A Machine Learning Approach for Outcome Prediction in Postanoxic Coma Patients Using Frequency Domain Features

Vijay Vignesh Venkataramani[1], Akshit Garg[1], Maitreya Maity[2], U Deva Priyakumar[1,2]

[1] Center for Computational Natural Sciences and Bioinformatics, IIIT, Hyderabad
[2]IHub Data, IIIT, Hyderabad

Abstract

In this work, we describe the creation of our machine-learning-based solution for coma prognosis after cardiac arrest using longitudinal EEG and ECG recordings for the "Predicting Neurological Recovery from Coma After Cardiac Arrest: The George B. Moody PhysioNet Challenge 2023". Our team, "ComaToast", had its best submission ranked 28 out of 36 teams selected worldwide, with a challenge score of 0.381 on the official leaderboard for the hidden test set. We use a combination of age and signal features from EEG and ECG recordings. Frequency domain features, specifically mean power spectral density from 4 different bands of frequencies (Delta, Theta, Alpha and Beta) and mean Burst Suppression Ratio, were extracted from pre-processed EEG recordings from the first and last available recording for a given patient. Features like mean and standard deviations were extracted along channels for ECG recordings. After imputing missing values, these features are fed to an XGBoost classifier for the final binary classification of the outcome prediction task. The features are fed to a random forest regressor to predict the CPC outcome for every patient. A solution like ours, which uses a simple model and training technique, may be more viable than deep-learning solutions in general use cases. In our final model, our approach achieved a 5-fold cross-validation score of 0.34 on the public train set.

1. Introduction

Cardiac Arrests are one of the leading causes of cardiac-related deaths worldwide. More than 350,000 people suffer from an out-of-hospital cardiac arrest every year in the United States alone [1,2]. Even if patients are successfully resuscitated after the arrest, other complications are very likely to arise. For patients surviving initial resuscitation, the most common condition that occurs after, often leading to death, is severe brain injury [3]. Doctors are often put in a difficult position to provide an accurate prognosis for such patients in coma, where a wrong prognosis may serve as a life sentence to the patient. With such a heavy responsibility in their hands, continuous brain monitoring can be beneficial in aiding doctors to make such decisions

In this paper, we present our machine learning based solution for Outcome Prediction in Postanoxic Coma Patients as part of the 2023 George B. Moody PhysioNet Challenge. Teams were invited to develop automated solutions that use longitudinal electroencephalogram (EEG) and other recordings to predict patient outcomes after cardiac arrest. [4,5]. An EEG is a procedure where the brain's electrical activity is recorded using electrodes placed on different locations of a patient's scalp. In addition to the EEG recordings, teams were also given access to Electrocardiogram (ECG) recordings, a simple, non-invasive procedure that records the heart's electrical activity. These EEG recordings are combined with the patient's ECG features and other features and are used to predict patients' neurological recovery following cardiac arrest.

We tackle this challenge by using a machine learning based approach, where a series of frequency domain features are extracted from the patient data. We make sure to use data collected from two different timestamps, one from the earliest time from when the data was available from the time of admission of the patient and also from the last available data for the given patient. These EEG frequency features are combined with other patient features like age, sex, location of arrest, etc., and also ECG features. We then employ our hyper-parameter tuned XGBoost machine learning classifier to predict patient outcomes from the extracted features for every patient individually. We will discuss our methodology and results in the coming sections of the study in upcoming sections.

2. Methods

In this section, we will go over the technical details of our study in detail. Please refer to 1 for visualisation of the overall pipeline. We shall first talk about the data preprocessing techniques we employed. From there, we will

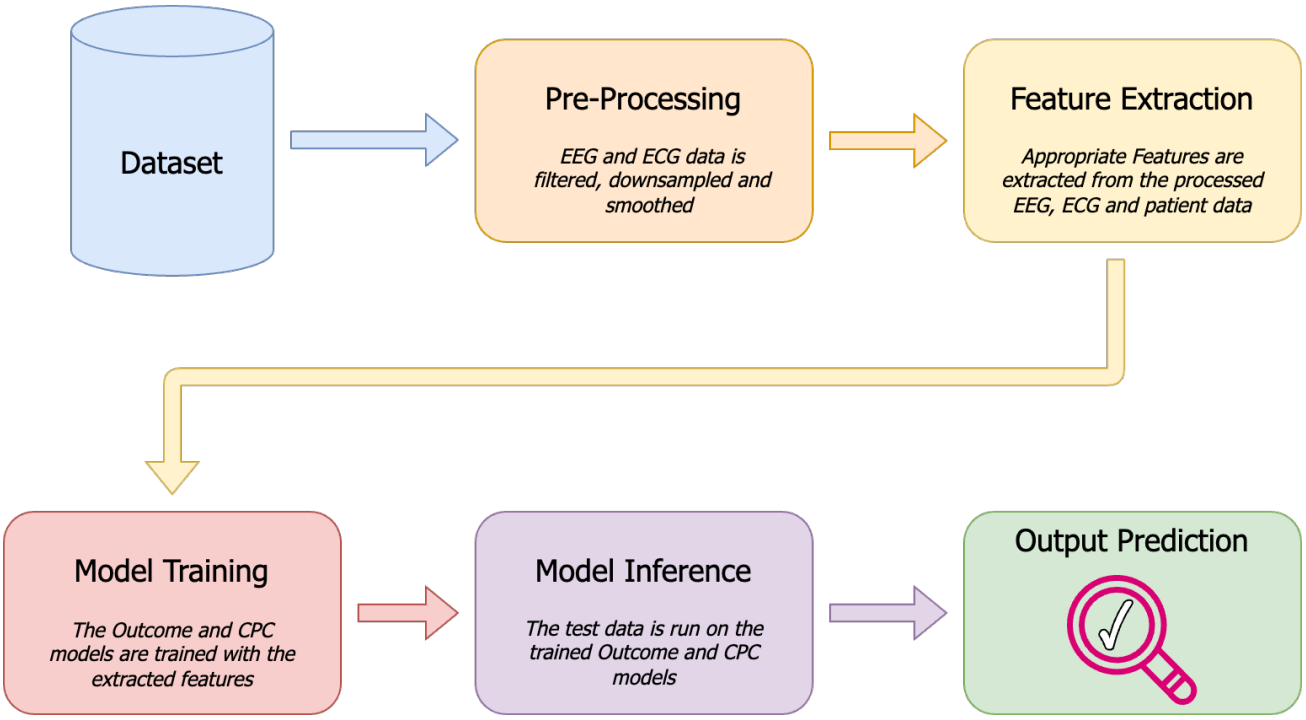

Figure 1. Overall Solution Pipeline

discuss our featurising approach, model selection and finally move on to our model testing techniques used to predict final patient outcomes. All our presented results come from a stratified 5-fold cross-validation, where we ensure all folds have the same ratio of positive and negative outcome patients.

2.1. Data Preprocessing

The dataset used for the challenge was collected from various hospitals across the US and Europe with the help of investigators from the International Cardiac Arrest Research Consortium (I-CARE)[6]. Even though data from 1020 different patients was collected, only 607 patients, or approximately 60% were provided to the teams. The challenge organisers provided teams with different modalities of data, which were divided into signal and clinical data. The signal data consisted of the EEG, ECG, and other modality recordings. In the clinical data, features like age, sex, a hospital identifier, the arrest location, and the type of cardiac rhythm recorded during resuscitation were present. For our approach, we decided to only use EEG and EEG data from the signal data provided and all the features from the clinical data. As different amounts of data were present for different patients, we need to ensure our pre-processing steps work for varying amounts of data. In our approach, we only utilise EEG and ECG data from the first and last hour of the available data for every patient, where the rest of the data is discarded. We believe such pruning is warranted as the model can now look at the difference in feature pattern from when the patient was admitted to when they left to determine the outcome and CPC score without being bogged down by data from other times.

The selected EEG data was first filtered using a bandpass filter of two parts. First, a Butterworth [7, 8] low pass filter is set to a frequency cutoff of 30Hz to eliminate high-frequency noise with more than 30Hz. A high pass filter followed this to eliminate low-frequency noise and baseline wander from the EEG recordings, where the cutoff frequency was set to 0.1Hz. This filtered signal was then resampled down to 60Hz to reduce the size of each recording while preserving only the information we may require. This value of 60HZ was chosen because of the Nyquist sampling theorem[9], which states that a wave must be sampled at twice the value of the frequencies of interest. As we only used frequencies up to 30Hz for creating our feature set physiologically, this resampling rate of 60 HZ was set. This data was then passed through a Savitzky–Golay filter from smoothing to remove any unwanted irregularities in the signal. The signal was finally scaled between -1 and 1 to negate the difference in signal amplitudes across patients. As quite a few channels were given for every patient, only a select 19 channels (C3, C4, Cz, F3, F4, F7, F8, Fp1, Fp2, Fz, O1, O2, P3, P4, Pz, T3,

T4, T5, T6) were chosen for featurisation. A similar pre-processing approach was also used for ECG recordings, where all the signals were filtered, resampled, smoothed, and scaled before featurisation. No pre-processing was performed on the clinical data, and it was directly passed to the featurising step.

2.2. Featurising

The goal of the featurising step is to convert the filtered data of every patient into concise information-rich features which can be used effectively for the classification tasks. The clinical data is first featurised by converting all categorical variables like age to numerical ones and are stacked into a list of 8 numbers. Next, Using the 19-channel EEG signals, we find nine signals made from the difference of the 19 channels (Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F3). Using these nine different channels, we compute five unique features, the mean Delta Power Spectral Density (PSD) with a frequency range from 0.5 to 8, mean Theta PSD with a frequency range from 4 to 8, mean Alpha PSD with a frequency range 8 to 12, mean Beat PSD with frequency range 12 to 30 and finally mean Burst Suppression Ratio (BSR) calculated with a threshold of 0.5 and duration of 1 second. Thus, we have 18 features for every hour for every patient, giving us 36 EEG features for a given patient, as only data from the first and last hour is used. Similarly, for the ECG recordings, the mean and standard deviations across the channels were used as the features, giving us 36 (18x2) unique ECG features. Thus, we are left with 80 (8 from clinical data, 36 from EEG data and 36 from ECG data) unique features to predict for every patient. However, in some cases, patients may only have a single hour of available data or even no available data. In these cases, the available features are padded with "nan" values. Finally, once a trainset of features is created from the training data, a simple imputer is run on the trainset to impute all missing values before training. Thus, for training, we were left with a 485x80 matrix of 485 patients and 80 features for every patient after combining four folds of the 5-fold cross-validation process. Similarly, a 122x80 matrix was used to test the final fold.

2.3. Model Selection

In this section, we will discuss how the right ML classifiers were chosen for both tasks. As the task of outcome prediction was a binary classification task, multiple viable classifier options were present. The challenge was that the classifier had to be simple enough to learn from a training set of approximately 480 data points and have a very low false positive ratio, as per the given challenge metric.

Training	Validation	Test	Ranking
0.34 ± 0.02	0.33	0.38	28/36

Table 1. True positive rate at a false positive rate of 0.05 (the official Challenge score) for our final selected entry (team ComaToast), including the ranking of our team on the hidden validation set. We used 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set.

We experimented with different machine learning models like Support Vector Machines (SVMs), Linear Regressors, K-Nearest Neighbours and various tree-based algorithms. Among the various tree-based machine learning algorithms, Decision Trees (DT), Random Forest [10], and Gradient Boosting Trees (GBT) each offered distinct strengths. However, XGBoost [11] stood out as the superior choice for its versatility and state-of-the-art performance in various tasks. We found that tree-based algorithms performed well for our use due to their simplicity of splitting the data based on feature values. XGBoost took this further by employing extreme gradient boosting. The scalability, flexibility, ability to handle missing data, and custom loss functions made XGBoost the ideal choice for our outcome prediction model. Similarly, with experimentation, we found that using a Random forest regressor worked best for the CPC prediction model.

2.4. Model Evaluation

In our model evaluation phase, several experiments were run to find the most appropriate hyperparameters for the XGBoost-based outcome prediction model and the random forest-based CPC prediction model. We found our ideal values for the outcome prediction model to be 144 for the number of estimators, 'hist' for the tree construction algorithm, 500 for the maximum depth, 100 for the maximum leaves and 0.85 for the L2 regularisation value. The CPC model used values 144 for the number of estimators and 460 for the maximum leaves in each tree.

3. Results

A 5-fold cross-validation setup was used to test both the outcome and CPC models. All folds were stratified to ensure they had the same ratio of good to bad outcome patients. Our models gave an average score of 0.34 for the official challenge score on our local testing. Our final models were tested in the Official Phase of the challenge and obtained a score of 0.381 on the official hidden test set under the name "ComaToast". Further details can be seen in Table 1

4. Discussion and Conclusions

We tested various machine and deep learning approaches to find the best solution to the presented challenge. We believed that deep learning approaches could work well due to the size of the data provided. We tested various convolutional-based architectures like AlexNets[12], ResNets[13] on the filtered signals and also by trying to convert them into 2D spectrograms but could not achieve competitive results. We observed that even though these methods seemed to be giving us very good accuracy, upwards of 80%, they could not meet the criteria of 5% false positive ratio required by the official challenge score. Even when techniques like class bias and custom loss functions were introduced, no significant improvements in results were observable. So, we settled on a relatively simple classifier-based approach using an XGBoost model for our final approach. We prioritised feature extraction and extracted 80 unique features per patient based on clinical information, ECG and EEG recordings.

Our solution achieved an official challenge score of 0.381 on the official leader board on the hidden test set. Even though this score is not very high, the false positive ratio of the implemented solution is very low. Thus, it is a good candidate for assisting doctors, where doctors can shortlist all the patients screened by the model and conduct thorough examinations on them alone. In the future, we wish to improve the model's performance by introducing transformer[14] based predictive models into the solution pipeline. Techniques like pre-training can help build transformer models which are highly accurate while also maintaining a very low false positive rate. To implement these kinds of solutions in real world use cases, the solutions must undergo rigorous testing and incrementally be updated to become multi-modal in nature, as different health centres might use different diagnostic tools for treatment.

Acknowledgments

We thank IHub-Data, IIIT Hyderabad for support.

References

[1] Kong MH, Fonarow GC, Peterson ED, Curtis AB, Hernandez AF, Sanders GD, et al. Systematic review of the incidence of sudden cardiac death in the united states. Journal of the American College of Cardiology 2011;57(7):794–801.

[2] Mehra R. Global public health problem of sudden cardiac death. Journal of Electrocardiology 2007;40(6):S118–S122.

[3] Rundgren M, Westhall E, Cronberg T, Rosen I, Friberg H. Continuous Amplitude-Integrated Electroencephalogram predicts outcome in Hypothermia-treated Cardiac Arrest patients. Critical Care Medicine 2010;38(9):1838–1844.

[4] 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.

[5] 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.

[6] 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 Electroencephalography Database. Critical Care Medicine 2023 (in press); doi:10.1097/CCM.0000000000006074 ;.

[7] Daud S, Sudirman R. Butterworth Bandpass and Stationary Wavelet Transform Filter comparison for Electroencephalography signal. In 2015 6th International Conference on Intelligent Systems, Modelling and Simulation. IEEE, 2015; 123–126.

[8] Gomez V. The use of Butterworth filters for trend and cycle estimation in economic time series. Journal of Business Economic Statistics 2001;19(3):365–373.

[9] Landau H. Sampling, Data Transmission, and the Nyquist rate. Proceedings of the IEEE 1967;55(10):1701–1706.

[10] Breiman L. Random Forests. Machine Learning 2001; 45:5–32.

[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] Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification with Deep convolutional Neural Networks. Advances in Neural Information Processing Systems 2012;25.

[13] 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.

[14] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in Neural Information Processing Systems 2017;30.

Address for correspondence:

U Deva Priyakumar
Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad, Gachibowli, 500032, Hyderabad, India
deva@iiit.ac.in