aEEG Signal Analysis with Ensemble Learning for Newborn Seizure Detection*

Yini Pan[1], Hongfeng Li[2], Lili Liu[3], Quanzheng Li[2,4,5], Xinlin Hou[3], and Bin Dong[6,4]

[1] Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China
[2] Center for Data Science in Health and Medicine, Peking University, Beijing, China
[3] Department of Pediatrics, Peking University First Hospital, Beijing, 100034 China
[4] Beijing Institute of Big Data Research, Beijing, China
[5] Gordon Center for Medical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[6] Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China

Abstract. Amplitude-integrated EEG (aEEG) has been widely used in neonatal seizure monitoring due to its convenience and broad applicability. However, due to the long length of aEEG signals, detecting seizures in aEEG signals is still a challenging and time-consuming task for experienced clinicians. In this paper, we propose an ensemble learning algorithm to tackle with this problem, aiming to assist clinicians to identify seizures more efficiently and effectively. Firstly, we employ wavelet denoising method to improve the signal-noise rate (SNR) of aEEG signal. Then, to reduce the high dimensionality of aEEG signals while retaining the essential information, we extract global and local features from aEEG signals based on visual features and entropy. Thereafter, we process our data with a feature augmentation algorithm to obtain an extended data set. Finally, an ensemble algorithm is utilized to perform seizure detection. We conduct experiments on real clinical data collected from Peking University First Hospital. Experimental results show that the proposed algorithm achieves excellent performance in seizure detection.

Keywords: Seizure detection · aEEG signal · ensemble learning.

1 Introduction

Neonatal seizure is one of the most critical symptoms of neonatal nervous system and one of the most common clinical manifestations of neonatal neurological abnormalities. The immature brain tissue of the newborn is susceptible to injury, and frequent seizures may cause convulsive brain damage. Therefore, timely detection of neonatal seizures and treatment can reduce the occurrence of convulsive brain injury and improve the neurodevelopmental prognosis.

* Yini Pan, Hongfeng Li and Lili Liu are equally contributed. Corresponding authors: Bin Dong (dongbin@math.pku.edu.cn) and Xinlin Hou (houxinlin66@163.com).

Electroencephalogram (EEG) is the gold standard for the diagnosis of neonatal seizures [11]. Traditional EEG requires that large number of electrodes be placed on the patients' scalp, which is difficult to perform for newborns. In addition, due to the large number of electrodes and lots of information collected, it is easy to adulterate some artifacts while collecting EEG signals. Therefore, the interpretation of conventional EEG requires a large amount of training.

The amplitude of EEG changes during convulsions, and can appear as a transient increase in the upper and lower boundaries on aEEG. In clinical work, it is recommended that neonatal doctors mark the aEEG fragments they believe to be suspicious as "onset" for confirmation in the corresponding original EEG. A number of studies showed that aEEG has comparable sensitivity and specificity with EEG but is much more practical in diagnosis [8]. Thus, we focus on aEEG signals for seizure detection.

Machine learning has been widely studied and applied in medical signal processing. There have been multiple studies on the topic of EEG classification [3]. However, aEEG classification has just developed recently and only a few methods have been proposed for seizure detection [16]. Although existing machine learning methods have achieved high accuracies in some tasks, these methods mainly rely on human interactions in the data preprocessing stage which limits their implementation in actual clinical workflow.

Classifying aEEG signals under a clinically implementable scenario remains a great challenge. Firstly, aEEG signals are often much longer than EEG signals. labeling every seizure onset in an aEEG signal is time-consuming due to its long length. Therefore, we have to deal with very long signals with only global (or weak) labels. Secondly, different from EEG signals, aEEG signals do not have fixed length. As a result, automatically extracting features from aEEG signals with deep learning methods is difficult. Thirdly, aEEG signals have only one channel, which means that they can be easier disturbed by the environment than the EEG signals.

To resolve the above issues, in this paper, we propose a novel method to extract features from aEEG signals and employ an ensemble algorithm to classify aEEG signals. Our algorithm is able to deal with aEEE signals with variable lengths and small size. Experimental results show that the proposed algorithm can achieve a promising performance in the task of aEEG signal classification for seizure detection. The highlights of this paper are listed as follows:

1) We propose using the fuzzy entropy to select the top and negative instances to effectively represent an aEEG signal and introduce feature augmentation to expand data set.
2) The proposed algorithm, i.e. feature selections followed by ensemble learning, achieves excellent classification performance and provides a good balance between the specificity and sensitivity.

The remainder of the paper is organized as follows. Section 2 introduces the collected data. The seizure classification algorithm is described in detail in Section 3. Evaluations of the proposed algorithm on the real clinical aEEG data are presented in Section 4. Finally, we draw conclusions in Section 5.

2 Data Acquisition and Introduction

The aEEG data used in this paper were collected from the neonatal ward of Peking University First Hospital, with 216 newborns with neonatal seizures and 310 newborns with normal aEEG results. aEEG was undertaken by bedside of the newborn and trained nurses in the neonatal ward were responsible for aEEG recording operation and observing clinical seizures during the examination. The aEEG reports were issued by neonatal physician with aEEG experience over 5 years, who were not informed about the condition of the patients and were not involved in the clinical diagnosis and treatment in order to ensure the objectivity of the reports. Fig. 1 demonstrates the characteristics of aEEG recorded.

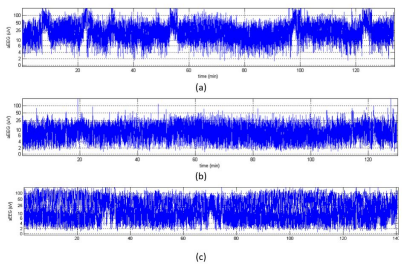

Fig. 1. Characteristics of aEEG. (a) Normal background pattern; (b) Discontinuous background pattern; (c) Burst suppression background pattern.

3 Seizure Detection Algorithm

In this section, we describe the proposed algorithm in detail. The flowchart of the algorithm is shown in Fig. 2.

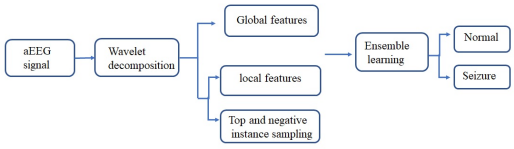

Fig. 2. The flowchart of the proposed algorithm.

3.1 Feature Selection

Feature extraction plays an important role in machine learning tasks. Here, we extract two types of features from the aEEG signals, i.e., global features and

local features. The global features include 4 basic features and total histogram. Basic features include minimum, maximum, skewness and kurtosis of the whole aEEG signal. Total histogram is calculated with the algorithm in [6]. As for the local features, a sliding window with the length of 3 minutes is applied to the aEEG signal to capture sudden changes. The overlap between two successive windows is set to 1.5 minutes. Then, local features are calculate in each window.

Here, we employ several entropies to extract the features of an aEEG signal. These include auto permutation entropy (APE) [13], sample entropy [2], approximate entropy [12], fuzzy entropy [15]. Besides, we introduce spectral entropy to add spectral information into our feature. We also compute the lower and upper border of an aEEG window. The lower border and upper border are defined as the mean of five points near the maximum and minimum values of an aEEG window's envelope, respectively. Together with the entropy features, the desired local features of an aEEG signal are obtained. Furthermore, with the window selection that will be described later in this paper, we get 22 windows for every signal. Thus, there are 236 columns for each local feature in total.

3.2 Top Instance and Negative Evidence for Windows Selection

Generally speaking, seizures can happen at any time in an aEEG signal, and the length of the signal can be very long. Thus, detecting seizures in an aEEG by going through every second of the whole signal is nearly impossible. However, cutting off the entire signal into shorter ones with the same length (for instance, 3 hours) may degrade the performance of classifiers.

Here we perform the following operations: 1) if an aEEG signal exceeds 3 hours, then we assume that there must be at least one seizure during the last 3 hours. So we truncate the aEEG signal and keep the last 3 hours. 2) if an aEEG signal is less than 3 hours, then we pad the upper and lower boundaries of aEEG by the corresponding upper and lower boundaries' mean values with random number in the range of $[0, 1]$. After truncating and padding, we observe that our AUC score improves by 3% compared to the original aEEG series.

Inspired by the top and negative sampling [7], we introduce a new window selection algorithm to tackle the problem caused by the variable aEEG lengths in seizure detection tasks.

We further demonstrate the top and negative windows of aEEG signals mentioned above in Fig. 3. The upper curve is the original aEEG data and the lower curve is the aEEG signal with selected windows. We use the red curve to denote the top seizure-like windows chosen by our algorithm, the green curve to denote the negative windows and the red arrow to denote the seizure onset pointed by doctors. As the figure shows, top windows appear in those places where there are abrupt changes, which exactly corresponds to the most probable seizure onsets.

We calculate certain entropies of the signal within each window and sort the windows according to their corresponding entropy values. Then, only the first k windows and last k windows are retained for further feature extraction. After many trails, we find optimal $k = 11$ and the best entropy is fuzzy entropy. Classification results with different window selection entropy are shown in Table. 1.

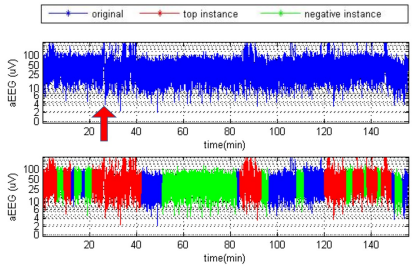

Fig. 3. aEEG signals with selected windows. The upper curve is the original signal and the lower curve is the signal after window selection. Red area indicates the top windows, green area indicates the negative evidences, and the red arrow indicates where seizure happens confirmed by clinicians.

Table 1. Classification results with different window selection entropy

Method	Accuracy	Specificity	Sensitivity	AUC
approximate entropy	77.38%	77.91%	75.27%	75.27%
sample entropy	77.24%	77.51%	75.10%	75.10%
shannon entropy	76.83%	76.95%	74.63%	74.41%
fuzzy entropy	79.73%	79.96%	79.01%	78.34%
permutation entropy	78.25%	78.55%	76.29%	76.29%

3.3 Feature Augmentation

Data augmentation is a powerful method in supervised learning. Here, we apply interpolation and extrapolation to augment our train set and measure on our test set as: 1) Interpolation: for each data point x, find k nearest neighbors with the same label and interpolate by $x_{new} = (x - x_k) * \lambda + x_k$; 2) Extrapolation: for each data point x, find k nearest neighbors with the same label and extrapolate by $x_{new} = (x_k - x) * \lambda + x$. Here, k and λ are trainable parameters and we set $k = 9$ and $\lambda = 0.73$. Experimental results show that extrapolation performs better than interpolation and improves the AUC score by 0.6%.

3.4 Classification Algorithms

When merging features, we apply min-max normalization to data to improve the speed and accuracy. Then, with the extracted features, we adopt several algorithms, i.e., support vector machine (SVM) [1], logistic regression (LR) [10], adaptive boosting (AdaBoost) [14], Xgboost [5], random forest (RF) [4] and gradient boosting trees (GBDT) [9], as classifiers to classify the aEEG signals.

3.5 Ensemble Algorithm

Ensemble learning algorithm combining multiple machine learning algorithms together commonly leads to better performance than any of the individual classifier. Therefore, we find the best single classifier and use bagging strategy to

form an ensemble learning algorithm for aEEG signal classification. Experimental results demonstrate that the ensemble algorithm can indeed further improve the classification accuracy.

4 Experiments and Analysis

In this section, we conduct experiments on the real data set described in Section 2 using the algorithm proposed in the previous section.

Prior to feature extraction, the aEEG signals are denoised with a band-pass filter to cut-off frequencies lower than 0.3Hz and higher than 30Hz to remove artifacts and a four-order wavelet is utilized to further improve the quality of the signals. In the experiments, we randomly divide the entire data set into three subsets, i.e., training, validation and testing sets. Specifically, we first randomly divide the entire data set into two parts with proportions of 80% and 20%, respectively. The 20% portion (105 samples) is used for testing. Then, 50 samples is further taken from the 80% portion for validation. Thus, the remaining 421 samples are used for training. We repeat the procedure 15 times and calculate the mean classification accuracy. All the experiments are carried out on a Dell laptop with Python 3.6 and MATLAB 7.0.

4.1 Feature Evaluation

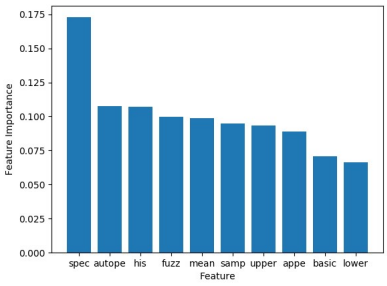

Fig. 4. Feature importance

Fig. 4 demonstrates the importance of features. We observe that the spectral entropy is of the most importance. Features like basic feature, lower border and approximate entropy are not as important as expected. As for the fuzzy entropy, it may be more significant in local context for window chosen in a single signal rather than sample comparison. The reason is that different patients may have a very different background signal level but possess a similar fuzzy entropy score.

4.2 Comparison of Classification Algorithms

Commonly, in medical image processing, positive means ill and negative means normal. We use following metrics for model evaluation and comparison: 1) $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100$; 2) $Sensitivity = \frac{TP}{TP+FN} * 100$; 3) $Specificity = \frac{TN}{TN+FP} * 100$; 4) Receiver operating characteristic curve (ROC). Here, TP indicates true positive, TN indicates true negative, FP indicates false positive and FN indicates false negative.

Table 2. Classification results with various classifiers.

Model	Accuracy	Specificity	Sensitivity	AUC
SVM	73.28%	73.46%	71.01%	71.01%
LR	74.08%	76.45%	78.26%	60.68%
Adaboost	76.79%	76.77%	75.46%	75.46%
Xgboost	78.12%	78.77%	78.70%	75.91%
Random Forest	74.53%	76.45%	76.54%	74.57%
GBDT	79.23%	79.27%	77.67%	77.67%
Ensemble	79.73%	79.96%	79.01%	78.34%

With the extracted features for aEEG signals, we compare the performance achieved with the six classification algorithms. Results are shown in Table 2. From Table 2 we conclude that the proposed feature extraction algorithm is effective for extracting semantic features from aEEG signals. Note that the ensemble algorithm made up of bagging of GBDTs obtains the best results in all categories. Table. 3 shows that our work is better than the past researches and improves the accuracy by about 3%.

Table 3. Classification results compared with other researches.

Model	Accuracy	Specificity	Sensitivity	AUC
Yu Wang[6]	76.31%	76.59%	74.59%	74.91%
Tao Yang[16]	76.23%	76.30%	74.86%	74.86%
Our work	79.73%	79.96%	79.01%	78.34%

5 Conclusions

aEEG can be effective for screening newborns with high risk factors of neonatal seizures. In this paper, we propose a novel algorithm to extract features from aEEG signals and performed seizure classification using an ensemble method. Firstly, we utilize a method based on visual features and entropy to extract the global and local features from aEEG signals. Then, we expand our data with feature augmentation method. Finally, a bagging of GBDT models is employed to perform seizure detection. Experimental results on a real aEEG data set show

that the ensemble algorithm can achieve a promising performance with the classification accuracy of 79.73%, the specificity score of 79.96%, the sensitivity score of 79.96% and the AUC score of 78.34%.

Acknowledgement. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0910700 and in part by the National Natural Science Foundation of China under Grants 11701018, 11831002 and 81801778.

References

1. Adankon, M.M., Cheriet, M.: Support vector machine. Computer Science **1**(4), 1–28 (2002)
2. Bai, D., Qiu, T., Li, X.: [the sample entropy and its application in eeg based epilepsy detection]. Journal of Biomedical Engineering **24**(1), 200 (2007)
3. Beyli, E.D.: Combined neural network model employing wavelet coefficients for EEG signals classification. Academic Press, Inc. (2009)
4. Breiman, L.: Random forest. Machine Learning **45**, 5–32 (2001)
5. Chen, T., He, T., Benesty, M., et al.: Xgboost: extreme gradient boosting. R package version 0.4-2 pp. 1–4 (2015)
6. Chen, W., Wang, Y., Cao, G., Chen, G., Gu, Q.: A random forest model based classification scheme for neonatal amplitude-integrated eeg. Biomedical engineering online **13**(2), S4 (2014)
7. Courtiol, P., Tramel, E.W., Sanselme, M., Wainrib, G.: Classification and disease localization in histopathology using only global labels: A weakly-supervised approach. arXiv preprint arXiv:1802.02212 (2018)
8. Frenkel, N., Friger, M., Meledin, I., Berger, I., Marks, K., Bassan, H., Shany, E.: Neonatal seizure recognition–comparative study of continuous-amplitude integrated eeg versus short conventional eeg recordings. Clinical Neurophysiology **122**(6), 1091–1097 (2011)
9. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. Annals of Statistics **29**(5), 1189–1232 (2001)
10. Menard, S.: Applied logistic regression analysis. Technometrics **38**(2), 192–192 (2002)
11. Nagarajan, L., Palumbo, L., Ghosh, S.: Classification of clinical semiology in epileptic seizures in neonates. european journal of paediatric neurology **16**(2), 118–125 (2012)
12. Pincus, S.: Approximate entropy (apen) as a complexity measure. Chaos An Interdisciplinary Journal of Nonlinear Science **5**(1), 110 (1995)
13. Ping, X., Xiuli, W., Yihao, D.: Feature extraction method of semg based on auto permutation entropy [j]. PR&AI **27**(6), 496–501 (2014)
14. Schapire, R.E.: The boosting approach to machine learning: An overview. In: Nonlinear estimation and classification, pp. 149–171. Springer (2003)
15. Xiang, J., Li, C., Li, H., Cao, R., Wang, B., Han, X., Chen, J.: The detection of epileptic seizure signals based on fuzzy entropy. Journal of Neuroscience Methods **243**, 18–25 (2015)
16. Yang, T., Chen, W., Cao, G.: Automated classification of neonatal amplitude-integrated eeg based on gradient boosting method. Biomedical Signal Processing and Control **28**, 50–57 (2016)