Semi-supervised Domain Adaptation for EEG-based Epileptic Seizure Classification

Xinru Chen and Dongrui Wu
School of Artificial Intelligence and Automation
Huazhong University of Science and Technology, Wuhan, China
Email: xrchen@hust.edu.cn, drwu@hust.edu.cn

***Abstract*—Epilepsy is a neurological disorder, which is usually detected by electroencephalogram (EEG) signals. Convolutional neural networks are the most widely used deep learning approaches in seizure classification. Automatic EEG-based epilepsy classification faces the challenges of inadequate labeled data, class imbalance, and individual differences. To address these issues, this paper proposes APE-cosine, a semi-supervised domain adaptation approach for cross-patient seizure classification. We adopted three networks as feature extraction layers and added data augmentation strategies to alleviate the influence of class imbalance. Experiments on three different convolutional neural networks demonstrated that APE-cosine outperformed state-of-the-art algorithms, including the original APE algorithm.**

***Index Terms*—Seizure classification, electroencephalography, class imbalance, semi-supervised domain adaptation, convolutional neural network**

I. Introduction

Epilepsy is a chronic non-communicable disease of the brain that affects people of all ages. Seizures negatively impact the patients' physical and emotional health, and hence their quality of life. Epileptic seizure classification is important to its diagnostic and treatment.

Electroencephalography (EEG) is the gold standard for epilepsy classification [1]. However, visual examination of EEG is labor intensive, and may produce inconsistent results [2]. Hence, automatic epileptic seizure classification has attracted great research interests.

Automatic epilepsy classification usually consists of four steps: signal acquisition, signal preprocessing, feature extraction, and classification. When deep learning (DL) is used, feature extraction and classification can be integrated into a single neural network. Popular deep learning approaches for EEG-based automatic epilepsy classification can be partitioned into three categories [3]: 1) convolutional neural networks (CNNs), 2) recurrent neural networks, and, 3) autoencoders.

CNNs are the most widely used deep learning approaches in seizure classification [4]. SeizureNet [5] uses Fourier transform to convert EEG signals into a time-frequency map, and then convolutional layers for epilepsy classification. Temporal Graph Convolutional Networks [6] convert raw EEG signals into a temporal graph and then use five different layers of CNNs for classification. Additionally, CNN models that have shown promising performances in brain-computer interface applications, e.g., EEGNet [7], Deep ConvNet [8] and Shallow ConvNet [8], may also be used for EEG-based seizure classification.

Challenges in EEG-based seizure classification include:

1) *Inadequate labeled samples*. In clinical practice, the acquisition of EEG signals is relatively easy, but their labeling is time-consuming. Therefore, typically in addition to a small amount of labeled EEG samples, there are many more unlabeled EEG samples, which also contain useful information and should be utilized. Semi-supervised learning could be a solution.
2) *Class imbalance*. Epilepsy datasets typically have high class-imbalance, as generally seizure periods last much shorter than non-seizure periods. If not properly coped with, the minority class samples may be easily classified into the majority class [9].
3) *Individual differences*. Different patients usually have different seizure patterns, and hence cross-patient seizure classification is challenging. Transfer learning can be used to reduce the discrepancy in data distribution between the training data (source domain) and testing data (target domain), and simultaneously solve the problems of individual differences and inadequate labeled training data [10].

Many recent works [11] have demonstrated the effectiveness of semi-supervised learning and transfer learning. However, few tried to integrate both of them [12].

This paper considers automatic EEG-based cross-patient seizure classification. We compare four training algorithms [directly combine source and target domain data for training (S+T) [13], Minimum Entropy Regularization (ENT) [14], Minimax Entropy (MME) [15], and Attract, Perturb, and Explore (APE) [16]] on three CNNs (EEGNet [7], Deep ConvNet, and Shallow ConvNet [8]) architectures, and also improved APE for better performance.

The remainder of this paper is organized as follows: Section II introduces related works on EEG-based seizure classification and transfer learning. Section III describes the seizure classification framework, baseline approaches, and our proposed APE-cosine. Section IV presents the experimental results. Finally, Section V draws conclusions.

II. Related Works

This section briefly introduces previous works on EEG-based seizure classification and transfer learning.

A. EEG-based Seizure Classification

EEG-based automatic epilepsy classification research dates back to the 1970s [17]. Shoeibi *et al.* [4] gave a comprehensive review on EEG-based automatic seizure classification using deep learning. There are multiple datasets available for automatic epileptic seizure classification, such as Hauz Khas [18], CHB-MIT [19], Bonn [20], Kaggle [21], and NICU [22]. Table I summarizes commonly used seizure classification datasets. NICU is used in this paper.

TABLE I
POPULAR PUBLICA SEIZURE EEG DATASETS.

Dataset	No. of Patients	No. of Seizures	Recording Type	Sampling Rate (Hz)	Time Length (h)
Hauz Khas [18]	10	–	Scalp EEG	200	–
CHB-MIT [19]	22	163	Scalp EEG	256	644
Bonn [20]	10	–	Intracranial EEG	256	708
Kaggle [21]	2 humans	48	Intracranial EEG	5000	21.3
	5 dogs	100	Intracranial EEG	400	658
NICU [22]	79	460	Scalp EEG	256	97.4

B. Transfer Learning

The assumption of conventional machine learning is that the training data and the test data belong to the same distribution. However, this assumption may not hold in real life, e.g., EEG data of different seizure patients demonstrate large individual differences. Transfer learning has been proposed to cope with this problem [10].

Wang *et al.* [23] achieved 92.77% accuracy on the CHI-MIT benchmark using deep transfer. Zhang *et al.* [24] used three deep transfer CNNs (VGG16, VGG19, and ResNet50) for automatic cross-subject seizure classification, achieving 97.75%, 98.26% and 96.17% accuracies, respectively.

Domain Adaptation, a special case of transfer learning, means that the source domains have labeled data, whereas the target domain does not. Semi-supervised domain adaptation adapts the source distributions to the target distribution, making use of partially labeled target samples [16]. There have been only a few studies in this direction [15], [16], [25]–[27], most proposed for image classification.

III. METHODOLOGY

This section introduces the APE algorithm [16], and our proposed APE-cosine, for EEG-based seizure classification.

A. APE

APE [16] consists of three main steps: Attraction, Perturbation, and Exploration. The target domain is divided into an aligned sub-distribution and an unaligned sub-distribution. Attraction aligns the unaligned target sub-distribution to the aligned one, through intra-domain discrepancy minimization. Perturbation perturbs the target sub-distributions to their intermediate regions, propagating the labels to the unaligned target sub-distribution. Exploration locally modulates the prototypes in a class-aware manner, complementary to attraction and perturbation.

Let $\mathcal{D}_s = \{(\mathbf{x}_i^s, {y_i}^s)\}_{i=1}^{m_s}$ be the source domain samples, and $\mathcal{D}_t = \{(\mathbf{x}_i^t, {y_i}^t)\}_{i=1}^{m_t}$ and $\mathcal{D}_u = \{\mathbf{x}_i^u\}_{i=1}^{m_u}$ be the labeled and unlabeled target domain samples, respectively.

APE uses four loss terms (cross-entropy loss, attraction loss, perturbation loss, and exploration loss), to align the intra-domain and inter-domain distributions:

$$\min_{p,\epsilon} \left(\mathcal{L}_{cls} + \alpha\mathcal{L}_a + \gamma\mathcal{L}_p + \beta\mathcal{L}_e\right). \tag{1}$$

The cross-entropy loss $\mathcal{L}_{cls}$ is used to train an embedding function $f_{\boldsymbol{\theta}}(\cdot)$ with parameters $\boldsymbol{\theta}$ and prototypes $\mathbf{p}_k$ $(k = 1, ..., K)$ to align the source domain samples and the labeled target domain samples:

$$\mathcal{L}_{cls} = \mathbb{E}_{(\mathbf{x},y)\in\mathcal{D}_s\cup\mathcal{D}_t}\left[-\log p(y|\mathbf{x},\mathbf{p})\right], \tag{2}$$

where

$$p(y|\mathbf{x},\mathbf{p}) = \frac{\exp(\mathbf{p}_y \cdot f_\theta(\mathbf{x})/T)}{\sum_{i=1}^{K}\exp(\mathbf{p}_y \cdot f_\theta(\mathbf{x})/T)}, \tag{3}$$

in which T is a temperature parameter adjusting the "smoothness" of the output.

The attraction loss $\mathcal{L}_a$ uses Maximum Mean Discrepancy (MMD) [28] to minimize the discrepancy between the labeled and unlabeled samples:

$$\mathcal{L}_a = d(\mathcal{D}_s \cup \mathcal{D}_t, \mathcal{D}_u). \tag{4}$$

The perturbation loss $\mathcal{L}_p$ reduces the model overfitting:

$$H_{\mathbf{p}}(\mathbf{x}) = -\sum_{i=1}^{K} p(y=i \mid \mathbf{x}) \log p(y=i \mid \mathbf{x},\mathbf{p}), \tag{5}$$

$$r_{\mathbf{x}} = \underset{\|r\|<\epsilon}{\operatorname{argmin}} \max_{\mathbf{p}} H_{\mathbf{p}}(\mathbf{x}+r), \tag{6}$$

$$\mathcal{L}_p = \mathbb{E}_{\mathbf{x}\in\mathcal{D}_u}\left[\sum_{i=1}^{K} D_{KL}\left[p(y=i \mid \mathbf{x},\mathbf{p}), p\left(y=i \mid \mathbf{x}+\mathbf{r}_{\mathbf{x}},\mathbf{p}\right)\right]\right] + \mathbb{E}_{\mathbf{z}\in\mathcal{D}_t}\left[\sum_{i=1}^{K} D_{KL}\left[p(y=i \mid \mathbf{z},\mathbf{p}), p\left(y=i \mid \mathbf{z}+\mathbf{r}_{\mathbf{z}},\mathbf{p}\right)\right]\right], \tag{7}$$

where $H_{\mathbf{p}}(\cdot)$ is an entropy function based on the similarity between a given feature and a prototype, $\mathbf{x}$ and $\mathbf{z}$ are samples, and y is the corresponding label.

The exploration loss $\mathcal{L}_e$ locally modulates the class prototype based on class perception for the attractor term and selectively aligns some unaligned sample features in the target domain:

$$\mathcal{L}_e = \mathbb{E}_{\mathcal{D}_u}\left[-\mathbf{1}_{M_\epsilon}(\mathbf{x}) \log p\left(y=\hat{y}_{\mathbf{x}} \mid \mathbf{x},\mathbf{p}\right)\right], \tag{8}$$

where

$$M_\epsilon = \left\{\mathbf{x}\in\mathcal{D}_u \mid H_{\mathbf{p}}(\mathbf{x}) < \epsilon\right\}, \tag{9}$$

$$\hat{y}_{\mathbf{x}} = \underset{i\in\{1,\ldots,K\}}{\operatorname{argmax}} p(y=i \mid \mathbf{x},\mathbf{p}), \tag{10}$$

M_ϵ is a set of unlabeled target samples with entropy smaller than a threshold ϵ, and $-\mathbf{1}_{M_\epsilon}(\mathbf{x})$ is an indicator function that filters out alignable samples from the given unlabeled target samples.

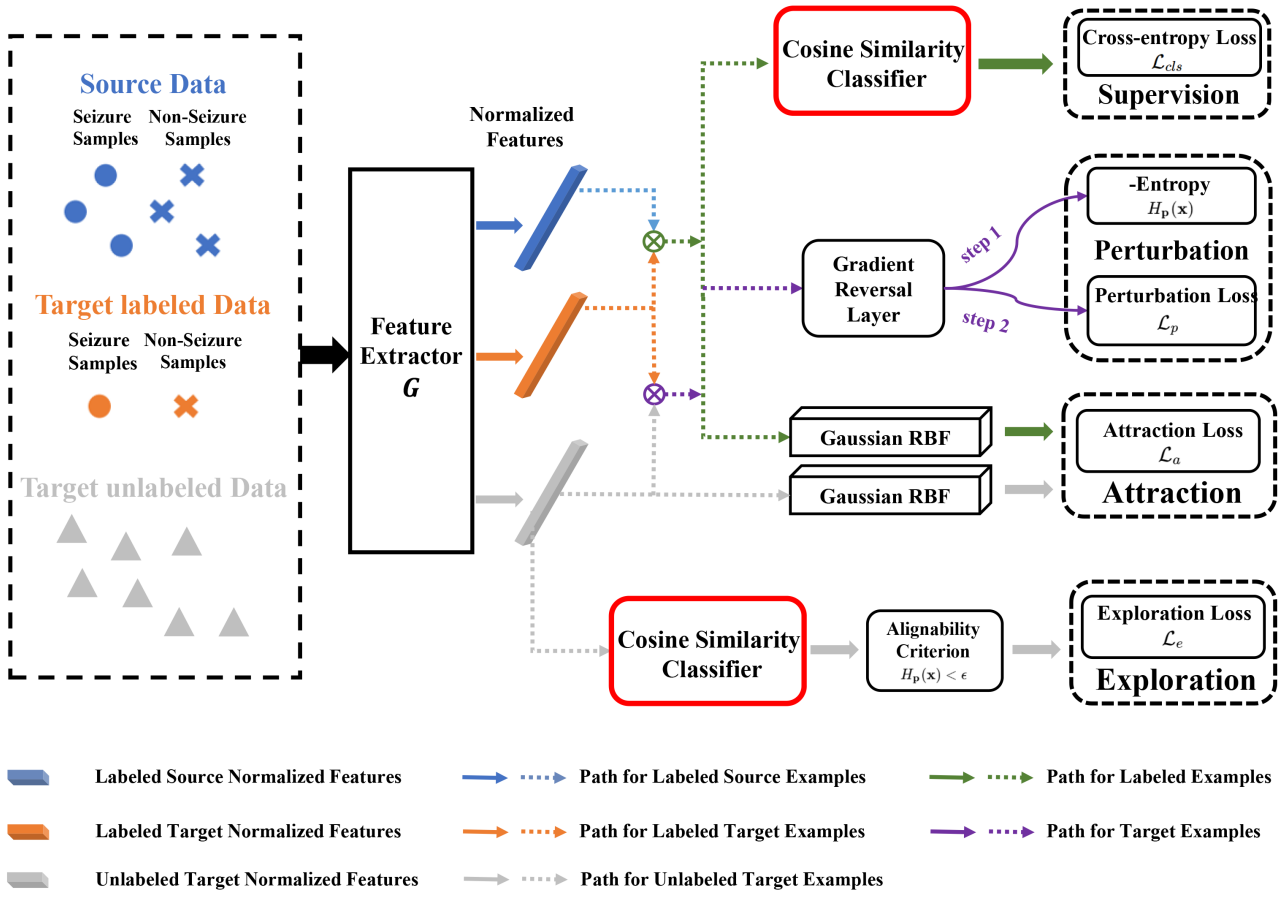

Fig. 1. The workflow of APE-cosine.

B. APE-cosine

APE computes the estimated label $\tilde{y}_i$ for input $\mathbf{x}_i$ using:

$$\tilde{y}_i = \sigma\left(\frac{1}{T}\frac{W^T f_\theta(\mathbf{x}_i)}{\|f_\theta(\mathbf{x}_i)\|}\right), \tag{11}$$

where W is the fully connected layer weight matrix, $f_{\boldsymbol{\theta}}(\mathbf{x}_i)$ the feature vector, and T is the temperature parameter.

To remove the extra temperature parameter T and further reduce intra-class variation, we propose APE-cosine, which computes the softmax of the cosine similarity between W and $f_{\boldsymbol{\theta}}(\mathbf{x}_i)$. We calculate the similarity scores $[s_{i,1}, s_{i,2}, ..., s_{i,c}]$ for all c classes, where

$$s_{i,j}\underset{j\in\{1,\ldots,c\}}{} = \frac{{W_j}^T f_\theta(\mathbf{x}_i)}{\|W_j\|\,\|f_\theta(\mathbf{x}_i)\|}, \tag{12}$$

in which W_j is the weight vector of class j. Prediction probability can be obtained by normalizing the similarity scores with a softmax function.

Fig. 1 illustrates the complete workflow of APE-cosine.

C. The Complete Cross-Subject Seizure Classification Framework

Fig. 2 shows our complete flowchart for cross-subject seizure classification, which includes preprocessing, feature extraction, APE-cosine classification, and evaluation.

First, raw EEG signals are transformed into fixed-length trials, after a series of pre-processing operations, including notch filtering, band-pass filtering, annotation, epoching, and data augmentation. Then, three CNN models (EEGNet, Deep ConvNet, and Shallow ConvNet) are used to extract deep features. The training data and test data come from different subjects. The test subject has a small number of labeled trials, and a large number of unlabeled trials. we use APE-cosine to train a model to label the unlabeled target domain trials.

IV. Experiments

This section presents experimental results to demonstrate the performance of APE-cosine.

A. Dataset

A publicly available dataset of annotated neonatal EEGs collected from human neonates admitted to the neonatal intensive care unit (NICU) at Helsinki University Hospital [22] was used. It includes multi-channel 256 Hz EEGs from 79 full-term neonates with a median recording time of 74 min (IQR: 64-96 min). Each neonate was independently labeled by three experts every second, 1 for seizure and 0 for non-seizure. On average, each expert labeled 460 seizures.

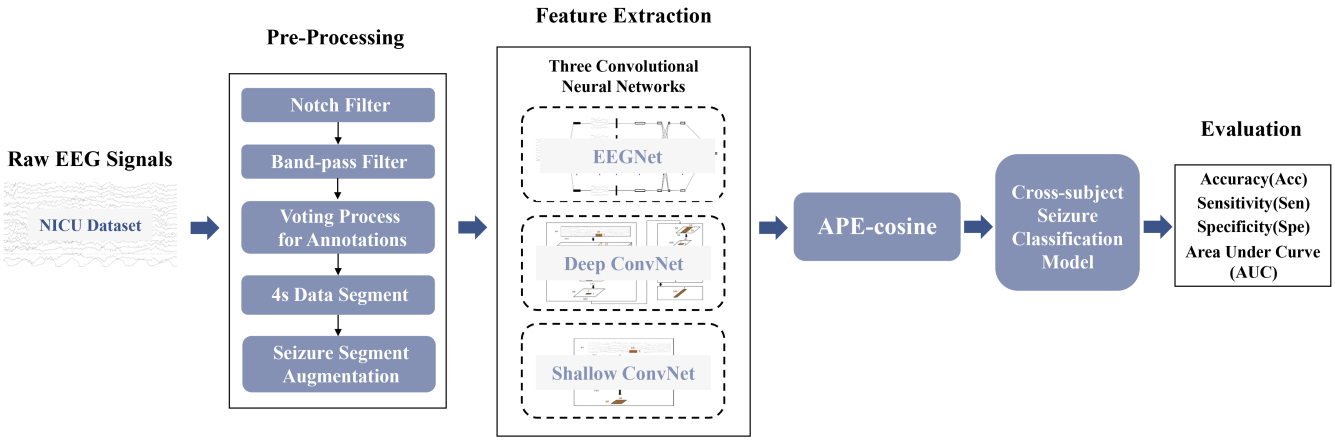

Fig. 2. EEG-based cross-subject seizure classification framework.

B. Preprocessing

A 50 Hz IIR notch filter was used to remove the power line interference, and a 0.5-50 Hz band-pass filter was employed to remove voltage drift and high frequency noise.

Each trial was labeled by three experts, and there were plenty of disagreements among them. We consider four different cases:

- '111', i.e., all three experts labeled the trial as seizure. The trial was recorded as noise-free segment P_1.
- '000', i.e., all three experts labeled the trial as non-seizure. The trial was recorded as noise-free segment P_0.
- '110', i.e., two of the three experts labeled the trial as seizure. This trial was recorded as noisy segment N_1.
- '100', i.e., only one expert labeled the trail as seizure. The trial was recorded as noisy segment N_0.

We then selected patients based on the following three criteria:

$$\frac{P_0}{\lambda N_1 + (1-\lambda)N_0} \geqslant w, \tag{13}$$

$$\frac{P_1}{\lambda N_0 + (1-\lambda)N_1} \geqslant w, \tag{14}$$

$$\frac{P_0}{P_1} \leqslant \theta, \tag{15}$$

(13) and (14) indicate the purity of non-seizure and seizure trials, respectively. Both have to be greater than or equal to a threshold $w = 10$ ($\lambda = 0.8$). (15) indicates the degree of class imbalance, which should be no larger than a threshold $\theta = 50$. Seven patients, as shown in Table II, were selected for subsequent algorithm evaluation, and voting of the three experts was used for final annotations [29].

We epoched the EEG time series using a 4s sliding window. Thus, each trial is a 18×1024 (18 channels, 1024 sampling points) matrix. In training, we increased the number of seizure trials using 25% sliding window overlap. No overlap was used for non-seizure trials in training, and all trials in validation and test.

C. Feature Extraction and Classification

Three networks, EEGNet, Deep ConvNet, and Shallow ConvNet, were used for feature extraction and classification.

TABLE II
CHARACTERISTICS OF THE SEVEN PATIENTS.

Patient Index	Recording Duration (min)	Ratio of Non-seizure to Seizure Duration
13	256.9	10.27
31	58.8	18.22
34	108.2	13.25
36	84.7	9.33
62	97.5	14.25
66	189.2	5.53
75	65.9	3.29
Avg	123.0	10.59

EEGNet [7] is a general and compact CNN specifically designed for general EEG recognition tasks. EEGNet uses a temporal convolution to learn frequency filters, then a depth-wise convolution to learn frequency-specific spatial filters, and finally separable convolutions to learn a temporal summary for each feature map individually and to optimally mix the feature maps.

Deep ConvNet [8] has achieved promising performances in multiple different BCI paradigms. It consists of five convolutional layers and a softmax classification layer, which can be used to extract temporal and spatial information from EEGs.

Shallow ConvNet [8] was specifically designed for oscillatory signal classification (by extracting features related to log band power). It consists of two convolutional layers (temporal and spatial), a squared nonlinearity [$f(x) = x^2$], a mean pooling layer, and a logarithmic nonlinearity [$f(x) = \log(x)$].

D. Patient-Specific Seizure Classification

For patient-specific seizure classification, we used the first 80% of the seizure and non-seizure segments as the training set, and the remaining 20% as the test set. 25% of the training set was further reserved for validation. Adam optimizer with initial learning rate 0.001 and batch-size 64 was used. The maximum number of training epochs was 50, and AUC was

used as an indicator to select and save the best model based on the validation set.

We repeated all three CNNs three times for each subject with the same experiment settings and report the average results in Table III.

TABLE III
PATIENT-SPECIFIC SEIZURE CLASSIFICATION AUCS (%).

	EEGNet	Deep ConvNet	Shallow ConvNet	Avg
w/o data augmentation	78.73	85.15	90.74	84.87
w/ data augmentation	91.58	92.61	92.48	92.22

Data augmentation of seizures was implemented by sliding windows with 25% overlap. Table III shows that on average it increased the AUC by 7.35%, and the AUC of every CNN model exceeded 90% after data augmentation. Thus, data augmentation was used in all subsequent experiments.

E. Semi-Supervised Cross-Patient Seizure Classification

Semi-supervised cross-patient seizure classification assumes a large amount of labeled EEGs from the source subjects, and 2-min labeled EEGs and a large amount of unlabeled EEGs from the target subject. For example, when Patient 13 is the target subject, he/she has 2 minutes of labeled EEGs and 254.9 minutes of unlabeled EEGs, and the remaining 6 patients were used as the sources subjects. Each patient was used as the target subject once, and the average results are reported.

Our proposed semi-supervised domain adaptation approach was compared with APE [16], MME [15], ENT [14], S+T [13], and Source-only. ENT trains a model using all labeled and unlabeled data, whose task classifier maximizes the entropy of the unlabeled data whereas the feature extractor minimizes the entropy. S+T trains a model on all labeled samples. Source-only uses data in the source domain for training and unlabeled data in the target domain for testing.

We adopted EEGNet, Deep ConvNet and Shallow ConvNet as the backbone network. Each mini-batch consisted of the same number of labeled source samples and labeled target samples, and twice the number of unlabeled target samples. Adam optimizer with initial learning rate 0.001 was used. We randomly combined 20% of the source domain data and 40% of the labeled target domain data as the validation set to select the best hyperparameters. The temperature parameter was set to 0.5. The batch size was selected from $\{32, 64, 128\}$. For MME, λ was selected from $\{0.01, 0.05, 0.1, 0.2\}$. For APE and APE-cosine, the threshold ϵ was selected from $\{0.1, 0.3, 0.5, 0.7\}$, and the loss term weight parameters (α, β, γ) was selected from $\{(10, 1, 10), (1, 0.1, 1)\}$.

The results are shown in Table IV and Fig. 3. Source-only had the worst performance, indicating the necessity of using some target domain data for patient-specific calibration, due to individual differences. S+T had the worst performance among all algorithms that used target domain data, as it only used the labeled target domain data, whereas others also used the unlabeled target domain data. APE achieved the best performance among existing approaches, but our proposed APE-cosine further improved it by about 2%.

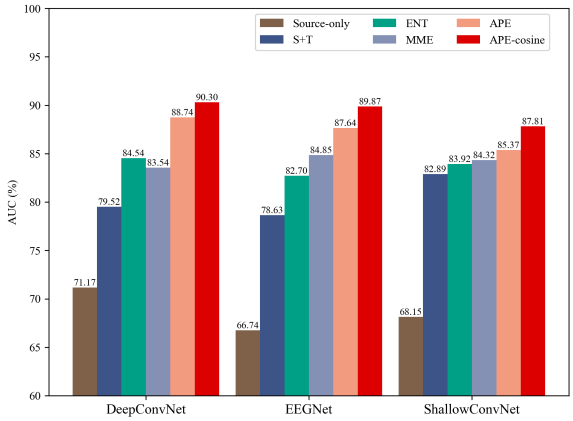

Fig. 3. Semi-supervised cross-subject seizure classification results, averaged over the seven subjects.

V. CONCLUSIONS

EEG-based automatic seizure classification plays a crucial role in the diagnosis and treatment of epilepsy. However, the machine learning algorithms have to cope with the challenges of inadequate labeled samples, class imbalance, and individual differences. This paper has proposed APE-cosine, a semi-supervised domain adaptation approach for cross-patient seizure classification. Experimental results using three different CNN models demonstrated its effectiveness.

ACKNOWLEDGEMENT

This research was supported by the IEEE Computational Intelligence Society Graduate Student Research Grant.

REFERENCES

[1] S. Noachtar and J. Rémi, "The role of EEG in epilepsy: a critical review," *Epilepsy & Behavior*, vol. 15, no. 1, pp. 22–33, 2009.

[2] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowledge-Based Systems*, vol. 45, no. 6, pp. 147–165, 2013.

[3] R. Peng, J. Jiang, G. Kuang, H. Du, D. Wu, and J. Shao, "EEG-based automatic epilepsy detection: Review and outlook," *Acta Automatica Sinica*, vol. 48, no. 2, pp. 335–350, 2022, in Chinese.

[4] A. Shoeibi, M. Khodatars, N. Ghassemi, M. Jafari, P. Moridian, R. Alizadehsani, M. Panahiazar, F. Khozeimeh, A. Zare, H. Hosseini-Nejad *et al.*, "Epileptic seizures detection using deep learning techniques: a review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, pp. 5780–5813, 2021.

[5] M. T. Avcu, Z. Zhang, and D. W. S. Chan, "Seizure detection using least EEG channels by deep convolutional neural network," in *Proc. IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, Brighton, UK, May 2019, pp. 1120–1124.

[6] I. C. Covert, B. Krishnan, I. Najm, J. Zhan, M. Shore, J. Hixson, and M. J. Po, "Temporal graph convolutional networks for automatic seizure detection," in *Proc. of Machine Learning for Healthcare Conf.*, Ann Arbor, Michigan, Oct. 2019, pp. 160–180.

[7] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

TABLE IV
CROSS-PATIENT SEIZURE CLASSIFICATION RESULTS. THE BEST PERFORMANCE IS MARKED IN BOLD, AND THE SECOND BEST BY AN UNDERLINE.

Backbone model	Approach	Patient							Avg
		13	31	34	36	62	66	75	
EEGNet	Source-only	0.77	0.4992	0.9691	0.7891	0.6531	0.4803	0.5109	0.6674
	S+T	0.8476	0.9641	0.9854	0.8378	0.8244	0.515	0.53	0.7863
	ENT	0.8403	0.9478	0.9871	0.9312	0.8633	0.5348	0.6846	0.827
	MME	0.8384	0.9701	0.9892	0.9176	0.8633	0.6058	0.7554	0.8485
	APE	0.8473	0.9642	0.9881	0.9331	0.9627	0.7608	0.6785	<u>0.8764</u>
	APE-cosine	0.8703	0.9612	0.9909	0.9421	0.9595	0.8282	0.7386	**0.8987**
Deep ConvNet	Source-only	0.8098	0.6787	0.9545	0.8256	0.6417	0.5362	0.5357	0.7117
	S+T	0.8232	0.879	0.9666	0.8495	0.7338	0.6808	0.6338	0.7952
	ENT	0.8327	0.9431	0.9655	0.9096	0.8809	0.6964	0.6894	0.8454
	MME	0.8373	0.9182	0.9805	0.8585	0.8587	0.7029	0.6915	0.8354
	APE	0.8298	0.9428	0.9791	0.9132	0.9581	0.8683	0.7202	<u>0.8874</u>
	APE-cosine	0.8816	0.9595	0.9705	0.9306	0.9607	0.8811	0.7371	**0.903**
Shallow ConvNet	Source-only	0.7238	0.6401	0.8968	0.8447	0.6555	0.507	0.5024	0.6815
	S+T	0.8297	0.9541	0.9876	0.8587	0.8602	0.7358	0.5765	0.8289
	ENT	0.8406	0.9478	0.9719	0.923	0.7827	0.7778	0.6306	0.8392
	MME	0.8452	0.9591	0.9704	0.9059	0.792	0.7668	0.6628	0.8432
	APE	0.8604	0.9734	0.9713	0.9242	0.8938	0.6765	0.6763	<u>0.8537</u>
	APE-cosine	0.8534	0.9461	0.9705	0.9443	0.9552	0.7775	0.6995	**0.8781**

[8] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[9] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, pp. 1–54, 2019.

[10] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 1–40, 2016.

[11] H. Zhou, A. Oliver, J. Wu, and Y. Zheng, "When semi-supervised learning meets transfer learning: Training strategies, models and datasets," *arXiv preprint arXiv:1812.05313*, 2018.

[12] Y. Jiang, D. Wu, Z. Deng, P. Qian, J. Wang, G. Wang, F. Chung, K.-S. Choi, and S. Wang, "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.

[13] W. Chen, Y. Liu, Z. Kira, Y. F. Wang, and J. Huang, "A closer look at few-shot classification," *arXiv preprint arXiv:1904.04232*, 2019.

[14] Y. Grandvalet and Y. Bengio, "Semi-supervised Learning by Entropy Minimization," in *Advances in Neural Information Processing Systems*, British Columbia,Canada, Dec. 2004, pp. 529–536.

[15] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko, "Semi-supervised domain adaptation via minimax entropy," in *Proc. of the IEEE/CVF Int'l Conf. on Computer Vision*, Seoul, Korea, Oct. 2019, pp. 8050–8058.

[16] T. Kim and C. Kim, "Attract, perturb, and explore: Learning a feature alignment network for semi-supervised domain adaptation," in *European Conf. on Computer Vision*, Glasgow, UK, Aug. 2020, pp. 591–607.

[17] von Albert Hh, "Automatic analysis of long-term EEG in epilepsy. First experiences with a new part-automatic analytic system," *EEG-EMG Zeitschrift fur Elektroenzephalographie, Elektromyographie und Verwandte Gebiete*, vol. 8, no. 2, pp. 105–111, 1977.

[18] E. B. Assi, D. K. Nguyen, S. Rihana, and M. Sawan, "Towards accurate prediction of epileptic seizures: A review," *Biomedical Signal Processing and Control*, vol. 34, no. 4, pp. 144–157, 2017.

[19] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," M.Sc. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2009.

[20] 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," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.

[21] J. J. Howbert, E. E. Patterson, S. M. Stead, B. H. Brinkmann, V. Vasoli, D. P. Crepeau, C. H. Vite, B. K. Sturges, V. Ruedebusch, J. Mavoori, K. W. Leyde, W. D. Sheffield, B. Litt, and G. A. Worrell, "Forecasting seizures in dogs with naturally occurring epilepsy," *PloS One*, vol. 9, no. 1, p. e81920, 2014.

[22] N. J. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo, "A dataset of neonatal EEG recordings with seizure annotations," *Scientific Data*, vol. 6, no. 1, pp. 1–8, 2019.

[23] Y. Wang, J. Cao, J. Wang, D. Hu, and M. Deng, "Epileptic signal classification with deep transfer learning feature on mean amplitude spectrum," in *Annual Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, Berlin, Germany, Oct. 2019, pp. 2392–2395.

[24] B. Zhang, W. Wang, Y. Xiao, S. Xiao, S. Chen, S. Chen, G. Xu, and W. Che, "Cross-Subject seizure detection in EEGs using deep transfer learning," *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 1–8, 2020.

[25] H. Daumé, A. Kumar, and A. Saha, "Co-regularization based semi-supervised domain adaptation," in *Proc. of the 23rd Int'l Conf. on Neural Information Processing Systems*, Red Hook, NY, Dec. 2010, p. 478–486.

[26] H. Daumé III, A. Kumar, and A. Saha, "Frustratingly easy semi-supervised domain adaptation," in *Proc. of the 2010 Workshop on Domain Adaptation for Natural Language Proc.*, Uppsala, Sweden, Jul. 2010, pp. 53–59.

[27] J. Donahue, J. Hoffman, E. Rodner, K. Saenko, and T. Darrell, "Semi-supervised domain adaptation with instance constraints," in *Proc. of the IEEE conf. on Computer Vision and Pattern Recognition*, Portland, OR, Jun. 2013, pp. 668–675.

[28] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 723–773, 2012.

[29] A. Borovac, S. Gudmundsson, G. Thorvardsson, and T. P. Runarsson, "Influence of human-expert labels on a neonatal seizure detector based on a convolutional neural network," in *Proc. of Conf. on Neural Information Processing Systems*, Sydney, Australia, Dec. 2021, pp. 1–5.