EEG-Based Seizure Prediction Using Transfer Learning

Ziwei Wang, Dongrui Wu School of Artificial Intelligence and Automation
Huazhong University of Science and Technology, Wuhan, China
Email: zwwang0606@hust.edu.cn, drwu@hust.edu.cn

***Abstract*—As one of the most common brain diseases, epilepsy affects nearly 1% of all human beings. Patients' central nervous systems are destroyed chronically, hardly differentiable from a normal state until epileptic seizures happen, and hence difficult to prevent. About 65 million people worldwide have epilepsy, more than the combined number of people diagnosed with Parkinson's disease, Alzheimer's disease and multiple sclerosis. It is practical and feasible to establish a reliable seizure prediction system. When the electroencephalogram signal is predicted to be pre-ictal, the system sends an alarm to remind patients, family members or doctors to take drug treatments, electrical stimulation or physical measures to prevent seizures. We introduce three transfer learning algorithms, including TCA, JDA and EasyTL, together with two ensemble learning algorithms, RF and GDBT, to predict seizures, constructing a seizure prediction system framework based on transfer learning algorithms. Three experiments are conducted consecutively on a public dataset: 1) single patient seizure prediction experiment reaches prediction of 95.18%; 2) cross-patient seizure prediction without transfer learning results in prediction accuracy of 32.71%. Compared with Experiment I, Experiment II confirmed that the individual differences of EEG among patients would lead to a significant decline in prediction performance; 3) cross-patient seizure prediction with transfer learning. After using transfer learning algorithms, the prediction accuracy was improved from 32.71% to 57.29%, demonstrating the effectiveness of our proposed approach and confirming the strength of transfer learning methods cross patient seizure prediction task.**

***Index Terms*—Seizure Prediction, Transfer Learning, Electroencephalogram, Interictal, Preictal, Ensemble Learning**

I. Introduction

Electroencephalogram (EEG) is a kind of time series signal collected by monitoring specific brain electromagnetic fields, which can be utilized to directly reflect brain activities [1]. Scalp electroencephalogram (sEEG) is the most commonly used and most economical non-invasive brain wave detection method [2]. Through placing electrodes on specific positions on the scalp, the microvolt level signals generated by synchronized neuronal activity in the brain are collected. Currently, in clinical practice, visual inspection and manual EEG annotation are the gold standards for epilepsy detection. However, the large amount of tedious work brings a heavy burden to professionals, and the accuracy of detection excessively depends on the subjective judgment of the examiner. Therefore, it is important to combine the expertise of computer scientist and medical scientist to build a predictive model for automatic epilepsy detection and seizure prevention. When the predictive model indicates signs of seizures, the patient only needs to take key medications for prevention. Obviously, these models require highly reliable algorithms to function.

With the help of automatic predictive model, it is practical and feasible to establish a reliable seizure prediction system. When the EEG signal is predicted to be pre-ictal, the system sends an alarm to remind patients, family members or doctors to take drug treatment, electrical stimulation or physical measures. This system is able to ensure the safety of patients, reduce the rate of accidental sudden unexpected death in epilepsy patients (SUDEP) [3], and allow patients to take drugs only when needed, avoiding drug dependence. In addition, it can prevent patients from engaging in high-risk activities (such as driving and high-altitude work) during seizures, and minimize the damage caused by seizures to themselves or others, so as to improve the quality of life of patients with epilepsy. The significance of seizure prediction is as follows: 1) reduce the workload of doctors and assist in epilepsy treatment; 2) reduce the pain caused by epilepsy and reduce the accidental mortality and drug dependence of patients with epilepsy; 3) it is beneficial fir patients' health and quality of life.

However, there are still some problems with currently available methods in the literature on epileptic seizure prediction, which have not been well resolved.

1) *There are few studies on the recognition of pre-ictal state*. A large number of studies only focus on the recognition and detection of ictal state or inter-ictal state [1], but research rarely on the recognition of pre-ictal state. If accurate identification of the pre-ictal state can be achieved, doctors and patients could take certain preventive measures before epileptic seizures and reduce the harm to patients, which is more realistic than traditional seizure detection [4].
2) *Good results cannot be obtained when performing cross-patient testing*. Traditional methods on epilepsy prediction focus on the within-patient scenario, that its train set and test set generally contain the EEG signals of the same patient. When the model performs cross-patient testing, i.e., the EEG signals in the test set are collected from a new patient that the model has never seen in the train set, the prediction accuracy (ACC) of the model will drop sharply. Since epilepsy is a very unstable phenomenon, EEG signals during epileptic seizures is very complex and often shows different distributional characteristics in different patients. Even for the same patient, EEG signals from different trials would often

vary significantly. Therefore, transfer learning methods must be adopted to handle cross-patient seizure prediction problem.

Because of the data scarcity issue in the field of seizure prediction, data from other patients must be utilized for the model to learn additional information. However, without specific transfer learning measures, such data would often have negative influence on the model because of high discrepancy of EEG signals between patients. Our research motivation is to reuse existing patients' data, decrease the discrepancy and realize cross-patient seizure prediction. In order to reduce differences and improve prediction accuracy, we introduce three transfer learning methods, including TCA [5], JDA [6] and EasyTL [7], together with two ensemble learning algorithms, RF [8] and GDBT [9], to predict seizures. We propose a novel method to improve the effectiveness of cross-patient seizure prediction and realize the reuse of different patients' data.

The remainder of this paper is organized as follows: Section II introduces related work on seizure prediction and transfer learning methods. Section III describes the seizure prediction system framework and the algorithms used. Section IV presents the experiments and evaluates the performance of seizure prediction before and after introducing transfer learning algorithms. Finally, Section V draws conclusions.

II. Related Work

In this section, we briefly introduce previous works on EEG-based seizure prediction and transfer learning.

A. Seizure Prediction

A great influential review published in 2007 stated that there is not enough evidence to prove that seizures can be predicted [4]. But since then, some progress has been made in the field. A clinical trial in 2013 showed that it is possible to implant devices in human patients to predict seizures in real time [10]. A review article on epilepsy prediction in 2018 also proposed a framework for online real-time prediction of seizures and summarized the progress in this field, including EEG databases, seizure prediction competitions, the prospective trial mentioned and advances in our understanding of the mechanisms of seizures [11].

There are many seizure databases and associated studies, international organization has an unprecedented effort on building publicly available databases. One is the EPILEPSIAE database that included non-continuous data sets [12]. The other is the IEEG.org database [13], [14], which has provided the basic framework to process data on GitHub. There are also other database like Bonn dataset [15] and CHB-MIT dataset [16] to help predict seizures.

According to the different location of EEG signal collection, datasets can be divided into scalp electroencephalogram data (sEEG) and intracranial electroencephalogram data (iEEG). In many cases, non overlapping windows are more suitable for predicting seizures, and researchers use them to predict seizures. The datasets commonly used in seizure prediction in recent years are listed in Table I.

In 2014 and 2016, large international competitions about seizure prediction were held on standard database and competitors were asked to use advanced algorithms to solve prediction problems. In 2014, the competition named American Epilepsy Society Seizure Prediction Challenge [17], [18], the competition database involved intracranial EEG data from human and dogs, there are a total of 942 seizures recorded over 500 days. Intracranial EEG was recorded from dogs with naturally occurring epilepsy using an ambulatory monitoring system. All EEG signals were segmented into 10 min segments of pre-ictal data and inter-ictal data used as train set for their algorithm, they also provided unlabeled 10 min segments for testing. In 2016, contest named Melbourne University AES/MathWorks/NIH Seizure Prediction [10] also used iEEG data, which collected from three patients about 1100 seizures records over 1300 days. The patients' seizures were hard to predict with existing algorithms.

It is worth mentioning that the two competitions both used AUC metric as the evaluation of seizure prediction algorithms, in 2014 the AUC value of the first place is 0.84 (nature 2018-39), and had strong generalization ability. When reused in a larger dataset, its algorithm performance decreased only 8%. In 2016 contest, the top AUC value is 0.81.

TABLE I
Several common international seizure EEG datasets

Name	Patient Num	Seizure Num	Data Type	Sample Rate(Hz)	Time Length(h)
Freiburg [19]	21	87	iEEG	256	708
CHB-MIT [16]	22	163	sEEG	256	644
Bonn [15]	10	100	iEEG	256	708
Kaggle [17]	2(human)	48	iEEG	5000	21.3
	5(dog)	100	iEEG	400	658
NICU [20]	79	460	sEEG	256	97.4
Barcelona [21]	5	3750	iEEG	512	83

Researchers have proposed different machine learning methods to predict seizures in recent years. Generally they can be divided into two categories, one is the traditional process, including preprocessing, feature extraction, classification and postprocessing four steps in total, and the other use convolutional neural network (CNN) to predict.

1) *Preprocessing*. Preprocessing of EEG is often conducted in various ways to remove noise and select channels in EEG signals [2]. Butterworth filter, wavelet transform, and Fourier transform are often used to get a better signal to noise ratio (SNR) when preprocessing. Common spatial filtering (CSP) can also reduce the SNR by selecting channels to decrease the computational cost [22].
2) *Feature Extraction*. Univariate and bivariate features, linear and non-linear features are extracted to classify pre-ictal and inter-ictal signals. They can be divided into four measures [4], including univariate linear measures, univariate non-linear measures, bivariate linear measures, and bivariate non-linear measures. Statistical features in time domain include standard deviation, mean, skewness, median, quartile, and so on. Spectral

features in frequency domain include variational coefficients and spectral skewness. For different types of signals, the performance of extraction features is also different. Spectral features have better performance in scalp EEG data, but the statistical features perform well in both scalp EEG and iEEG data. When the train data is sufficient, convolutional neural network is a good feature extraction method.

3) *Classification*. Support vector machine is a widely used EEG classification method. There are also some other classifiers can be used, such as k-nearest neighbor, random forest and ensemble learning classifier.
4) *Postprocessing*. Researchers have proposed many methods for seizure prediction but only a few have done statistical validation [2]. The postprocessing methods include k-fold cross validation, and moving average filter [23].

With the development of deep learning, CNN is more commonly used and has obtained better sensitivity in scalp EEG and intracranial EEG datasets. In a scalp EEG dataset, Truong et al. [24] and Hussein et al. [25] have applied CNN and obtained 81.2% and 93% sensitivity, respectively. Khan et al. [3] have applied CNN to scalp EEG dataset and classified it with 87.8% sensitivity and 85.8% specificity. As for iEEG datasets, there are also some researchers who used CNN. Yu et al. [26] used PCA and CNN to extract features in iEEG signals and got 87.7% sensitivity. Acharya et al. [27] applied CNN to extract features and classify which achieved 73.9% sensitivity. However, due to the differences among patients, it is worth noticing that although CNN has achieved good prediction results in single patient, it cannot obtain good results in cross-patient prediction and the prediction performance will be poor. It is therefore necessary to introduce transfer learning for cross-patient prediction.

B. Transfer Learning

Transfer learning uses the knowledge learned in the existing environment and applies it in the new environment to complete learning tasks, so it can solve the inconsistent distribution of train set and test set [5]. Researchers proposed transfer learning algorithms to use rich labeled data in the source domain to build a reliable classifier for the target domain which has a few or even no labeled data.

Due to the individual differences between patients and complexity of EEG signals, it is difficult to establish a general model to predict seizures for many specific patients. This leads to long-term data collection and processing for each patient, which is time-consuming and labor-consuming. By introducing transfer learning algorithms, the prediction model based on one patient who has sufficient data and labels can be reused in the prediction of another patient with less data and classification labels. Therefore, the study of seizure prediction based on transfer learning algorithms is of great significance to improve the health and living standards of epileptic patients and assist patients in treatment.

Some researchers have applied transfer learning to the prediction of seizures. Dhulekar et al. [28] present a novel approach to predicting EEG seizures by cross-learning public knowledge and using transfer learning in patients' records, they also designed a novel transformation to improve the efficiency of transfer learning. Raghu et al. [29] used transfer learning to pre-train the network and using SVM to classify EEG signals. Daoud et al. [30] introduced a semi-supervised approach based on transfer learning to improve their optimization problem which aims to select the most discriminative features and improve the accuracy of classification and reduce prediction time.

III. Methodology

A. Data Preprocessing

In the Kaggle2014 dataset, intracranial EEG data were collected from human and dogs. There are totally 942 seizures recorded over 500 days [17], [18]. Intracranial EEG was recorded from dogs with naturally occurring epilepsy using an ambulatory monitoring system. All EEG signals were segmented into 10 min segments of pre-ictal data and inter-ictal data used as the train set of their algorithm, also with unlabeled 10 min segments for testing. The sampling rate of human EEG signal is 5000Hz, of 15 channels. The sampling rate of dog EEG signal is about 400Hz of 16 channels. The train data segments are numbered in order, and the test data are numbered in random order to ensure the generalization performance.

The time range of train and test data set is from one hour and five minutes before the seizure to five minutes before the seizure. On one hand, this division of data can ensure enough time to alert and inform patients to take measures for prevention and treatment as soon as possible. On the other hand, if missed information annotated by epilepsy diagnosis before seizure, it will not affect the prediction results.

In this work, discrete wavelet transform (DWT) [31] and low pass filtering are used to remove high-frequency noise, and the sub-bands obtained by wavelet decomposition are also used for the following feature extraction.

B. Feature Extraction

The input signal frequency band is 0-128Hz. After 7-level wavelet decomposition, EEG signals can be divided into seven different frequency bands, 7 sub-bands are selected for feature extraction respectively. The extracted features are statistical features, crossing features and entropy features. Details are described below.

Calculate n5, n25, n75, n95, mean, median, standard deviation, variance and root mean square, are calculated, totaling 9 statistical features. Zero crossing rate and mean crossing rate are the times that the signal passes through the baseline $y = 0$ and the average level $y = u$ per unit time respectively. Entropy features are characteristic measure of signal complexity. Approximate Entropy (ApEn) is a measure of data regularity. Irregular time series correspond to higher non negative ApEn values, while regular and predictable time series signals correspond to lower ApEn values. Sample entropy (SampEn) is a method to measure self-similarity. Compared with ApEn, SampEn is an improved parameter, and seizures will lead to the reduction of these two entropy parameters [32].

C. Classification Algorithms

An important branch of machine learning is ensemble learning. Ensemble learning is to build many base learners, and then combine these learners to increase the generalization performance of the model and get better performance in machine learning tasks.

1) RF: Breiman proposed the idea of Random Forest (RF), which is a kind of bagging ensemble learning [8]. It takes the decision tree as the base learner, and constructs the bagging ensemble learner. When selecting the optimal partition attribute of decision tree, the random attribute selection is carried out. The traditional decision tree is to select an optimal attribute in the set of attributes of the current node and do continuous division. In RF, the strategy of selecting the optimal partition attribute for each internal node of the base decision tree is different from the traditional decision tree. In the set with d attributes in the current node, k attributes are randomly selected to generate a subset of the original set, and then select an optimal attribute in the new subset as the basis for branching.

2) GBDT: Gradient Boosting Decision Tree (GBDT) classifier uses the boosting and gradient boosting algorithm which aims to promote a weak learner to a strong learner. AdaBoost [33] is one of the boosting algorithms. Shortly after AdaBoost was proposed, Friedman, a machine learning and statistician, explained from a statistical perspective and explained that AdaBoost was actually based on an additive model to optimize the exponential loss function similar to the Newton iterative method. Inspired by this, Friedman [9] proposed the idea of gradient boosting. Gradient boosting decision tree is to use the decision tree as a weak learner to fit the gradient in the process of gradient boosting, then fuse it into the whole gradient boosting process.

D. Transfer Learning Algorithms

In this work, we apply transfer learning methods to achieve cross-patient seizure prediction, which meets the treatment requirements in more realistic scenarios. Below, we introduce three methods adopted in this work.

1) Transfer Component Analysis (TCA): TCA [5] proposed to learn a transformation that decreases the marginal distributions discrepancy between the source domain and the target domain. The main idea of TCA is to use a dimension reduction method to re-extract the information in the sample, so that the new information not only reduces the distance between two domains, but also maintains the integrity of the main information.

TCA method uses Maximum Mean Discrepancy (MMD) as the measurement of domain discrepancy [34]. So, below we first introduces MMD. Borgwardt et al. [35] proposed the MMD as a relative standard of comparing distributions based on Reproducing Kernel Hilbert Space (RKHS). Let $X = \{x_1, ..., x_{n_1}\}$ and $Y = \{y_1, ..., y_{n_2}\}$ be two varible sets with different distributions P and Q. Then the distance between P and Q are defined by MMD as follows:

$$\mathrm{MMD}(X, Y) = \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} \phi(x_i) - \frac{1}{n_2} \sum_{i=1}^{n_2} \phi(y_i) \right\|_{\mathcal{H}}^2, \tag{1}$$

where $\mathcal{H}$ is a RKHS. By minimizing this quantity, a nonlinear mapping function ϕ from sample space χ to $\mathcal{H}$ can be found marked as $\chi \mapsto \mathcal{H}$.

Let $X \in \mathbb{R}^{n \times d}$ denote the input space and $Y \in \mathbb{R}$ denote the output space, where d is the dimension of the input features and n is the number of samples. We consider the unsupervised DA scenario that we have access to labeled samples of the source domain $D_s = \{X_s, Y_s\}$, where $\{X_s, Y_s\} = \{(\mathbf{x_i^s}, y_i^s)\}_{i=1}^{n_s}$, and unlabeled samples of the target domain $D_t = \{X_t\}$, where $\{X_t\} = \{\mathbf{x_i^t}\}_{i=1}^{n_t}$. Typically, there exists difference between distributions of source and target domains, i.e. $P(X_s, Y_s) \neq P(X_t, Y_t)$, $P(X_s) \neq P(X_t)$. Let $X_s \bigcup X_t = X$ be the input transformed from source domain, target domain and combined domain, respectively. Our task is to predict the unknown outputs $\{y_i^t\}$.

For transfer learning problems,MMD is used as a measurement, and the distance between source domain D_s and D_t can be calculated as follows:

$$\mathrm{Dist}(X_s, X_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(x_{s_i}) - \frac{1}{n_t} \sum_{i=1}^{n_t} \phi(x_{t_i}) \right\|_{\mathcal{H}}^2. \tag{2}$$

It is worth noticing that TCA only considers the marginal distributions discrepancy, the setting is that $P(X_s) \neq Q(X_t)$, but $P(Y_S \mid \phi(X_S)) = P(Y_T \mid \phi(X_T))$, that is to say they only consider the marginal distributions discrepancy $P(X_s) \neq Q(X_t)$ between the source and target domain.

Pan et al. [5] proposed to transform the nonlinear transformation ϕ to a kernel learning problem. By using the kernel trick, such as $k(x_i, x_j) = \phi(x_i)^{'} \phi(x_j)$, the distance can be written as follows:

$$\mathrm{Dist}(X_s, X_t) = \mathrm{tr}(KL), \tag{3}$$

where

$$K = \left[\begin{array}{cc} K_{s,s} & K_{s,t} \\ K_{t,s} & K_{t,t}. \end{array} \right] \tag{4}$$

Then, by virtue of the parametric kernel map for unseen patterns, the empirical means of source domain and target domain distance can be rewritten as follows:

$$\mathrm{Dist}(X_s, X_t) = \mathrm{tr}\left(W^\top KLKW\right) \tag{5}$$

Finally, the learning problem for transfer learning using kernel trick can be simplified to the following form:

$$\begin{array}{ll} \min_W & \mathrm{tr}\left(W^\top W\right) + \mu \, \mathrm{tr}\left(W^\top KLKW\right) \\ \text{s.t.} & W^\top KHKW = I \end{array} \tag{6}$$

2) Joint Distribution Adaptation (JDA): JDA [6] is proposed to minimize the joint distribution discrepancy of input data and the label between source and target domains. That is, conditional distribution adaptation is introduced based on TCA to adapt both marginal and conditional distributions discrepancy, and construct new feature representation which is effective for reducing distribution difference. In most transfer learning tasks, the labels of target domain are unknown. JDA use the pseudo label strategy, by virtue of the classifier trained in the source domain to help with the classification tasks in the target domain. Then gradually repeat iteration to improve the accuracy of the pseudo labels. Given marked auxiliary domain,the source domain $D_s = \{(x_1, y_1), \ldots, (x_{n_s}, y_n)\}$ and unlabeled target domain $D_t = \{(x_{n+1}, \ldots, x_{n+m})\}$. It is necessary to meet these conditions, firstly the feature space and the label space of two domains should be same, $F_s = F_t, Y_s = Y_t$. Secondly, the marginal distribution and conditional distribution are both different: $P(X_s) \neq Q(X_t)$, but $P(Y_S \mid \phi(X_S)) \neq P(Y_T \mid \phi(X_T))$. JDA method study a feature representation T and classification model f to minimize both marginal and conditional distributions discrepancy at the same time. JDA modifies MMD to measure the distance between the class-conditional distributions, and the distance between source domain D_s and D_t can be calculated as follows:

$$\operatorname{Dist}(X_s, X_t) = \left\| \frac{1}{n_s^{(c)}} \sum_{i=1}^{n_s^{(c)}} \phi(x_{s_i}) - \frac{1}{n_t^{(c)}} \sum_{i=1}^{n_t^{(c)}} \phi(x_{t_i}) \right\|_{\mathcal{H}}^2 . \tag{7}$$

The kernel method is also introduced to solve the distance minimization problem:

$$\min \frac{\sum_{c=0}^{C} \operatorname{tr}\left(A^{\mathrm{T}} X M_c X^{\mathrm{T}} A\right) + \lambda \|A\|_F^2}{A^{\mathrm{T}} X H X^{\mathrm{T}} A}. \tag{8}$$

E. Seizure Prediction System Framework

Traditional prediction model for seizure prediction includes four parts: data preprocessing, feature extration, seizure state classification and postprocessing. Based on the traditional model, this work proposed EEG-Based transfer learning seizure prediction system framework which is shown in Fig. 1. In our framework, the architecture consists of five parts, preprocessing, feature extraction, feature transformation, classification, and alarm generation. Firstly, The raw signal is transformed into a relatively pure signal by removing high-frequency noise through low-pass filtering and wavelet transform. Then, eleven different features are extracted from the seven sub-bands obtained by wavelet decomposition, including time domain, frequency domain and nonlinear features. The train set and test set come from different patients. After feature extraction, feature vectors are not directly input the classifier. Instead, features are aligned from source domain and target domain through transfer learning methods including TCA, JDA, etc., and obtain the aligned two new feature vectors $X_s^{'}$ and $X_t^{'}$, the new feature vectors are then input into GBDT and RF classifiers. Transfer learning methods reduce the difference by aligning the feature space of the source domain and the target domain. In this way, the model trained in the source domain can be reused in the target domain to realize the knowledge transformation.

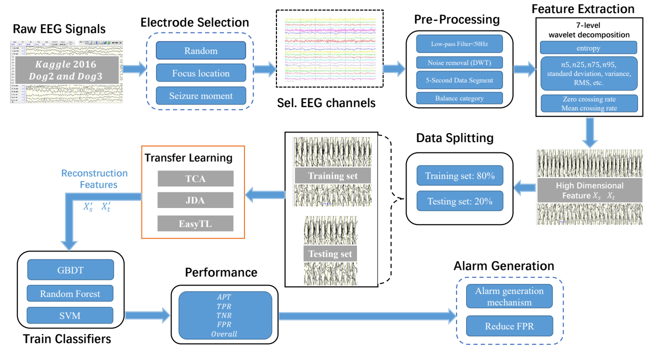

Fig. 1. EEG-based transfer learning seizure prediction system framework

IV. Experiments

This section presents three experiments and their results. The main contents are as follows: experiment settings, experimental results, performance verification, and model effectiveness.

A. Experiment settings

In all experiments, we choose two patients as the data for train and test named Dog_2 and Dog_3 from the original EEG records. In Dog_2 records, the inter-ictal EEG signals have 500 sets, but the pre-ictal EEG signals only have 42 sets. In Dog_3 records, there are 1440 inter-ictal sets and 72 pre-ictal sets. In order to solve data imbalance problem, we randomly select two kinds of data with the ratio of $4:1$. Totally, for Dog_2 we select 210 sets of which 168 are inter-ictal and 42 are pre-ictal, for Dog_3 we select 360 sets of which 288 are inter-ictal and 72 are pre-ictal. Every set has 10 minutes data segments, the sample rate is 400Hz and we select 5s as the time window's length. Therefore, the total number of samples of two patients are 25200 and 43200, respectively. Our task is a binary classification task of inter-ictal and pre-ictal. The settings in experiment are listed in Table II.

TABLE II
Parameter and Experiment Settings of Our Tasks

Parameters	Settings
Database	Kaggle Competition database 2016
Patients	Dog_2, Dog_3
Sampling rate	400Hz
Channels	16
Window length	5s
Sample number	Dog_2: 25200 Dog_3: 43200
Per record length	10 minutes
Feature dimension	1×1344
Classifiers	GBDT, RF

B. Experiment I: Single patient seizure prediction based on RF and GBDT

Our task is a binary classification task of inter-ictal and pre-ictal. The allocation ratios of train and test set are 80% and

20%. We repeat the experiment ten times and calculate the average prediction accuracy on the test set. The experiment results on two patients are listed in Table III.

TABLE III
SINGLE PATIENT SEIZURE PREDICTION ACCURACY BASED ON GBDT AND RF

	Dog2		Dog3	
	GBDT	RF	GBDT	RF
1	0.9417	0.8835	0.9341	0.9340
2	0.9826	0.9082	0.9213	0.8839
3	0.9518	0.9277	0.9192	0.9091
4	0.9643	0.9732	0.92	0.9
5	0.9417	0.9320	0.949	0.9388
6	0.9438	0.9551	0.8804	0.9347
7	0.9036	0.9157	0.9681	0.9255
8	0.9467	0.92	0.91	0.91
9	0.9479	0.9270	0.9135	0.8858
10	0.9574	0.9255	0.9271	0.9375
Average	0.9428	0.9268	0.9243	0.9159

The single patient experiment results using the data of Dog_2 and Dog_3 can be found in Table III. Using ensemble learner GBDT and RF to classify inter-ictal and pre-ictal states can achieve high accuracy. The average classification accuracy of GBDT in Dog_2 is more than 0.94 and more than 0.92 in Dog_3; The average accuracy of RF in Dog_2 is more than 0.92 and more than 0.91 in Dog_3.

We also draw the confusion matrix of inter-ictal and pre-ictal classification in Fig. 2.

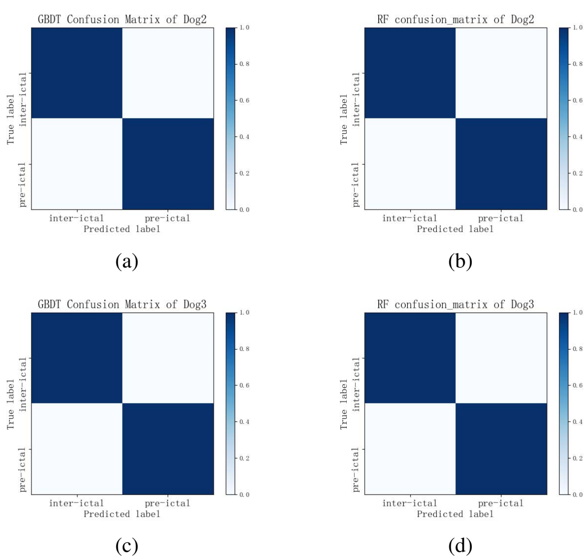

Fig. 2. The confusion matrix of binary classification task with GBDT and RF ensemble learners. (a) GBDT learner for Dog_2; (b) RF learner for Dog_2; (c) GBDT learner for Dog_3; (d) RF learner for Dog_3

C. Experiment II: Cross-patient seizure prediction (baseline of transfer learning methods)

Three evolutional indexes of cross-patient seizure prediction experiment are calculated and listed in Table IV. When predicting seizures cross patients, because of the discrepancy between patients Dog2 and Dog3, the evolutional indexes ACC, TPR, and TNR are all decreasing a lot. The reason for the poor performance is that the train data have a negative influence on the test data, so binary classification probability will even be lower than the random prediction probability 0.5. An improved method based on transfer learning will be proposed in Experiment III.

TABLE IV
CROSS-PATIENT SEIZURE PREDICTION ACCURACY BASED ON GBDT AND RF

	Dog2→Dog3		Dog3→Dog2	
	GBDT	RF	GBDT	RF
ACC	0.4604	0.3271	0.4583	0.3438
TPR	0.6667	0.2833	0.6542	0.2583
TNR	0.2542	0.3708	0.2625	0.4292

D. Experiment III: Cross-patient seizure prediction based on Transfer Learning

When the data of source domain and target domain come from different individuals respectively, i.e., the source domain is Dog2 and the target domain is Dog3, or the source domain is Dog3 and the target domain is Dog2. As the results shown in Experiment II, only using the traditional machine learning classifier will inevitably lead to bad performance, and performed worse than single patient prediction in Experiment I. So we introduce transfer learning methods. Before putting the feature vectors into classifiers, the feature vectors from source and target domain are aligned using the transfer learning methods, mapped to a new common feature space, and then the two newly generated feature vectors $X_s^{'}$ and $X_t^{'}$ can decrease the difference between source and target domain distribution and improve the accuracy of cross-patient classification. The comparison of prediction results before and after introducing transfer learning methods (TCA, JDA, EasyTL [7]) is shown in Table V.

TABLE V
CROSS-PATIENT SEIZURE PREDICTION ACCURACIES BEFORE AND AFTER INTRODUCING TRANSFER LEARNING METHODS

	Before TL		After TL			
	GBDT	RF	TCA +GBDT	TCA + RF	JDA	EasyTL
Dog2→Dog3	0.4604	0.3271	0.5354	0.5167	0.525	0.5229
Dog3→Dog2	0.4583	0.3438	0.5729	0.5583	0.5125	0.5646
Average	0.4594	0.3355	0.5542	0.5375	0.5188	0.5438

Besides the prediction accuracy (ACC), we select other three indexes to evaluate the experiment results, including sensitivity (TPR), specificity (TNR), and false positive rate (FPR). Their calculation formulas are as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$TPR = \frac{TP}{TP + FN} \tag{10}$$

$$TNR = \frac{TN}{TN + FP} \tag{11}$$

$$FPR = \frac{FP}{TP + FN} \tag{12}$$

The values of four evolutional indexes are calculated by above formulas and listed in Table VI.

TABLE VI
THE VALUES OF FOUR EVOLUTIONAL INDEXES FOR THREE EXPERIMENT BEFORE AND AFTER INTRODUCING TRANSFER LEARNING METHODS

	ExperimentI		ExperimentII		ExperimentIII			
	GBDT	RF	GBDT	RF	TCA +GBDT	TCA + RF	JDA	EasyTL
ACC	0.9518	0.9277	0.4583	0.3271	0.5729	0.5583	0.525	0.5646
TPR	0.9348	0.913	0.6542	0.2833	0.6958	0.65	0.1625	0.8375
TNR	0.9730	0.9459	0.2625	0.3708	0.45	0.4667	0.8875	0.2917
FPR	0.0230	0.0455	0.3458	0.7167	0.3042	0.35	0.8375	0.1625

We analyze the results of Table VI:

1) *Experiment I*. This experiment is about single patient seizure prediction, the accuracy based on GBDT and RF are 0.9518 and 0.9277, respectively. Besides, the sensitivity and specificity are better than cross-patient experiments, FRP is lower than cross-patient experiments, which is less than 0.04. Because of the discrepancy between patients, when predicting cross patients, the data from one patient has a negative influence on the other patient's seizure prediction. It is also the motivation of this work to decrease the discrepancy and realize cross-patient seizure prediction.
2) *Experiment II*. In order to identify the effectiveness of transfer learning methods, we carry out some control experiments. The experiment setting is cross-patient but does not introduce any transfer learning methods. It can be seen from the experimental results that the results are poor when it comes to cross patient prediction corresponding to the result of Experiment II in Table VI.
3) *Experiment III*. In order to reduce differences and improve prediction accuracy, we introduce transfer learning methods. In summary, we use four transfer learning methods: 1) TCA transfer learning methods and GBDT classifier named TCA+GBDT in Table VI; 2) TCA transfer learning methods and RF classifier named TCA+RF in Table VI; 3) JDA transfer learning method; 4) EasyTL transfer learning method. Among the four methods, TCA+GBDT has the best results. Compared with ExperimentII, the accuracy of TCA + GBDT method can be improved from 32.71% to 57.29% and the sensitivity is improved from 28.33% to 69.58%, also the accuracy of TCA + RF method can be improved from 32.71% to 55.83%, which is slightly lower than TCA+GBDT. The two methods also improves the specificity by a small margin, with the improvement rate of more than 20%, realizing large improvement relatively.
4) *Biased Results*. For JDA and EasyTL, they have 'biased' promotion. Although they both improve the prediction accuracy, the prediction sensitivity of JDA method is lower than that of no transfer method, and its false positive rate is high. The accuracy is improved by specificity, but our optimization goal is to improve sensitivity and specificity at the same time, also to reduce the false positive rate. As for EasyTL method, its specificity is lower. Therefore, JDA and EasyTL are both biased approach, but they still have some improvement in accuracy, the improvement rates of ACC are 54.64% and 62.10%, respectively.

V. CONCLUSION

A reliable seizure prediction system is significant for practical application. This can ensure the safety of patients, reduce the rate of accidental sudden unexpected death in epilepsy patients, and allow patients to take drugs only when needed, avoiding drug dependence. So we introduce three transfer learning and two ensemble learning algorithms to predict seizures, and construct s framework of seizure prediction system based on transfer learning algorithms. Experiment results on a public dataset demonstrated the effectiveness of our proposed approach, confirming that introducing transfer learning methods has a good effect on the cross-patient seizure prediction task.

ACKNOWLEDGEMENT

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

REFERENCES

[1] 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, pp. 147–165, 2013.

[2] S. M. Usman, S. Khalid, R. Akhtar, Z. Bortolotto, Z. Bashir, and H. Qiu, "Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: Review of available methodologies," *Seizure*, vol. 71, pp. 258–269, 2019.

[3] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Trans. on Biomedical Engineering*, vol. 65, no. 9, pp. 2109–2118, 2017.

[4] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007.

[5] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[6] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int'l Conf. on Computer Vision*, Sydney, Australia, Dec. 2013, pp. 2200–2207.

[7] J. Wang, Y. Chen, H. Yu, M. Huang, and Q. Yang, "Easy transfer learning by exploiting intra-domain structures," in *Proc. IEEE Int'l Conf. on Multimedia and Expo*, Shanghai, China, Jul. 2019, pp. 1210–1215.

[8] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[9] J. H. Friedman, "Greedy function approximation: A gradient boosting machine." *The Annals of Statistics*, vol. 29, no. 5, pp. 1189 – 1232, 2001.

[10] M. J. Cook, T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka *et al.*, "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study," *The Lancet Neurology*, vol. 12, no. 6, pp. 563–571, 2013.

[11] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction—ready for a new era," *Nature Reviews Neurology*, vol. 14, no. 10, pp. 618–630, 2018.

[12] M. Ihle, H. Feldwisch-Drentrup, C. A. Teixeira, A. Witon, B. Schelter, J. Timmer, and A. Schulze-Bonhage, "EPILEPSIAE–a european epilepsy database," *Computer methods and programs in biomedicine*, vol. 106, no. 3, pp. 127–138, 2012.

[13] J. B. Wagenaar, G. A. Worrell, Z. Ives, M. Dümpelmann, B. Litt, and A. Schulze-Bonhage, "Collaborating and sharing data in epilepsy research," *Journal of Clinical Neurophysiology*, vol. 32, no. 3, p. 235, 2015.

[14] L. G. Kini, K. A. Davis, and J. B. Wagenaar, "Data integration: Combined imaging and electrophysiology data in the cloud," *Neuroimage*, vol. 124, pp. 1175–1181, 2016.

[15] 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.

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

[17] J. J. Howbert, E. E. Patterson, S. M. Stead, B. Brinkmann, V. Vasoli, D. Crepeau, C. H. Vite, B. Sturges, V. Ruedebusch, J. Mavoori *et al.*, "Forecasting seizures in dogs with naturally occurring epilepsy," *PloS One*, vol. 9, no. 1, p. e81920, 2014.

[18] B. H. Brinkmann, J. Wagenaar, D. Abbot, P. Adkins, S. C. Bosshard, M. Chen, Q. M. Tieng, J. He, F. J. Muñoz-Almaraz, P. Botella-Rocamora, J. Pardo, F. Zamora-Martinez, M. Hills, W. Wu, I. Korshunova, W. Cukierski, C. Vite, E. E. Patterson, B. Litt, and G. A. Worrell, "Crowdsourcing reproducible seizure forecasting in human and canine epilepsy," *Brain*, vol. 139, no. 6, pp. 1713–1722, 2016.

[19] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, J. Timmer, and A. Schulze-Bonhage, "Do False Predictions of Seizures Depend on the State of Vigilance? A Report from Two Seizure-Prediction Methods and Proposed Remedies," *Epilepsia*, vol. 47, no. 12, pp. 2058–2070, 2006.

[20] 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.

[21] R. G. Andrzejak, K. Schindler, and C. Rummel, "Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients," *Physical Review E*, vol. 86, no. 4, p. 046206, 2012.

[22] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int'l Joint Conf. on Neural Networks*, Rio de Janeiro, Brazil, Jul. 2008, pp. 2390–2397.

[23] F. Mormann, R. G. Andrzejak, T. Kreuz, C. Rieke, P. David, C. E. Elger, and K. Lehnertz, "Automated detection of a preseizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients," *Physical Review E*, vol. 67, no. 2, p. 021912, 2003.

[24] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Networks*, vol. 105, pp. 104–111, 2018.

[25] M. Alhussein, G. Muhammad, M. S. Hossain, and S. U. Amin, "Cognitive IoT-cloud integration for smart healthcare: Case study for epileptic seizure detection and monitoring," *Mobile Networks and Applications*, vol. 23, no. 6, pp. 1624–1635, 2018.

[26] Z. Yu, W. Nie, W. Zhou, F. Xu, S. Yuan, Y. Leng, and Q. Yuan, "Epileptic seizure prediction based on local mean decomposition and deep convolutional neural network," *The Journal of Supercomputing*, vol. 76, no. 5, pp. 3462–3476, 2020.

[27] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.

[28] N. Dhulekar, S. Nambirajan, B. Oztan, and B. Yener, "Seizure prediction by graph mining, transfer learning, and transformation learning," in *Proc. Int'l Conf. on Machine Learning and Data Mining in Pattern Recognition*, Hamburg, Germany, Jul. 2015, pp. 32–52.

[29] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, and P. L. Kubben, "EEG based multi-class seizure type classification using convolutional neural network and transfer learning," *Neural Networks*, vol. 124, pp. 202–212, 2020.

[30] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Trans. on Biomedical Circuits and Systems*, vol. 13, no. 5, pp. 804–813, 2019.

[31] M. J. Shensa *et al.*, "The discrete wavelet transform: Wedding the a trous and Mallat algorithms," *IEEE Trans. on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, 1992.

[32] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. 2039–2049, 2000.

[33] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

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

[35] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. 49–57, 2006.