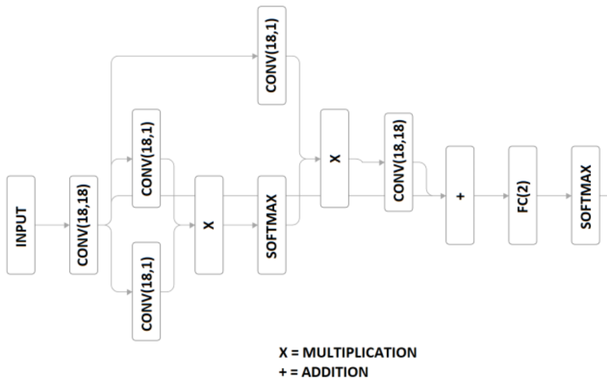

Figure 1. The deep learning model used for classification of categories good and poor. The model for CPC was formed by changing FC(2) to FC(1) and removing SoftMax layer.
CONV(f,m) = Convolutional layer (filter size, feature maps)
FC = Fully connected layer(neurons).

3. Results

Two proposed methods were used for predicting neurological recovery after cardiac arrest from coma. The first approach was based on using shallow classifier and in particular the treebagger algorithm, and another was dependent on deep learning model. Table 2 shows the scores and the accuracy for the proposed approaches. A higher performance can be obtained from the shallow classifier compared to the deep-learning model.

Table 2. Result using publicly available database.

MODEL	CHALLENGE SCORE	5-FOLD CROSS ACCURACY	Negative Class Accuracy
CNN-attention	0.33 (Local Score)	0.65	0.70
Shallow model	**0.51 (Official Score)**	0.75	0.74

Table 2 shows the results achieved by testing the shallow model and the CNN model on the publicly available dataset.

The challenge score, as described by Physionet [7], is the specificity (True Positive Rate) at a False Positive Rate (FPR) of less than 0.05. Mathematically written as,

$$TPR = \frac{TP}{FP + FN}$$

$$where\ FPR \leq 0.05$$

The accuracy is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\begin{gathered} where: \\ TPR = True\ Positive\ Rate \\ TP = True\ Positive \\ TN = True\ Negative \\ FP = False\ Positive \\ FN = False\ Negative \end{gathered}$$

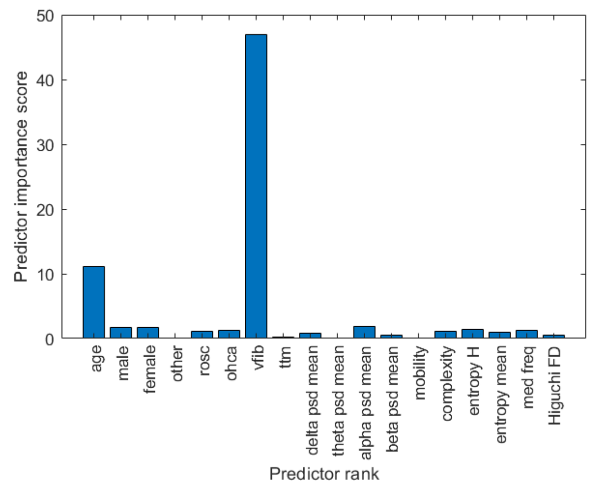

Figure 2: Feature Importance

4. Discussions and Limitations

The CNN attention model is less accurate than the shallow model, which may be due to the small amount used for training. Both models have a low challenge score and FPR (as defined by the conditions challenge score. The accuracy is good, but a low challenge score means there would still be a noticeable chance of poor prognosis for the wrong patient. This would counter the main aim of this work.

These results, nevertheless, are promising as they show that shallow models can be very useful in giving a prognosis for the recovery of a comatose patients. One of the drawbacks to these models could be that the features were extracted from only the first 5 minutes of the last recording of the patients. There could have been better results if a small section of each hour was used. This would require more computing power.

The deep learning model result may improve if more data were used for training. Therefore, when using a small amount of data, additional data augmentation is preferred. For example, a synthetic data generator would be a suitable solution. Furthermore, the dataset was not balanced, which affected the training. But the feature-based inputs for the deep learning model is an interesting choice to consider.

It is important to note the information obtained from the feature importance algorithm. Chi squared algorithm was used to plot the feature importance in predicting a recovery prognosis. It was found that ventricular fibrillation and age played the most important role in determining a good and poor prognosis. This finding rendered the other features used in this study close to inconsequential for a feature-based prediction.

The dataset size (number of samples) was a challenge in our training setup. Therefore, augmented data would be desired. Some suitable methods exist, for example, semi-supervised learning using labelled and unlabelled data and synthetic data using a data generator. However, the semi-supervised method would require acquiring new live data. Therefore, synthetic data could be a feasible solution. The requirement is to generate samples to improve the model training as efficiently as possible. Such a method could address the lack of labelled samples problem and replacement of low-quality samples.

5. Conclusion

This work proposed two approaches to predict neurological recovery after coma due to cardiac arrest. Shallow model and deep learning models were trained using the same features. The higher performance was obtained from shallow model officially as compared to deep learning model locally with challenge scores 0.51 and 0.33 respectively; 5-fold cross validation accuracy of 0.75 and 0.65; and negative class accuracy of 0.74 and 0.70, respectively. The two most important features for the training, as seen from the chi square algorithm, are ventricular rate and age.

Further work into a feature-based approach, for coma recovery prognosis, should recognise that ventricular rate and age play a vital role in the prediction while EEG categorical and complexity features are inconsequential.

References

[1] M. Rundgren, E. Westhall, T. Cronberg, I. Rosén, and H. Friberg, “Continuous amplitude-integrated electroencephalogram predicts outcome in hypothermia-treated cardiac arrest patients,” *Critical care medicine,* vol. 38, no. 9, pp. 1838-1844, 2010.

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

[3] Reyna MA*, Amorim E*, Sameni S, Weigle J, Elola A, Bahrami Rad A, Seyedi S, Kwon H, Zheng, WL and Ghassemi M, 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.

[4] J. L. Greenfield, P. R. Carney, and J. D. Geyer, *Reading EEGs*, Philadelphia: Philadelphia: Wolters Kluwer Health, 2020.

[5] L. J. Hirsch, S. M. LaRoche, N. Gaspard, E. Gerard, A. Svoronos, S. T. Herman, R. Mani, N. Jetté, Y. Minazad, J. F. Kerrigan, P. Vespa, S. Hantus, J. Claassen, G. B. Young, E. So, P. W. Kaplan, M. R. Nuwer, N. B. Fountain, and F. W. Drislane, “American Clinical Neurophysiology Society's standardized critical care EEG terminology: Interrater reliability and 2012 version,” *Journal of the neurological sciences,* vol. 333, pp. e15-e16, 2013.

[6] Y. Zhang, *Machine Learning*, Rijeka, Croatia: IntechOpenInTech, 2010.

[7] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation,* vol. 101 (23), pp. e215–e220, 2000.

[8] C. S. Nayak, and A. C. Anilkumar, "EEG Normal Waveforms. StatPearls," Treasure Island, FL: StatPearls Publishing. https://europepmc.org/article/nbk/nbk537023, 2019.

[9] M. Kafashan, S. Ryu, M. J. Hargis, O. Laurido-Soto, D. E. Roberts, A. Thontakudi, L. Eisenman, T. T. Kummer, and S. Ching, “EEG dynamical correlates of focal and diffuse causes of coma,” *BMC Neurol,* vol. 17, no. 1, pp. 197, Nov 15, 2017.

[10] S.-H. Oh, Y.-R. Lee, and H.-N. Kim, “A Novel EEG Feature Extraction Method Using Hjorth Parameter,” *International Journal of Electronics and Electrical Engineering*, pp. 106-110, 2014.

[11] C. E. Shannon, and W. Weaver, "Shannon. The Mathematical Theory of Communication," University of Illinois Press, pp. 1-10, 1998.

[12] A. Accardo, M. Affinito, M. Carrozzi, and F. Bouquet, “Use of the fractal dimension for the analysis of electroencephalographic time series,” *Biological cybernetics,* vol. 77, no. 5, pp. 339-350, 1997.

[13] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature (London),* vol. 521, no. 7553, pp. 436-444, 2015.

[14] Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” *Neurocomputing (Amsterdam),* vol. 452, pp. 48-62, 2021.

Address for correspondence:

Dr Xin Li.
School of Engineering
University of Leicester, UK
xin.li@leicester.ac.uk