A Neurological Recovery Prediction Algorithm based on Multi-Feature Extraction and Bagging Aggregation

Ke Jiang[1], Zirui Wang [1], Runze Shen [1], Sibo Wang[1], Yang Liu [1], Yizhuo Feng [1], Xiaohe Lisun[1], Zhenfeng Li [1]

[1]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Science, Beijing, China

Abstract

Objective: For the George B. Moody PhysioNet Challenge 2023. We aim to predict neurological recovery from coma after cardiac arrest (CA) by using multi-domain feature extraction methods and the ensemble model of machine learning (ML) classifiers.

Methods: We employ methods such as wavelet packet transform (WPT) and short-timed fourier transform (STFT) to extract features from electroencephalogram (EEG) signals. Gradient boosting (GB), random forest (RF), and constructive covering algorithm (CCA) are selected to build models for predicting neurological recovery conditions.

Results: Our team AHU lab, obtained a Challenge score of 0.418 (ranked 37th out of 62 teams) on the public validation set and 0.238 (ranked 33th out of 36 teams) on the hidden test set.

Conclusions: We propose a prediction algorithm that combines multiple feature extraction and classifiers for predicting results and providing specific suggestions based on EEG analysis.

1. Introduction

Cardiac arrest (CA) is a severe cardiac condition that often leads to loss of consciousness and poses a threat to life. With over six million occurrences of CA worldwide each year, the survival rate remains below ten percent. Therefore, predicting the probability of awakening after CA is crucial [1]. Patient biometrics such as blood pressure, respiratory rate, body temperature, and blood oxygen saturation are related to this condition, and the continuous evolution of electroencephalogram (EEG) signals over time can provide additional predictive information. However, analyzing and processing these signals effectively is essential to establish a high-quality predictive model.

This study aims to utilize EEG data features as inputs for the model and employ machine learning (ML) techniques to predict the patient's neurological recovery level [2]. The focus is on using a constructive cover algorithm (CCA) as the predictive classifier for the cerebral performance category (CPC). Experimental results demonstrate the effectiveness of this classifier in the prediction process, with a high overall model prediction accuracy.

2. Methods and Materials

As shown in Figure 1, the methods used in this study include feature extraction of EEG signals, feature selection, and dimensionality reduction using the ReliefF feature selection algorithm and principal component analysis (PCA). Building an appropriate classification algorithm to achieve prediction.

2.1. Dataset and Pre-processing

The challenge data was collected from seven academic hospitals in the U.S. and Europe, including 1020 patients [3]. Each patient had hours of continuous EEG data recording, which was divided into three sets: training, validation, and test sets[4].

We performed the following preprocessing steps. Firstly, interpolation was applied to the data from 19 channels (Fp1, Fp2, F7, F8, F3, F4, T3, T4, C3, C4, T5, T6, P3, P4, O1, O2, Fz, Cz, Pz) to obtain 18 channels differences (Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F4, F4-C4, C4-P4, P4-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fz-Cz, Cz-Pz) as the raw data. Secondly, the raw data was filtered using bandpass filters (0.5-30Hz) and resampling to 128 Hz. By differentiating the electrodes and using bandpass filters, they can help remove noise, highlight local variations, and reduce the correlation between electrodes, thereby improving the quality and interpretability of EEG signals.

2.2. Feature Extraction

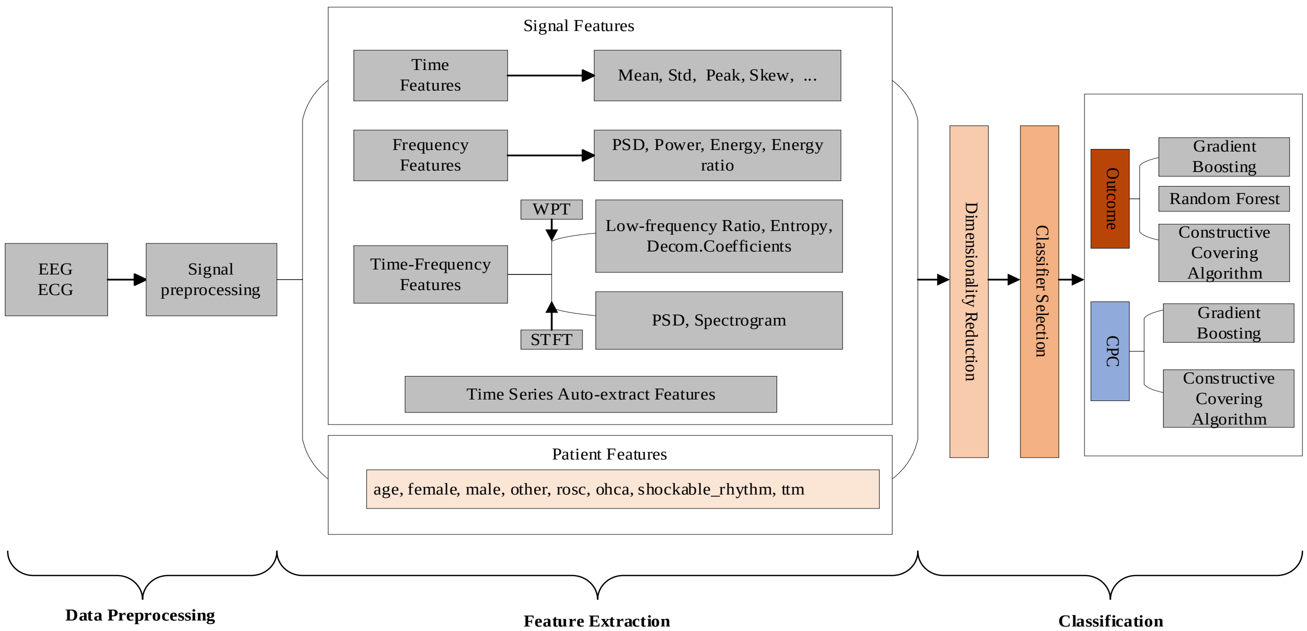

Figure 1. Procedure of the Neurological Recovery Prediction

The common EEG features can be mainly categorized into time-domain, frequency-domain, and time-frequency domain features. Time-domain features are the most intuitive and easily obtained features in EEG signals. We use an $18 \times n$ (relying on the last hour of patients' EEG signal length) array that has been preprocessed as the original data, and feature extraction is performed in the order of each row. We compute the mean, standard deviation, peak value, skewness, and amplitude differences between the electrodes of each row.

Frequency-domain features involve transforming the original time-domain EEG signal into the frequency domain and extracting features from it. We select delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz) frequency bands. The time-domain signal is transformed into these four frequency bands, and corresponding features are extracted. We calculate the energy ratio of different frequency bands, average power, power spectral density and so on.

Time-frequency features involve transforming the time-domain signal using short-time fourier transform (STFT) and wavelet packet transform (WPT) [5]. Unlike the previously mentioned method of extracting features row by row, when using WPT and STFT, we transform the preprocessed data $18 \times n$, which takes all of it as input. In this way, the 18 electrode differences are directly taken as the vertical axis, while the time series is taken as the horizontal axis. This transformation can better represent the local characteristics of the EEG signal in a more comprehensive manner. Besides, we selected db3 wavelet as the basis function and performed 5 levels of decomposition to extract features such as energy, power, decomposition coefficients, and frequency band ratios.

Finally, we utilized the Tsfresh package [6] to automatically extract features from the EEG data. These features are mostly made of statistical features (such as mean, variance, maximum value), kurtosis, skewness, autocorrelation, etc.

In total, we extracted 808 features from the data, including 107 time-domain features, 354 frequency-domain features, 157 time-frequency domain features, and 190 features extended by Tsfresh. These features will serve as inputs for further classification and prediction models.

2.3. Feature Selection

Due to the long duration of the EEG data used for training, it is necessary to select features that are highly correlated with the prediction of CA. We employ two methods for feature selection.

1. After the previous step, we get an array of n (depending on the amount of data for training) $\times$ 808 with labels, we use the ReliefF feature selection method, which calculates weights based on the differences between features, to identify the features that better reflect their importance. At last, we reduced 30 percent of features to get a $n \times 565$ reshaped array for training.

2. Utilizing feature dimensionality reduction to project the features onto dimensions with higher weights related

to the target [7]. We processed 808 features using PCA (n_components = 600) and used the fit function before the prediction of the test set. PCA finds a new set of features through linear transformations that maximally preserves the information in the original data while reducing redundancy in the features.

2.4. Classifier and Evaluation Metrics

We utilized the PyCaret low-code platform, which allows for quick deployment of models and automates the machine learning process [8]. We input the tagged data after PCA processing and performed 10-fold cross-validation to select GB and RF as some of the classifiers with better performance [9]. The detailed test results are shown in Table 1, which is ranked by Area Under the Curve (AUC).

CCA is based on the idea of M-P neurons and achieves coverage by utilizing spherical neighbourhood coverage of samples. This algorithm maps the samples onto the surface of a sphere in an n-dimensional space and increases the number of hidden layer neurons to achieve coverage [10]. It uses inner products instead of Euclidean distance and generates test cases with high coverage rates.

Table 1. The Performance of Model

Model	AUC	Accuracy	Recall	Precision
GB	0.7446	0.6939	0.8362	0.7244
CatBoost	0.7181	0.6690	0.8243	0.7035
AdaBoost	0.7150	0.6665	0.7842	0.7190
XGBoost	0.7061	0.6540	0.7647	0.7096
LightGBM	0.7046	0.6565	0.7681	0.7102
LDA	0.7006	0.6740	0.7765	0.7280
ET	0.6775	0.6514	0.8326	0.6830
RF	0.6423	0.6490	0.8683	0.6704
QDA	0.5545	0.4484	0.2967	0.5986
DT	0.5496	0.5735	0.6448	0.6699
LR	0.5346	0.6266	1.0000	0.6266
NB	0.5214	0.4562	0.3766	0.6276
KNN	0.5049	0.5890	0.7765	0.6428
MLP	0.5020	0.5041	0.5048	0.4141
SVM	0.5000	0.6266	1.0000	0.6266
GPC	0.5000	0.3734	0.0000	0.0000
DC	0.5000	0.6266	1.0000	0.6266

As shown in Figure 2, there are a total of n neurons in the input layer, representing each dimension of the sample data.

For example, x_i^1 represent the First-dimensional data for the sample x_i. The neurons of the hidden layer represent the coverage. Such as c_i^j representing the j th coverage of the i th sample. The output layer aggregates the neurons in the hidden layer that belong to the same category. Each neuron represents a group of covers belonging to the same category. For example, O_t presents the output for the t th class sample.

The evaluation metric is the AUC. AUC is defined as the area enclosed by the axis under the Receiver Operating Characteristic curve (ROC). ROC represents the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) at different classification thresholds. The AUC value ranges from 0 to 1, which is defined in Eq. 1.

$$AUC = \int_0^1 ROCcurve(x)dx \qquad (1)$$

AUC of 0.5 indicates random guessing, AUC > 0.5 suggests better than random guessing, and AUC close to 1 indicates excellent predictive performance with perfect class distinction.

3. Results

In this experiment, we focused only on the data from the last hour as the raw data of the signals. We tested all combinations locally and found that the combination of GB and CCA produced the best results, which ranked by scores, are shown in Table 2. The classifiers CCA and GB exhibit excellent performance in terms of overall performance and the most important scoring metric （0.2956）.

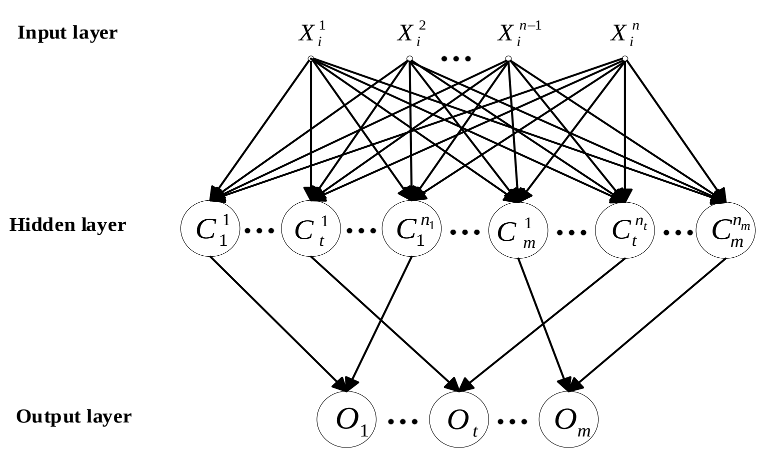

Figure 2. The Architecture of Constructive Covering Algorithm

Table 2. Performance of different combinations of classifiers

Classifier	Score	Accuracy	MSE	MAE
CCA, GB	0.2956	0.6773	4.5274	1.6593
CCA, RF	0.2212	0.6703	3.6833	1.6503
CCA	0.1909	0.6153	5.5879	1.7527
CCA, RF, GB	0.1379	0.6758	4.0300	1.7000
RF	0.1376	0.6593	3.4646	1.5704
GB, RF	0.1196	0.6483	3.5583	1.4586
GB	0.1176	0.7032	4.5824	1.2857

The Challenge scores on both the public training set, hidden validation set, and hidden test set that our final selected entry obtained are shown in Table 3.

Training	Validation	Test	Ranking
0.798	0.418	0.238	33/36

Table 3. The official Challenge score for our final selected entry (team AHU lab), including the ranking of our team on the hidden test 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.

4. Discussion and Conclusions

This study proposes a novel method that combines multiple feature extraction techniques and integrates multiple classifiers for predicting neurological recovery after CA. In the testing phase, only the time domain features scored 0.358 on the public validation set, which rose to 0.379 after adding the frequency domain features, and finally reached 0.418 after adding the time domain decomposition. It can be seen that the time-frequency domain feature has a great weight in the signal processing of EEG. Meanwhile, the increase in the number of channels can also lead to an increase in the score. When we changed the number of channels in the demo code from 4 to 18, the score increased from 0.179 to 0.358. Although our model's performance on the public training dataset is not satisfactory, we can see from Table 2 that the predictive power of CCA far exceeds that of RF and GB, because CCA can introduce diversity by gradually adding and adjusting classifiers, avoiding the problem of relying too much on a single classifier.

Besides, ECG analysis can provide an assessment of the degree of ischemia and heart function, correlating with the level of nerve recovery, and the ability to predict it may be better if ECG signals can be processed accordingly.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62371441), National Natural Science Foundation of China (No. 62331025), National Natural Science Foundation of China (No. 62071451), CAMS Innovation Fund for Medical Sciences (2019-I2M-5-019).

References

[1] 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.

[2] Kang X, Zeng R, Yi H, et al. Prediction of Delivery Mode from Fetal Heart Rate and Electronic Medical Records Using Machine Learning[C]//2022 Computing in Cardiology (CinC). IEEE, 2022, 498: 1-4.

[3] 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; Oct. 2023; doi:10.1097/CCM.0000000000006074.

[4] Goldberger AL, Amaral LA, 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; 101(23): e215-e220.

[5] Morales S, Bowers M E. Time-frequency analysis methods and their application in developmental EEG data[J]. Developmental cognitive neuroscience, 2022, 54: 101067.

[6] Christ M, Braun N, Neuffer J, et al. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package)[J]. Neurocomputing, 2018, 307: 72-77.

[7] Uddin M P, Mamun M A, Hossain M A. PCA-based feature reduction for hyperspectral remote sensing image classification[J]. IETE Technical Review, 2021, 38(4): 377-396.

[8] Iqbal F B, Biswas S, Urba R. Performance analysis of intrusion detection systems using the PyCaret machine learning library on the UNSW-NB15 dataset[D]. Brac University, 2021.

[9] Mahesh B. Machine learning algorithms-a review[J]. International Journal of Science and Research (IJSR).[Internet], 2020, 9(1): 381-386.

[10] Yan, Y., et al. "SMOTE oversampling method for constructive coverage algorithm." Computer Science and Exploration 14.06 (2020): 975-984.

Address for correspondence:

Zhenfeng Li
State Key Laboratory of Transducer Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, No19. North Fourth Ring West Road, Zhongguancun, Haidian District, Beijing, China
lizhenfeng@aircas.ac.cn