2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Semi-Supervised Few-Shot Learning Model for Epileptic Seizure Detection

Zheng Zhang[1], Xin Li[1], Fengji Geng[2], and Kejie Huang[1]

Abstract— **In the past decade, the rapid development of machine learning has dramatically improved the performance of epileptic detection with Electroencephalography (EEG). However, only a small amount of labeled epileptic data is available for training because labeling requires numerous neurologists. This paper proposes a one-step semi-supervised epilepsy detection system to reduce the labeling cost by fully utilizing the unlabeled data. The proposed neural network training strategy enables a more robust and accurate decision boundary by forcing the consistency of the double predictions on the same unlabeled data. The results show that the Area Under Receiver Operating Characteristic (AUROC) curves of our proposed model are 10.3% and 4.9% higher than the supervised methods on CHB-MIT and Kaggle datasets, respectively.**

Index Terms— **epileptic seizure detection, semi-supervised learning, EEG, machine learning, double predictions**

I. Introduction

Epilepsy is a common chronic disease that causes a repetitive and irregular abnormal discharge of brain neurons, leading to transient brain dysfunction. Nearly one percent of people worldwide have epilepsy, and the number of patients is growing at a rate of 5% every year. Therefore, studying epilepsy, including clinical trials, drug treatments, and signal analysis, has received significant attention. Due to high detection accuracy, Deep Neural Network (DNN) models, including Long-Short-Term-Memory (LSTM) [1], Convolution Neural Networks (CNNs) [2], and Graph Neural Networks (GNNs) [3], have been widely used to analyze epileptic signals. However, the models trained on publicly available datasets are inapplicable to actual patients due to personality and individual differences. Therefore, the DNN model is often trained with the patients' self-epileptic data to achieve higher performance. However, labeled epileptic data are always insufficient because of the shortage of epilepsy experts.

Recently, semi-supervised learning has demonstrated powerful capabilities to reduce the demand of the labeled data with comparative high recognition accuracy [4] [5]. In the field of seizure detection or prediction systems, Truong et al. [6] trained a GAN in an unsupervised manner, and the discriminator was used as a feature extractor. The model achieved the Area Under Receiver Operating Characteristic (AUROC) curves of 77.68% and 75.47% for the CHB-MIT scalp EEG dataset [7] and Freiburg Hospital intracranial EEG dataset [8], respectively. Ahmed et al. [9] employed a mixing of Variational Autoencoder (VAE) together with a supervised classifier to achieve the 99% overall accuracy on the Department of Epileptology at Bonn University (DEBN) dataset [10].

Though conventional works have achieved promising results with semi-supervised learning methods, sufficiently large label data are still required for the training. The reason is that the existing semi-supervised methods for seizure detection require two steps. Unlabeled data are used for feature extraction in the first step, and then in the second step, labeled data are used for classification training. To significantly reduce the amount of the labeled data, we propose a one-step epilepsy detection system. A random enhancing method is applied to both the unlabeled data and labeled data to generate two different representations in our system. One is for training the student model, and another is for the teacher model. Our system learns from a small amount of labeled data while trying to give a robust recognition on a large amount of test data, which can minimize the diagnostic tasks of specialists. The proposed method is evaluated on the CHB-MIT dataset, and the UPenn & Mayo Clinic's Seizure Detection Challenge database [11].

The rest of the paper is organized as follows: Section II introduces the datasets and our proposed semi-supervised learning method. Section III demonstrates the evaluation results of our approach. The comparison with other state-of-the-art methods is also discussed in Section III. Finally, the conclusion is drawn in Section IV.

II. PROPOSED METHOD

In conventional supervised training, abundant labeled data are always needed to learn robust decision boundaries. The proposed system enhances the data twice to training the teacher network and student network, respectively. The design minimizes the difference between the two predicted results of the unlabeled data after data enhancements in the training process. As a result, the boundary of decision-making can be pushed away from the labeled data points, increasing the robustness of the model.

A. Datasets

Table I and II show the summary of two EEG datasets used in this paper. The UPenn and Mayo Clinic's Seizure Detection Challenge database is provided in a competition by Kaggle (2014), which consists of intracranial EEG (iEEG) recordings of four dogs and eight patients. The canine recordings were sampled from 16 electrodes at 400 Hz, and

[1]Zheng Zhang, Xin Li, and Kejie Huang are with College of Information Science and Electrical Engineering, Zhejiang University, 38 Zheda Road, Hangzhou, 310027, China {zheng_zhang, i.lixin, huangkejie}@zju.edu.cn

[2]Fengji Geng is with the College of Education, Zhejiang University, Hangzhou, China gengf@zju.edu.cn

TABLE I
SUMMARY OF DATASET PROVIDED BY KAGGLE

	No. of subjects	No. of channels	Ictal data(s)	Interictal data(s)	Unlabeled data(s)
Kaggle	12	16-72	2477	23445	32915

TABLE II
SUMMARY OF CHB-MIT DATASET

	No. of patients	No. of channels	No. of seizures	Interictal hours(h)
CHB-MIT	10	23	52	215.7

the human recordings were sampled from varying numbers of electrodes (ranging from 16 to 72) at 500 Hz or 5 kHz. All data were pre-organized into 1s iEEG segments as ictal or interictal. Note that the unlabeled data in Table I is the test dataset in the competition.

CHB-MIT dataset contains 23 scalp EEG (sEEG) records collected from 22 pediatric patients with 844 hours of continuous EEG signals, including 198 seizures. These EEG data are obtained at the sampling rate of 256 Hz. Similar to the binary classification problem in the Kaggle dataset, the preictal period is defined as 30 minutes before seizures onset, and the interictal period is at least 4 hours away from seizures. We combine the closed seizures that the interval of seizures is fewer than 30 minutes. In pre-processing, we divide the EEG records in the CHB-MIT into 30 seconds fragments similar to [6].

B. Pre-processing

The EEG signal frames are first filtered through a band-pass filter to avoid the interference of baseline and remove the high-frequency data containing little information. After filtering, the one-second segments are transformed by Fast Fourier Transform (FFT), denoted as $X(f)$. Power spectrum analysis based on FFT provides the center frequency of rhythmic fluctuations without phase information. The power spectrum of the segment is mapped onto the Mel scale [12] to produce the features of the EEG signals. The power spectrum $|X(f)|$ with a filter bank with L filters is defined as

$$W_l(k) = \begin{cases} \frac{k-o(l)}{c(l)-o(l)} & o(l) \leq k \leq c(l) \\ \frac{h(l)-k}{h(l)-c(l)} & c(l) < k \leq h(l) \\ 0 & else \end{cases} \quad (1)$$

Final features $m(l)$ are obtained by performing the logarithm to the results of each triangle filter, which is defined as

$$m(l) = \sum_{k=o(l)}^{h(l)} W_l(k) \, |X(k)| \qquad l = 1, 2, ..., L \quad (2)$$

where the center frequency $c(l)$ of each triangle filter is uniformly spaced in Mel-scale, and $o(l)$, $c(l)$, and $h(l)$ are the lowest frequency, center frequency, and the highest frequency for the l_{th} filter, respectively. The relationship between adjacent triangle filters is given by

$$c(l) = h(l-1) = o(l+1) \qquad l = 2, 3, ..., L-1 \quad (3)$$

For a 30s segment, Short-Time Fourier Transform (STFT) and power spectrum on Mel scale of frequency are applied to transfer the data to the time-frequency domain.

C. Semi-supervised Learning Method

As shown in Fig. 1, Mean Teacher taken from [13] is used as a semi-supervised classifier in this paper. When there is a lack of high-quality labels, the limited training data cannot describe the data distribution completely. Therefore, additional distribution information is supplemented by using unlabeled data with consistency cost that is defined as

$$L_{consistency}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left\| f(x_i, \theta, \eta) - f(x_i, \theta', \eta') \right\|^2 \quad (4)$$

Where $f(x, \theta, \eta)$ represents the prediction of the student model (with weights θ and noise η). The other $f(x, \theta', \eta')$ is the prediction of the teacher model.

The semi-supervised model is trained as follows. Firstly, the input batch, which consists of half of unlabeled data and half of labeled data, is processed by a random augmentation to generate two different representations for each fragment. Secondly, the student model and the teacher model take these two representations to obtain each prediction. Note that the student model has the same structure as the teacher model but with different parameters. The total loss is the combination of classification loss and consistency loss, while we only accept classification loss for labeled data. Finally, the backpropagation only works on the student model. The teacher model uses the Exponential Moving Average (EMA) weight of the student model during training, which is defined in (5).

$$\theta'_t = \alpha \theta'_{t-1} + (1-\alpha)\theta_t \quad (5)$$

To adapt EEG fragments with different lengths, we achieve augmentation by adding noise to the model's inputs to simulate the Gaussian noise, which is inevitable when recording the EEG signals.

As the features have been extracted in the pre-processing, a Convolution Neural Network (CNN) is applied as the classifier. The proposed CNN consists of three convolutional layers followed by a fully connected layer with a Softmax activation function. Meanwhile, we use Mean Squared Error (MSE) as the classification loss.

$$L_{classification}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left\| y_i - f(x_i, \theta, \eta) \right\|^2 \quad (6)$$

After training, the teacher model is used to complete prediction and evaluation, as recommended in [13].

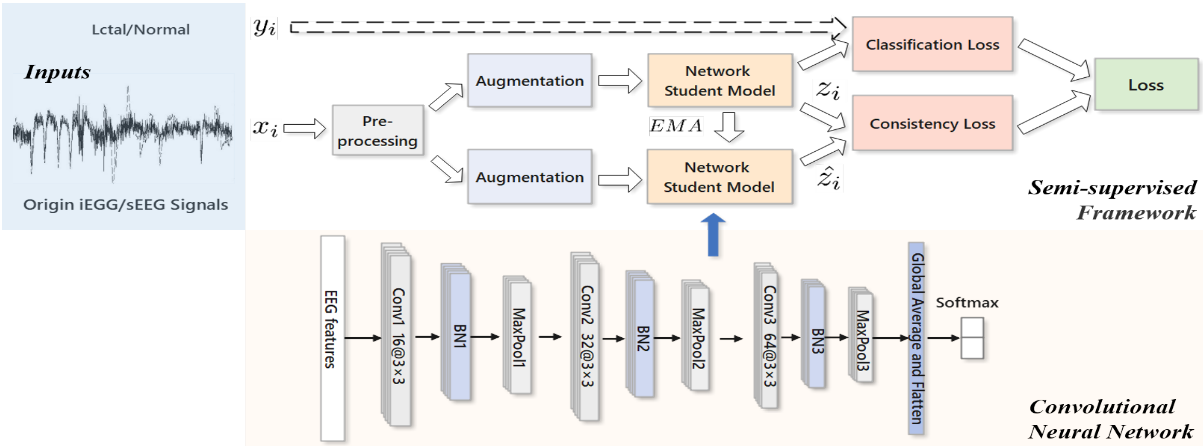

Fig. 1. Semi-supervised learning method

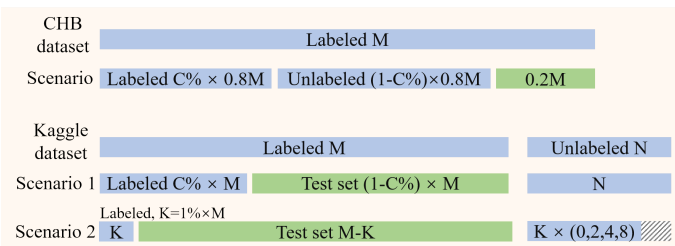

Fig. 2. Evaluation method. The blue and green blocks indicate the data used for training and testing, respectively. Gray slashed block means data are not used in that scenario.

D. System evaluation

Since the two datasets used in this paper are constructed in different ways, and the unlabeled data can be directly represented by the test data in the Kaggle competition, we carried out different evaluation strategies for the two datasets, as shown in Fig. 2. When testing the proposed method on CHB-MIT, the dataset is processed as follows. 20% of the whole dataset is chosen randomly for verification. Furthermore, we remove a certain number of labels from the training set, leaving C% (e.g., 1% and 5%) of the training data with labels, and the remaining data is treated as unlabeled data. The evaluation is equal to 5-fold cross-validation when C%=100%, so we can compare ours with other state-of-the-art methods.

As for the training and test set provided by Kaggle, the test set at the competition is treated as unlabeled data to train the model. Moreover, the training set, which has labels, would be divided into training and test sets. We conduct experiments from two scenarios. The first is that C% (e.g., 10% and 30%) of labeled data is applied to train the model and the rest for evaluation. Note that when C%=90%, this situation is equal to 10-fold cross-validation. The second is that the model uses a different number of unlabeled data while training with only 1% of labeled data, which can test the impact of the increase in unlabeled data. We adopt AUROC value as the measurement of system performance.

TABLE III

COMPARISON OF AVERAGE PERFORMANCE USING CHB-MIT

	1%	5%	10%	100%
Truong et al. [6]				83.79
Zeng et al. [17]				**99.6**
Supervised-only	75.69	81.05	91.35	98.57
Proposed method	**85.99**	**94.88**	**97.04**	**98.80**

TABLE IV

COMPARISON BETWEEN THE PROPOSED AND SUPERVISED-ONLY METHOD ON EACH PATIENTS USING CHB-MIT

Subject	Proposed method				Supervised-only			
	1%	5%	10%	100%	1%	5%	10%	100%
chb01	97.9	99.0	99.6	99.9	88.6	96.4	89.9	99.8
chb03	79.6	99.2	99.1	99.1	62.3	59.9	93.4	99.1
chb05	84.6	92.8	97.0	99.1	77.0	68.0	91.4	99.9
chb09	94.3	99.0	99.8	99.4	77.1	96.1	98.9	96.3
chb10	73.8	92.1	94.8	98.1	52.2	78.6	84.7	99.4
chb14	74.3	85.3	89.4	97.8	70.6	59.9	75.2	96.9
chb18	78.5	91.2	96.0	97.6	68.3	67.2	90.2	94.9
chb20	93.7	99.6	99.9	99.6	97.8	96.5	98.8	99.9
chb21	87.7	90.6	94.4	97.0	74.9	89.4	92.1	99.1
chb23	94.9	99.5	99.9	99.9	87.5	98.1	98.5	99.9

III. EXPERIMENTAL RESULT

In subsequent chapters, supervised-only represents the classification system using the same neural network as the proposed method but does not apply the semi-supervised learning framework, which means it can only learn from the labeled data.

A. Performance on CHB-MIT

Table III shows the results of the proposed method compared with the supervised-only method and other state-of-the-art methods using the CHB-MIT dataset. Table IV lists the AUROC values on each patient when providing a different amount of labeled data. Our method has outperformed the supervised-only method on almost all patients. When 1% labels are available, our method achieves 85.99% AUROC on average, which is 10.3% and 2.2% higher than the supervised-only method and the method proposed by Truong

TABLE V

COMPARISON OF AVERAGE PERFORMANCE USING KAGGLE DATASET

	10%	30%	50%	70%	90%
Wang et al. [14]					98.39
Truong et al. [15]					95.85
Hills [16]					95.91
Supervised-only	94.52	97.04	98.28	98.87	99.12
Proposed method	**96.45**	**97.59**	**98.60**	**98.97**	**99.38**

TABLE VI

COMPARISON BETWEEN THE PROPOSED AND SUPERVISED-ONLY METHOD USING KAGGLE DATASET

Subject	Proposed Method				Supervised-only			
	10%	30%	50%	70%	10%	30%	50%	70%
Dog1	94.3	96.6	97.5	96.8	90.2	96.6	96.1	98.0
Dog2	99.1	98.9	99.0	99.7	98.5	99.5	99.3	99.9
Dog3	99.5	99.5	99.4	99.6	98.8	99.5	99.5	99.0
Dog4	97.3	96.5	96.8	97.8	95.2	96.8	96.7	97.6
Pat1	82.5	88.3	99.4	100	72.7	80.7	96.4	98.1
Pat2	99.5	99.5	99.4	99.8	98.2	99.6	99.6	98.3
Pat3	95.2	95.7	95.5	96.8	94.8	95.6	95.6	99.3
Pat4	92.6	98.8	99.0	99.1	91.4	99.0	98.6	98.3
Pat5	99.5	99.7	99.7	99.8	98.3	99.6	99.7	99.4
Pat6	99.9	99.9	99.9	99.9	99.7	99.9	99.9	100
Pat7	99.6	99.1	99.0	99.3	99.2	99.1	99.5	100
Pat8	98.3	98.6	98.5	99.0	97.3	98.5	98.5	98.6

et al. [6], respectively. The proposed method achieves the performance of 98.80% AUROC value when all labels are provided. The results demonstrate that the proposed system can effectively improve the detection performance, especially when few labels can be obtained.

B. Performance on UPenn and Mayo Clinic's Seizure Detection Challenge database

- Scenario 1: As shown in Table V, the results indicate that the proposed method can achieve 96.45% AUROC value when only use 10% of the labeled data, which is 0.6% higher than Truong et al. [15].
- Scenario 2: Table VII shows the performance while providing the model different amounts of unlabeled data. When ×8 unlabeled data is provided, our proposed method can reach 91.2% AUROC, which is 4.9% higher than the supervised method. It can be inferred that our proposed method can learn more information from additional unlabeled data to improve the prediction results.

IV. CONCLUSION

In this paper, a one-step semi-supervised seizure detection system with recorded EEG signals has been proposed. We have shown that extracting the power spectrum of Mel scale of frequency and the Mean Teacher can achieve superior performance when there are few labels. Meanwhile, we have used additional unlabeled data to obtain a better score than the benchmark, which shows a possible future promotion direction. It is feasible to use easy-to-obtain unlabeled data to improve the accuracy of seizure prediction and detection.

TABLE VII

PREDICTION PERFORMANCE WHILE APPLYING DIFFERENT AMOUNTS OF UNLABELED DATA ON KAGGLE DATASET

	×0	×2	×4	×8
Supervised-only	0.863	0.863	0.863	0.863
Proposed method	**0.869**	**0.892**	**0.905**	**0.912**

REFERENCES

[1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, pp. 1735-1780, 1997.

[2] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in neural information processing systems Curran Associates Inc. 2012.

[3] F. Scarselli, A. C. Tsoi, M. Gori, and M. Hagenbuchner, "Graphical-based learning environments for pattern recognition," Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR). Springer, Berlin, Heidelberg, 2004.

[4] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "MixMatch: A Holistic Approach to Semi-Supervised Learning," arXiv:1905.02249 [cs, stat], Oct. 2019, Accessed: Mar. 19, 2021. [Online]. Available: http://arxiv.org/abs/1905.02249.

[5] Q. Xie, Z. Dai, E. Hovy, M.-T. Luong, and Q. V. Le, "Unsupervised Data Augmentation for Consistency Training," arXiv:1904.12848 [cs, stat], Nov. 2020, Accessed: Mar. 19, 2021. [Online]. Available: http://arxiv.org/abs/1904.12848.

[6] N. D. Truong, L. Zhou and O. Kavehei, "Semi-supervised Seizure Prediction with Generative Adversarial Networks," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 2369-2372.

[7] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[8] U. o. Freiburg, "EEG Database at the Epilepsy Center of the University Hospital of Freiburg, Germany," 2003. [Online]. Available: http://epilepsy.uni-freiburg.de

[9] A. M. Abdelhameed and M. Bayoumi, "Semi-Supervised EEG Signals Classification System for Epileptic Seizure Detection," IEEE Signal Process. Lett., vol. 26, no. 12, pp. 1922–1926, Dec. 2019.

[10] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Phys. Rev. E, vol. 64, no. 6, 2001, Art. no. 061907.

[11] J. B. Wagenaar, B. H. Brinkmann, Z. Ives, G. A. Worrell and B. Litt, "A multimodal platform for cloud-based collaborative research," 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), San Diego, CA, USA, 2013, pp. 1386-1389.

[12] V. Tyagi and C. Wellekens, "On desensitizing the mel-cepstrum to spurious spectral components for robust speech recognition," in Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on. IEEE, 2005, vol. 1, pp. I–529.

[13] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," arXiv:1703.01780 [cs, stat], Apr. 2018, Accessed: Mar. 19, 2021. [Online]. Available: http://arxiv.org/abs/1703.01780.

[14] D. Wang, Y. Fang, Y. Li and C. Chai, "Enhance feature representation of electroencephalogram for Seizure detection," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 1230-1234.

[15] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, J. Yang, A. Faulks, and O. Kavehei, "Supervised learning in automatic channel selection for epileptic seizure detection," Expert Systems with Applications, vol. 86, pp. 199–207, 2017.

[16] M. Hills, "Seizure detection using fft, temporal and spectral correlation coefficients, eigenvalues and random forest," in Technical Report. GitHub, 2014.

[17] D. Zeng, K. Huang, C. Xu, H. Shen and Z. Chen, "Hierarchy Graph Convolution Network and Tree Classification for Epileptic Detection on Electroencephalography Signals," in IEEE Transactions on Cognitive and Developmental Systems.