Computers in Biology and Medicine 150 (2022) 106169

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A general sample-weighted framework for epileptic seizure prediction

Yikai Gao [a,b], Aiping Liu [b,c,*], Xinrui Cui [b], Ruobing Qian [a], Xun Chen [a,b,c,*]

[a] Epilepsy Center, Department of Neurosurgery, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230001, China

[b] School of Information Science and Technology, University of Science and Technology of China (USTC), Hefei 230027, China

[c] USTC IAT-Huami Joint Laboratory for Brain-Machine Intelligence, Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China

ARTICLE INFO

Keywords:
Epileptic seizure prediction
Machine learning
Deep learning
Sample-weighted
Genetic algorithm

ABSTRACT

Objective: Effective epileptic seizure prediction can make the patients know the onset of the seizure in advance to take timely preventive measures. Many studies based on machine learning methods have been proposed to tackle this problem and achieve significant progress in recent years. However, most studies treat each EEG training sample's contribution to the model as equal, while different samples have different predictive effects on epileptic seizures (e.g., preictal samples from different times). To this end, in this paper, we propose a general sample-weighted framework for patient-specific epileptic seizure prediction.

Methods: Specifically, we define the mapping from the sample weights of training sets to the performance of the validation sets as the fitness function to be optimized. Then, the genetic algorithm is employed to optimize this fitness function and obtain the optimal sample weights. Finally, we obtain the final model by using the training sets with optimized sample weights.

Results: To evaluate the effectiveness of our framework, we conduct extensive experiments on both traditional machine learning methods and prevalent deep learning methods. Our framework can significantly improve performance based on these methods. Among them, our framework based on Transformer achieves an average sensitivity of 94.6%, an average false prediction rate of 0.06/h, and an average AUC of 0.939 in 12 pediatric patients from the CHB-MIT database with the leave-one-out method, which outperforms the state-of-the-art methods.

Conclusion: This study provides new insights into the field of epileptic seizure prediction by considering the discrepancies between EEG samples. Moreover, we develop a general sample-weighted framework, which applies to almost all classical classification methods and can significantly improve performance based on these methods.

1. Introduction

As one of the most common neurological diseases, epilepsy affects about 50 million people worldwide. With the development of modern medicine during the decades, 70% of patients with epilepsy can become seizure-free with appropriate treatment. However, there are still 30% of patients who suffer from intractable epilepsy [1]. Therefore, research of epileptic seizure prediction is significantly valuable for these patients, which could make the patients know the onset of the seizure in advance to take timely preventive measures.

Electroencephalography (EEG) is an efficient technique to record electrical activity of the brain and is often used clinically to diagnose epilepsy [2–4]. In the past few years, researchers have shown that EEG signals could be used to predict seizures by machine learning methods [5–7]. Generally, physicians divide the consecutive EEG signals of epileptic patients into four states: preictal (the period before the seizure), ictal (the period of seizure), postictal (the period following the seizure), and interictal (the period between seizures) [8]. Hence, we can describe the epileptic seizure prediction task as a binary classification problem differentiating the preictal and interictal states. Many advanced methods based on machine learning have been developed to address this challenging issue following this pattern.

Traditional EEG-based seizure prediction methods mainly focus on two parts: feature extraction and classification. For feature extraction, researchers generally construct features of EEG signals manually based on experience or observations (e.g., time domain, frequency domain, and time–frequency domain features). A classifier is designed for classification subsequently. For example, Chisci et al. used the autoregressive

* Corresponding authors.
E-mail addresses: aipingl@ustc.edu.cn (A. Liu), xunchen@ustc.edu.cn (X. Chen).

https://doi.org/10.1016/j.compbiomed.2022.106169
Received 13 June 2022; Received in revised form 16 September 2022; Accepted 1 October 2022
Available online 5 October 2022

coefficient as the features of EEG signals. Then, a support vector machine (SVM) method was adopted for classification [9]. Bedeeuzzaman et al. built a statistical feature set, including mean absolute deviation and inter quartile range, and classified them by a linear classifier [10]. Zhang et al. extracted features of EEG signals by using spectral analysis, such as absolute spectral powers, relative spectral powers, and spectral power ratios. Then a two-step feature selection method is employed on these features and input into a SVM classifier [11]. These studies offer a strong foundation for EEG-based epileptic seizure prediction.

Recently, deep learning methods have attracted significant attention due to their powerful feature extraction capability. Truong et al. used the short-time Fourier transform (STFT) to obtain the time–frequency features of EEG signals and classified them with a convolutional neural network (CNN) classifier [12]. Ozcan et al. extracted features by considering the position of the electrodes and adopted a 3D-CNN method to predict seizures [13]. Although these studies have achieved satisfactory performance, they still depend on manually extracting features. To this end, Daoud et el. took the raw EEG signal as the input without any preprocessing. Specifically, they used a CNN to extract features automatically and classified them with a recurrent neural network (RNN) method [14]. Ma et al. developed an end-to-end model with improved Long Short Term Memory networks and channel-and-spatial attention for automatic seizure prediction [15]. These deep learning-based methods have greatly improved the performance of epileptic seizure prediction.

Back to the paradigm for seizure prediction, current studies rely on the hypothesis that a "preictal" brain state exists, which would lead to a seizure in the following period [16]. To this end, most studies define a fix-length period empirically before the seizure as the preictal state (e.g., 30 min before the seizure). According to this definition, researchers label EEG samples from different time windows as the same class and treat each EEG training sample's contribution to the model as equal. However, different samples have different predictive effects on epileptic seizures. For example, the EEG clip taken three minutes before seizure onset and the EEG clip taken twenty minutes before seizure onset were both labeled preictal, but these two samples may have completely different predictive effects on epileptic seizures. Hence, an efficient epileptic seizure prediction framework should consider the discrepancies between samples.

Several studies have look at the problem above, caused by the definition of preictal and interictal. For instance, Li et al. tried to adaptively infer the subject-specific optimal preictal period by a semi-supervised active learning strategy [17]. But their method is still limited since the precursory symptoms of a seizure may not occur consecutively over time [18,19]. Nasseri et al. proposed a semi-supervised method to select preictal training data, improving the prediction performance [20]. However, the preictal samples they deleted may also help predict seizures. Besides, these studies all ignore the discrepancies between training samples. Ozcan et al. assigned a time-varying weight to each preictal sample while training the model. Nevertheless, their sample reweighting strategy is linear over time, so it cannot effectively improve the performance effectively [13]. Hence, we develop a general sample-weighted framework for epileptic seizure prediction in this study. The diagram of our proposed framework is shown in Fig. 1.

Specifically, our framework would assign a weight to each training sample for controlling its contribution to the classifier. To optimize the weights of training samples, we define the mapping from the sample weights of training sets to the performance of the validation sets as the fitness function to be optimized. Intuitively, we adjust the sample weights so that our classifier trained on the training set performs better on the validation set. A genetic algorithm (GA) is used to optimize this fitness function in this study. It is worth mentioning that our framework can be applied to almost any machine learning classifier, and the algorithm for optimizing sample weights is not limited to GA. To evaluate the effectiveness of our framework, we conduct experiments on both traditional machine learning methods and prevalent deep learning methods (e.g., SVM, CNN, and Transformer) on the CHB-MIT scalp EEG database with the leave-one-out strategy.

Table 1
Subject information of the CHB-MIT scalp EEG dataset.

Case	Gender	Age (years)	No. of seizures	Interictal times (h)
Chb-01	F	11	7	14.3
Chb-03	F	14	6	25.4
Chb-05	F	7	5	14.4
Chb-09	F	10	4	48.5
Chb-10	M	3	7	24.3
Chb-13	F	3	7	15.1
Chb-14	F	9	8	4.7
Chb-16	F	7	8	5.6
Chb-18	F	18	4	25.0
Chb-20	F	6	8	18.0
Chb-21	F	13	4	23.4
Chb-23	F	6	7	14.2
Total	–	–	75	232.9

The main contributions are as follows:

- We propose a general sample-weighted framework to solve the noisy label problem caused by the definition of preictal and interictal. Our framework can be applied to almost any machine learning classifier and significantly improve their performance.
- Our framework based on Transformer achieves state-of-the-art performance on the CHB-MIT scalp EEG dataset. Furthermore, our proposed framework is robust to the class imbalance problem by adjusting the sample weight.
- The insight into why our proposed framework works can be well interpreted. Particularly when training the model, our method can select and focus on more valuable samples for seizure prediction. These samples prompt the model to perform better on the validation set, thus further improving the generalization of the model.

The rest paper is organized as follows. Section 2 presents the datasets used in this study and our proposed methods. Section 3 describes our experimental results and comparison. Section 4 provides discussion of our methods. Finally, we conclude our work in Section 5.

2. Methodology

2.1. Data description

In this work, we use the CHB-MIT scalp EEG database to evaluate the effectiveness of our framework for epileptic seizure prediction [21,22]. The CHB-MIT database recorded the long-duration EEG signals of 23 pediatric subjects with intractable epilepsy. In the CHB-MIT dataset, all EEG signals were collected with a sampling rate of 256 Hz. The electrodes were placed according to the 10–20 international system, using a bipolar montage. To guarantee the consistency of our approach, according to [23], we select 18 channels common to each patient in this study, including FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, and CZ-PZ.

Before introducing our methods, we define some relevant parameters. According to [8], the preictal period is defined as 30 min before the occurrence of a seizure. The interictal period is defined as more than 240 min after a seizure and more than 240 min before the subsequent seizure. If two consecutive seizures are less than 15 min apart, the subsequent seizure is excluded due to a lack of preictal data. Besides, in order to prevent overfitting problems, we select patients from the CHB-MIT database who have more than three seizures and whose interictal duration was more than three hours. Consequently, 12 patients are selected in this study, and Table 1 presents the data information. Finally, we divide the consecutive EEG recordings into 30-second windows for classification according to [12].

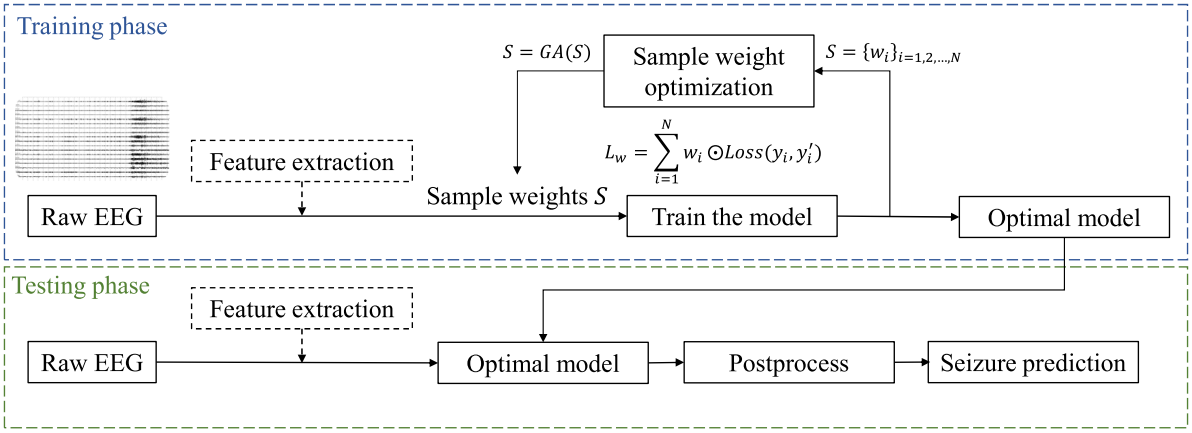

Fig. 1. The diagram of our proposed weight-sample-based framework. Since the "feature extraction" step is often used in traditional machine learning methods, but is not required in deep learning-based methods, we place it in a dashed box. $GA(S)$ represents the optimization process of sample weight S by genetic algorithm, and the detailed optimization process is shown in Fig. 2. The L_w denotes the loss function during classification model training. The operator $\odot$ represents the way the sample weight affects the loss function, which could have different meanings in different classification algorithms (e.g., the multiplication in the deep learning-based method).

2.2. Weight-sample-based framework with a genetic algorithm

By considering the discrepancies between samples, we develop a sample-weighted machine learning framework with a genetic algorithm. Fig. 1 presents the diagram of our proposed framework. Our framework improves the training phase by optimizing the training sample weights compared with general seizure prediction methods, allowing the model to pay more attention to the more valuable samples. Furthermore, it is worth mentioning that the trained model in our framework applies to almost all classical classification methods. To show that, in our study, we use the SVM, CNN, and Transformer to complete the classification, respectively.

2.2.1. Sample weight optimization — genetic algorithm

The genetic algorithm (GA) is a heuristic search algorithm, which has certain advantages in solving nonlinear problems and has been applied in different research fields [24–26]. In this study, we employ GA to optimize the sample weights.

Specifically, we first divide the data set into a training, validation, and test set and define the mapping from the sample weights of the training set to the performance of the validation set as the fitness function to be optimized. Then GA codes the sample weight S in the parameters population and randomizes the initial parameters population. After the coding and initialization, GA improves the fitness function through generations. Intuitively, GA optimizes the training sample weights to prompt the model to perform better on the validation set, thus further improving the generalization of the model. Fig. 2 shows the flowchart of sample weight optimization with GA.

During the optimization process, we use three main operators: selection, crossover, and mutation.

(1) Selection: Selection is an operator of selecting some better individuals from a population of individuals and the selected individuals will be used for the crossover. For our work, the better sample weight vector S will be selected in this step (i.e., perform better on the validation set). We employ the tournament selection strategy since it has lower time complexity.

(2) Crossover: In this step, the selected individuals (coded S) will be used for crossover to generate new offspring. Specifically, we use the two-point crossover in this work [27], and the schematic diagram of the two-point crossover is presented as Fig. 3.

(3) Mutation: Mutation is the process of generating new chromosomes by changing a part of a chromosome. It can improve population diversity and reduce the risk of evolutionary algorithms falling into local optimal solutions. In this study, we use an improved form of intermediate crossover and line crossover named Breeder mutation operator.

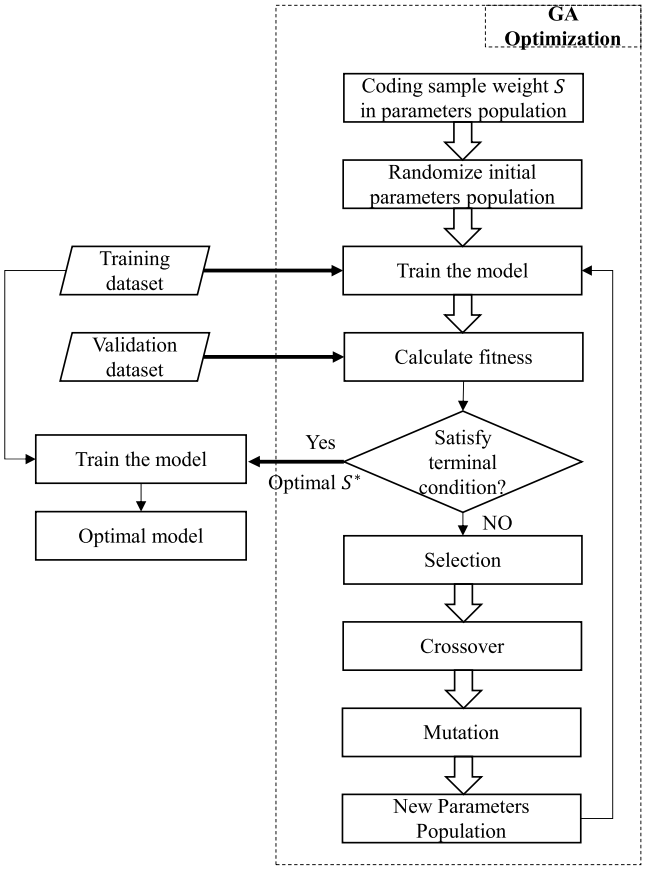

Fig. 2. The flowchart of sample weight optimization with GA.

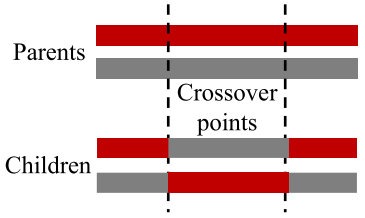

Fig. 3. The diagram of two-point crossover.

Table 2
Features used for SVM classification.

	Features
Statistical moment	Mean
	Variance
	Skewness
	Kurtosis
Spectral band power	Delta (0.5 – 4 Hz)
	Theta (4–8 Hz)
	Alpha (8–13 Hz)
	Beta (13–30 Hz)
	Gamma-1 (30–50 Hz)
	Gamma-2 (50–75 Hz)
	Gamma-3 (75–100 Hz)
	Gamma-4 (100–128 Hz)
Hjorth parameters	Mobility
	Complexity

2.2.2. Manual feature extraction and SVM classification

As one of the most prevailing classification algorithms, SVM has been widely used in epileptic seizure prediction [9,28,29]. Hence, SVM is very suitable for verifying whether our framework is effective to traditional methods. We first extracted features manually from EEG signals and classified these features with an SVM classifier.

(1) Feature extraction: Feature extraction plays a significant role in the pipeline. The quality of extracted features directly affects the classification performance. Specifically, we utilize the statistical moments, spectral band power and Hjorth parameters as features according to [13]. The statistical moments include mean, variance, skew, and kurtosis. For about spectral band power, we use the traditional periodogram method to estimate power spectral density, and the window length and the number of FFT points are set to the length of the input signal. We separate the EEG signal into 8 spectral bands to further characterize the frequency domain features. The mobility and complexity in Hjorth parameters are also employed as our features. The detailed features information is described as Table 2.

(2) SVM classification: In this study, we adjust the importance weights of different samples by modifying the penalty coefficients of different samples in the SVM loss function.

After manual feature extraction, each EEG sample is converted into a training vector $x_i \in \mathbb{R}^d, i = 1, 2, \ldots, n$, where n is the number of samples, $d = 14 * c$, and c is the number of channels. And there is also a label vector $y \in \{-1, 1\}^n$ that represents the label of the samples. Our objective is to compute $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$ so that the prediction results obtained by $sign(\omega^T \phi(x) + b)$ is correct for most samples. Therefore, we can get the following SVM optimization problem:

$$\begin{aligned} &\min_{\omega, b, \zeta} \ \frac{1}{2}\omega^T\omega + C_i \sum_{i=1}^{n} \zeta_i \\ &\text{s.t.} \ \ y_i \left(\omega^T \phi\left(x_i\right) + b\right) \geq 1 - \zeta_i \\ &\zeta_i \geq 0, i = 1, \ldots, n \end{aligned} \tag{1}$$

where ξ_i indicates the distance from the sample x_i to its corresponding correct margin boundary, and C_i denotes the strength of penalty for the sample x_i. The mapping $\phi(\cdot)$ is a kernel-dependent feature transformation, and the linear kernel is applied in our study.

2.2.3. Deep learning method - CNN

In recent years, CNN has been widely used as an end-to-end machine learning method, automatically extracting discriminative features of raw EEG signals. We design a 1D-CNN model to classify EEG signals in this study. The structure of our model is presented on Fig. 4 and Table 3 shows the specific parameters.

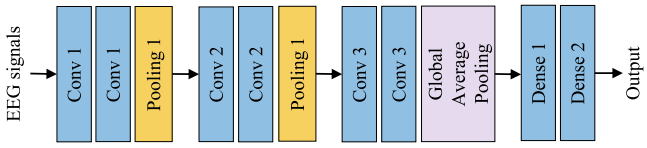

Fig. 4. The architecture of our CNN model.

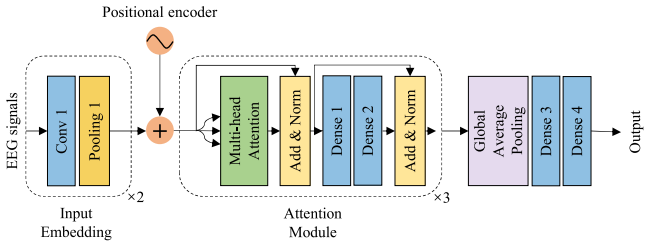

Fig. 5. The architecture of our Transformer model.

2.2.4. Deep learning method — transformer

Transformer is a deep learning architecture relying on self-attention to compute representations, which has achieved great success in various artificial intelligence fields [30,31]. Since the EEG signal is a classical time series signal, the transformer model has advantages over other deep learning models in predicting epileptic seizures. In this work, a transformer-based model is designed to predict epileptic seizures. Since the ultimate goal of our model is classification, we only use the encoder part of the transformer architecture. There are three primary parts in our Transformer-based model (i.e., input embedding, positional encoding, and attention module), and we present the detailed structure of our model in Fig. 5.

(1) Input embedding: The input embedding is the first step of our model, and it is to reduce the input dimension and obtain a preliminary feature from raw EEG signals. Specifically, we apply a 1D-CNN layer to extract preliminary features from raw input EEG signals, and the output size of the input embedding is defined as d_{model}.

(2) Positional encoding: To better use the order of the sequence, we use the learned positional encoding scheme to add positional information to the input embedding [32].

(3) Attention module: The attention module is composed of two components: the multi-head attention and the feed-forward network. For the multi-head attention, We generate the query, key, and value vectors by scaled dot-product attention. [30]. Specifically, these vectors are generated by multiplying the input by the learned weight matrices W_q, W_k and W_v (i.e. query, key and value). Then, we pack the queries, keys and values individually to create Q, K, and V. Finally, the attention values are calculated as below:

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{2}$$

where $Q \in \mathbb{R}^{l \times d_k}, K \in \mathbb{R}^{l \times d_k}$ and $Q \in \mathbb{R}^{l \times d_k}$. The l denotes the length of sequence and d_k represents the embedding dimension of our model here. In addition, We employ two fully connected layers for the feed-forward network. And 50% dropout is used in the attention module to prevent overfitting problem. The value of d_{model} and d_k are both set as 64 in our study. Besides, the additional model specifications are described in Table 3.

2.3. Training and testing

In this work, the model is trained and tested for specific patients. When training the SVM classifier, we solve the optimization problem (1). According to the formulation, we can control the importance weight of samples in SVM by adjusting the value of C_i for each sample.

Table 3
The parameters of our designed CNN and transformer model.

Layers	Sizes	Channels	Activation function
CNN-based model:			
Conv1	5	64	ReLU
Conv2	3	32	ReLU
Conv3	3	256	ReLU
Pooling1	2	–	–
Dense1	32	–	sigmoid
Dense2	1	–	sigmoid
Transformer-based model:			
Conv1	5	64	ReLU
Pooling1	4	–	–
Dense1	128	–	ReLU
Dense2	64	–	ReLU
Dense3	32	–	ReLU
Dense4	1	–	sigmoid

In this study, $S = \{C_i\}_{i=1,2,\ldots,n}$ is defined as the sample weights to be optimized in our framework, and we find that the model works better when $C_i \in \{0.1, 1, 10, 100\}$. The fitness of an individual is the classification accuracy on the validation set.

While training the CNN and transformer model, we used a weight-sample based cross-entropy function as the training loss. Specifically, the standard binary form of cross-entropy function is formulated by:

$$L_b = -\sum_{i=1}^{N} -\left(1-y_i\right)\log\left(1-\tilde{y}_i\right) - y_i \log\left(\tilde{y}_i\right) \tag{3}$$

and the sample-weighted cross-entropy form is described as:

$$L_w = -\sum_{i=1}^{N} w_i \left(-\left(1-y_i\right)\log\left(1-\tilde{y}_i\right) - y_i \log\left(\tilde{y}_i\right)\right) \tag{4}$$

where N is the number of training samples, $\tilde{y}_i$ and y_i are the network output and its true label of the ith sample, respectively. The term w_i represents the weight to be optimized assigned to the ith sample. In our experiments, we set $w_i \in \{0, 0.25, 0.5, 0.75, 1\}$. The AUC value on the validation set is taken as the fitness of an individual. Besides, the deep learning models are trained for five epochs in each GA iteration. For the parameter setting of GA, we set the population size as 50, and the maximum number of generations is 30. The optimization of sample weights by GA stops at the maximum number of generations (i.e. terminal condition).

In addition, we apply the leave-one-out cross-validation (LOOCV) method to evaluate our framework. Specifically, the dataset is divided into three parts: the training set, validation set, and test set. For instance, assuming that a subject has a total of N seizures, there are N corresponding preictal periods. Then, we randomly separate all interictal data into N equal parts and combine them with N preictal periods to be N pairs. As the LOOCV method indicates, in each round, we take out one of the pairs as a test set. The training and validation sets are randomly divided from the remaining N-1 pairs according to the ratio 3:1. Therefore, the model trained and tested N times for this subject.

Finally, the optimal model is selected by an early stopping scheme. The model stops training once the loss of validation set does not decrease for ten consecutive epochs. We select the model with the least loss on the validation set. In this work, our proposed model is developed in Python 3.8.11 environment and Keras 2.6.0 with a Tensorflow 2.5.0 backend. We use the python package Geatpy 2.6.0 for GA and implemented SVM by the python package Scikit-learn in this study [33,34].

2.4. Postprocess

To reduce the impact of false prediction, we apply the k-of-n methods in [12] to generate the alarm for patients. Specifically, an alarm will occur only if the predicted value is preictal in at least k of n consecutive EEG samples. We set the values of k and n to 8 and 10, respectively. Besides, the refractory period is set to 30 mins in this study to prevent continuous alarm for a short time.

2.5. Comparative methods

To further measure the efficiency of our proposed framework, we compare several state-of-the-art approaches with our model. All of the following methods are evaluated on the CHB-MIT dataset.

SVM with Phase Locking Value [35] used the phase locking value of EEG signals as features and classified these features by an SVM classifier.

CNN with STFT Spectral Analysis [12] used the STFT to extract EEG signals' time–frequency features and classified them with a CNN method.

3D CNN with Manual Features [13] designed the model by considering the location of the electrodes in EEG signals. Spectral power, statistical moments, and Hjorth parameters are used as manual features of EEG signals, which are classified by a 3D CNN method.

CNN with Common Spatial Pattern Statistics [23] utilized the common spatial pattern method to design spatial filters and extract features to maximize the discriminability of preictal and interictal. Then, a CNN model was performed to classify these features.

RDANet with STFT Spectral Images [8] applied STFT to EEG signal analysis to obtain the time–frequency features. Then, they classified the features by designing a dual self-attention residual network (RDANet).

GCN with active preictal interval learning [17] proposed a hierarchical graph convolutional network for capturing the preictal transitions from the EEG signals. Furthermore, a semi-supervised active preictal interval learning scheme is developed to infer the optimal patient-specific preictal interval.

3. Experiments and results

To evaluate the effectiveness of our proposed framework, we conduct a series of experiments on the CHB-MIT dataset. In this part, we depict the details of our experiments and the evaluation metrics. Furthermore, the experimental results and comparisons are described below.

3.1. Experiments and evaluation metrics

In this study, we first conducted benchmark experiments without sample weighting based on SVM, CNN, and Transformer, respectively. Then, we carry out the experiments with the above three models but optimize the samples weights based on our proposed framework.

For evaluation metrics, we adopt four parameters to evaluate the performance of our framework: sensitivity (Sens), false prediction rate (Fpr), area under the curve (AUC), and p-value. Sens is formulated as the proportion of seizures correctly predicted to the total number of seizures. Fpr is formulated as the number of false alarms per hour. AUC is defined as the area under the receiver operating characteristic curve, which is a metric commonly used for binary classification.

To measure the significance of an improvement over chance-level for our model, we calculate the p-value according to [13]. Specifically, assuming that the interval between two successive alarms follows an exponential distributed Poisson process, the probability that at least one alarm rises randomly in an interval of Δt is approximately equal to $\lambda_w \Delta t$, independent of t, where λ_w is called the Poisson rate parameter. In this context, the sensitivity of the chance predictor, S_{nc}, is defined as follows:

$$S_{nc} = 1 - \exp\left(-\lambda_w \tau_w + \left(1 - e^{-\lambda_w \tau_{w0}}\right)\right) \tag{5}$$

Table 4
Comparison to recent seizure prediction methods on CHB-MIT scalp EEG dataset.

Authors	Dataset	Features	Classifier	No. of seizures	No. of subjects	Fpr (/h)	Sens (%)	AUC	Interictal distance (minutes)	Preictal length (minutes)
Cho et al., 2017 [35]	CHB-MIT	Phase locking value	SVM	65	21	–	82.44	–	30	5
Truong et al., 2018 [12]	CHB-MIT	STFT spectral images	CNN	64	13	0.16	81.2	–	240	30
Ozcan et al., 2019 [13]	CHB-MIT	Spectral power Statistical moments Hjorth parameters	3D CNN	77	16	0.09	87.2	0.886	240	60
Zhang et al., 2020 [23]	CHB-MIT	Common spatial pattern statistics	CNN	156	23	0.11	93.6	0.900	–	30
Yang et al., 2021 [8]	CHB-MIT	STFT spectral images	RDANet	64	13	–	89.3	0.913	240	30
Li et al., 2021 [17]	CHB-MIT	ICs from ICA	GCN	98	19	0.11	92.9	0.938	Adaptive	Adaptive
Our work	**CHB-MIT**	**End-to-end**	**Sample-weighted transformer**	**75**	**12**	**0.06**	**94.6**	**0.939**	**240**	**30**

Fpr: false prediction rat; Sens: sensitivity; AUC: area under the receiver operating characteristic curve; ICA: independent components analysis; ICs: independent component.

Table 5
Seizure prediction performance based on SVM for All 12 patients.

Pats	SVM without sample weighted				SVM with sample weighted			
	Sens (%)	Fpr (/h)	AUC	*p*-value	Sens (%)	Fpr (/h)	AUC	*p*-value
Pat-1	100.0	0.07	0.981	<0.001	100.0	0.00	0.979	<0.001
Pat-3	50.0	0.11	0.742	0.005	66.7	0.04	0.784	0.001
Pat-5	60.0	0.14	0.915	0.005	80.0	0.14	0.910	0.006
Pat-9	50.0	0.10	0.601	0.012	50.0	0.06	0.649	0.026
Pat-10	71.4	0.29	0.775	0.002	71.4	0.21	0.719	0.002
Pat-13	85.7	0.27	0.846	0.002	100.0	0.27	0.837	<0.001
Pat-14	87.5	0.21	0.864	0.011	100.0	0.00	0.867	0.001
Pat-16	75.0	0.36	0.789	0.006	87.5	0.18	0.798	0.007
Pat-18	75.0	0.16	0.491	<0.001	75.0	0.08	0.624	0.003
Pat-20	100.0	0.06	0.970	<0.001	100.0	0.06	0.975	<0.001
Pat-21	100.0	0.04	0.863	<0.001	100.0	0.00	0.872	<0.001
Pat-23	100.0	0.14	0.967	<0.001	100.0	0.07	0.963	<0.001
Ave	79.6	0.16	0.817	n.a.	**85.9**	**0.09**	**0.831**	n.a.

Table 6
Seizure prediction performance based on CNN for all 12 patients.

Pats	CNN without sample weighted				CNN with sample weighted			
	Sens (%)	Fpr (/h)	AUC	*p*-value	Sens (%)	Fpr (/h)	AUC	*p*-value
Pat-1	100.0	0.14	0.989	<0.001	100.0	0.14	0.992	<0.001
Pat-3	83.3	0.16	0.855	<0.001	100.0	0.11	0.876	<0.001
Pat-5	80.0	0.14	0.898	0.004	80.0	0.07	0.901	<0.001
Pat-9	50.0	0.10	0.775	0.210	75.0	0.06	0.862	0.010
Pat-10	71.4	0.08	0.838	<0.001	71.4	0.00	0.850	<0.001
Pat-13	100.0	0.47	0.818	<0.001	100.0	0.33	0.862	<0.001
Pat-14	75.0	0.43	0.890	0.095	87.5	0.00	0.893	0.012
Pat-16	87.5	0.18	0.928	0.008	100.0	0.00	0.944	<0.001
Pat-18	75.0	0.24	0.708	0.006	75.0	0.12	0.667	0.005
Pat-20	100.0	0.11	0.969	<0.001	100.0	0.06	0.975	<0.001
Pat-21	100.0	0.00	0.916	<0.001	100.0	0.00	0.904	<0.001
Pat-23	100.0	0.14	0.974	<0.001	100.0	0.14	0.974	<0.001
Ave	85.2	0.18	0.880	n.a.	**90.7**	**0.09**	**0.892**	n.a.

where the detection interval τ_{w0} denotes the seizure prediction horizon (SPH), while τ_w represents the sum of SPH and the seizure occurrence period (SOP). The difference between observed and chance sensitivity depend on ρ_w is a strong measure of predictability [36]. For a seizure prediction method with sensitivity S_n and proportion of time-in-warning ρ_w, the sensitivity improvement-over-chance metric is given as below:

$$S_n - S_{nc} = S_n - 1 + \exp\left(-\lambda_w \tau_w + \left(1 - e^{-\lambda_w \tau_{w0}}\right)\right) \tag{6}$$

where

$$\lambda_w = -\frac{1}{\tau_w} \ln\left(1 - \rho_w\right) \tag{7}$$

Assuming that the proposed prediction method correctly identifies n out of N seizures for an individual subject, the significance of an improvement over chance is evaluated by the one-sided p-values, as given in (8).

$$p = 1 - \sum_{i=0}^{n-1} \binom{N}{i} S_{nc}^{i} \left(1 - S_{nc}\right)^{N-i}, \quad \text{for } \frac{n}{N} \geq S_{nc} \tag{8}$$

In this case, the corresponding hypotheses can be described as follows:

H_0 : median $((S_n - S_{nc})$ for algorithm$) = 0$

H_1 : median $((S_n - S_{nc})$ for algorithm$) \neq 0$

When we calculate the p-value, the significant level p is set to 0.05.

3.2. Results and comparison

To better illustrate the effectiveness of our proposed framework, we compare the seizure prediction performance of three models (SVM, CNN, and Transformer) with and without sample weighting for 12 patients on the CHB-MIT dataset. The comparison results are shown in Table 5, Table 6, and Table 7, respectively.

According to the results in Tables 5–7, we can find that the performance of all three models is improved significantly when the sample weights are optimized by our framework. Specifically, our framework improves sensitivity by an average of 6.1%, reduces the false prediction rate by an average of 0.07/h, and improves AUC by an average of 0.026. Furthermore, our transformer-based model can achieve an average Sens of 94.6%, an average Fpr of 0.06/h, and an average AUC of 0.939. When we compare our results with chance predictor for each patient, 9 out of 12 patients s have p-values less than 0.001.

Also, we compare our experimental performance with those reported in other recently published methods using the same CHB-MIT scalp EEG database. It is worth mentioning that not every study on seizure prediction using the CHB-MIT dataset used the same patients. Hence, to make a valid comparison, for a study with more subjects than we did, their performance obtained when they select the same patients as in our study is compared with ours. Otherwise, their optimal performance is compared with ours. Table 4 list the comparison results. Our framework achieves the optimal performance in Sens, Fpr and AUC. It is shown that our proposed framework is superior to all other state-of-the-art approach.

4. Discussion

The above experimental results show the superiority of our proposed framework. Whether based on traditional or deep learning methods, our framework can improve prediction performance by using a sample re-weighting strategy. Nevertheless, the reasons behind our framework's improved performance are worth discussing. Therefore, we carefully examined the EEG samples assigned with high weights

Table 7
Seizure prediction performance based on transformer for all 12 patients.

Pats	Transformer without sample weighted				Transformer with sample weighted			
	Sens (%)	Fpr (/h)	AUC	*p*-value	Sens (%)	Fpr (/h)	AUC	*p*-value
Pat-1	100.0	0.07	0.995	<0.001	100.0	0.07	0.997	<0.001
Pat-3	83.3	0.08	0.924	<0.001	100.0	0.08	0.959	<0.001
Pat-5	80.0	0.14	0.908	<0.001	100.0	0.07	0.958	<0.001
Pat-9	75.0	0.42	0.763	0.012	75.0	0.21	0.874	0.005
Pat-10	71.4	0.04	0.812	<0.001	85.7	0.04	0.919	<0.001
Pat-13	85.7	0.27	0.901	0.001	100.0	0.00	0.964	<0.001
Pat-14	100.0	0.21	0.917	0.001	100.0	0.21	0.922	0.012
Pat-16	87.5	0.00	0.951	<0.001	100.0	0.00	0.956	<0.001
Pat-18	75.0	0.04	0.717	0.010	75.0	0.00	0.808	0.007
Pat-20	100.0	0.06	0.983	<0.001	100.0	0.06	0.989	<0.001
Pat-21	100.0	0.00	0.920	<0.001	100.0	0.00	0.937	<0.001
Pat-23	100.0	0.00	0.991	<0.001	100.0	0.00	0.987	<0.001
Ave	88.2	0.11	0.899	n.a.	**94.6**	**0.06**	**0.939**	n.a.

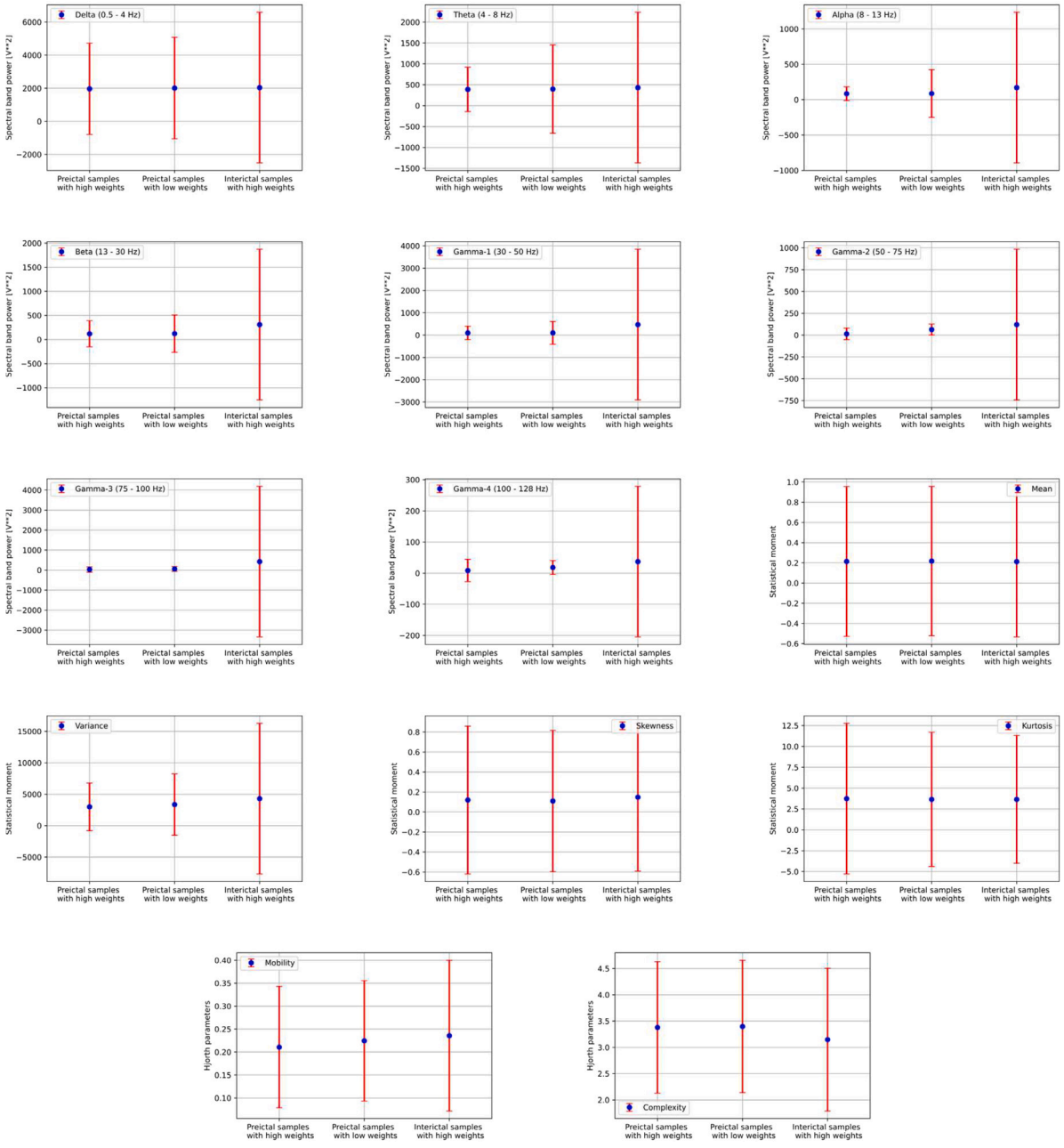

Fig. 6. Statistical analysis of training samples with different weights.

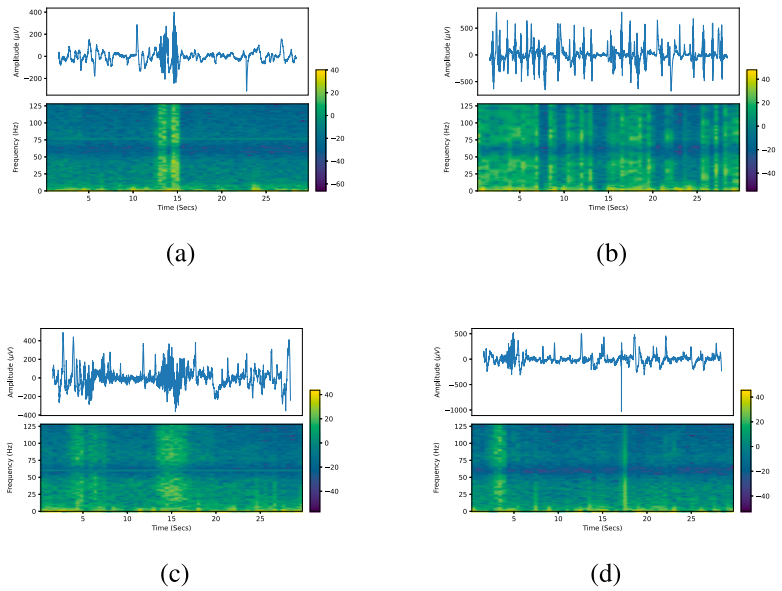

Fig. 7. Original preictal EEG samples assigned with highest weights and their corresponding spectrogram. Data belongs to Pat −3 (a, b) and Pat −23 (c, d), and the FP1-F7 channel is presented.

to explore the particularity of these samples. Specifically, We compare the differences between three types of training samples: 1. the preictal samples with high weights; 2. the preictal samples with low weights; 3. The interictal samples with high weights. We choose the samples with $C_i = 100$ as high-weight samples, those with $C_i = 0.1$ as low-weight samples. We understand the differences between these samples by analyzing the mean and variance of 14 features (8 spectral band power, 4 statistical moments and two Hjorth parameters) for the 12 patients used in this study. The comparison results are shown in Fig. 6.

According to Fig. 6, we can find that for the spectral band power, the features of the preictal samples have smaller standard deviation than the interictal samples. When the frequency is less than 8 Hz, there is no significant difference in the mean values of the samples. However, when the frequency is greater than 8 Hz, the mean of spectral band power of the preictal sample is lower than that of the interictal sample, especially when the frequency is greater than 30 Hz, the spectral band power of the preictal sample is concentrated near 0. That is to say, the spectral band power of EEG signals in the Gamma band (30–128 Hz) plays an essential in seizure prediction, which is consistent with the conclusion in [37,38]. Besides, we can find that the low-weight preictal samples are almost distributed between the high-weight preictal samples and the high-weight interictal samples. For the statistical moment, there is no significant difference in the distribution of the mean, skewness, and kurtosis of three types of samples. In terms of the variance, the preictal sample has a more concentrated distribution than the interictal sample. For the Hjorth parameters, the mean mobility of preictal sample is lower than that of interictal, and the mean complexity is higher than that of interictal. Overall, the low-weight preictal samples are almost distributed between the high-weight preictal samples and the high-weight interictal samples.

Furthermore, we select high-weight EEG samples based on the following criteria: the sample is assigned the highest weight during the training of the three models (SVM, CNN, and Transformer). Surprisingly, we find some significant sharp waves and even spike waves in these EEG samples. Sharp wave and spike wave are two abnormal EEG signal waveform closely associated with epileptic seizures, which is characterized as short duration and high amplitude [39,40]. In Fig. 7, we present some preictal EEG samples from one channel and their corresponding spectrogram.

According to Fig. 7, we can infer that our framework could capture EEG samples that are more valuable for epileptic seizure prediction. Since EEG signal is a continuous-time signal, the definition of preictal and interictal inevitably leads to the existence of noisy label. However, to the best of our knowledge, most studies on epileptic seizure prediction by EEG cannot deal with this problem well. To this end, our proposed framework make the model more robust to noisy label by optimizing the weight of each training sample.

Besides, the preictal samples are very scarce compared to interictal samples in real life, leading to another main problem in epileptic seizure prediction: the class imbalance problem. To address this issue, researchers usually adopt two strategies. One strategy is to reduce the number of interictal samples [13], but this will reduce available training samples. Another strategy is to increase the number of preictal samples (e.g., generative method) [6], but this will increase the uncertainty of the model. In our study, we keep the original sample number unchanged and neither generate preictal samples nor delete interictal samples. By assigning weight to each training sample, we avoid the problem caused by sample imbalance and improve the prediction performance.

Finally, we evaluate the computation efficiency of our proposed method (e.g., Transformer-based model). Specifically, Our Transformer model takes about 110 ms in a training step (i.e., one gradient update, including 64 30-second EEG training samples). For a specific patient, there are about 1500 training samples (e.g., patient-1), so it takes about 2.6 s for our model to train an epoch. Moreover, since our model is trained for five epochs in each GA iteration, and the population size is set to 50 and the maximum number of generations is set to 30 in the GA parameters, the total training time on one patient is about 4000 s (a single RTX 3080 GPU). In the testing phase, obtaining predictions for one hour of EEG data takes about only 0.5 s with our trained Transformer model. Therefore, our current model is computationally fast enough to be used for real time detection on the bedside.

5. Conclusion

In this work, we develop a general sample-weighted machine learning framework with a genetic algorithm for patient-specific epileptic seizure prediction. The proposed framework applies to almost all classical classification methods and can significantly improve prediction performance based on these methods. We conduct experiments on 12 pediatric epilepsy patients from the CHB-MIT scalp EEG dataset. After the LOOCV measurement, our framework based on Transformer

achieved an average Sens of 94.6%, an average Fpr of 0.06/h, and an average AUC of 0.939. Compared with the state-of-the-art methods using the same dataset, we achieve the highest Sens, lowest Fpr and highest AUC. Furthermore, our framework is robust to the class imbalance problem in seizure prediction without augmenting or deleting samples.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFF1200600, in part by the National Natural Science Foundation of China (Grants 32271431, 61922075 and 82272070), by Anhui Huami Information Technology Co., Ltd., and by the University Synergy Innovation Program of Anhui Province under Grants GXXT-2019-025

References

[1] W.H. Organization, et al., Epilepsy: A Public Health Imperative, World Health Organization, 2019.
[2] A. Subasi, Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction, Comput. Biol. Med. 37 (2) (2007) 227–244.
[3] A. Zahra, N. Kanwal, N. ur Rehman, S. Ehsan, K.D. McDonald-Maier, Seizure detection from EEG signals using multivariate empirical mode decomposition, Comput. Biol. Med. 88 (2017) 132–141.
[4] K.T. Tapani, P. Nevalainen, S. Vanhatalo, N.J. Stevenson, Validating an SVM-based neonatal seizure detection algorithm for generalizability, non-inferiority and clinical efficacy, Comput. Biol. Med. 145 (2022) 105399.
[5] K.M. Tsiouris, V.C. Pezoulas, M. Zervakis, S. Konitsiotis, D.D. Koutsouris, D.I. Fotiadis, A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals, Comput. Biol. Med. 99 (2018) 24–37.
[6] S.M. Usman, S. Khalid, S. Bashir, A deep learning based ensemble learning method for epileptic seizure prediction, Comput. Biol. Med. 136 (2021) 104710.
[7] Y. Gao, X. Chen, A. Liu, D. Liang, L. Wu, R. Qian, H. Xie, Y. Zhang, Pediatric seizure prediction in scalp EEG using a multi-scale neural network with dilated convolutions, IEEE J. Transl. Eng. Health Med. 10 (2022) 1–9.
[8] X. Yang, J. Zhao, Q. Sun, J. Lu, X. Ma, An effective dual self-attention residual network for seizure prediction, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 1604–1613.
[9] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, F. Fuggetta, Real-time epileptic seizure prediction using AR models and support vector machines, IEEE Trans. Biomed. Eng. 57 (5) (2010) 1124–1132.
[10] M. Bedeeuzzaman, T. Fathima, Y.U. Khan, O. Farooq, Seizure prediction using statistical dispersion measures of intracranial EEG, Biomed. Signal Process. Control 10 (2014) 338–341.
[11] Z. Zhang, K.K. Parhi, Low-complexity seizure prediction from iEEG/sEEG using spectral power and ratios of spectral power, IEEE Trans. Biomed. Circuits Syst. 10 (3) (2015) 693–706.
[12] N.D. Truong, A.D. Nguyen, L. Kuhlmann, M.R. Bonyadi, J. Yang, S. Ippolito, O. Kavehei, Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram, Neural Netw. 105 (2018) 104–111.
[13] A.R. Ozcan, S. Erturk, Seizure prediction in scalp EEG using 3D convolutional neural networks with an image-based approach, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (11) (2019) 2284–2293.
[14] H. Daoud, M.A. Bayoumi, Efficient epileptic seizure prediction based on deep learning, IEEE Trans. Biomed. Circuits Syst. 13 (5) (2019) 804–813.
[15] M. Ma, Y. Cheng, Y. Wang, X. Li, Q. Mao, Z. Zhang, Z. Chen, Y. Zhou, Early prediction of epileptic seizure based on the BNLSTM-CASA model, IEEE Access (2021).
[16] H. Khan, L. Marcuse, M. Fields, K. Swann, B. Yener, Focal onset seizure prediction using convolutional networks, IEEE Trans. Biomed. Eng. 65 (9) (2017) 2109–2118.
[17] Y. Li, Y. Liu, Y.-Z. Guo, X.-F. Liao, B. Hu, T. Yu, Spatio-temporal-spectral hierarchical graph convolutional network with semisupervised active learning for patient-specific seizure prediction, IEEE Trans. Cybern. (2021).
[18] A.S. Zandi, R. Tafreshi, M. Javidan, G.A. Dumont, Predicting epileptic seizures in scalp EEG based on a variational Bayesian Gaussian mixture model of zero-crossing intervals, IEEE Trans. Biomed. Eng. 60 (5) (2013) 1401–1413.
[19] W. Hu, J. Cao, X. Lai, J. Liu, Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks, J. Ambient Intell. Humaniz. Comput. (2019) 1–11.
[20] M. Nasseri, V. Kremen, P. Nejedly, I. Kim, S.-Y. Chang, H.J. Jo, H. Guragain, N. Nelson, E. Patterson, B.K. Sturges, et al., Semi-supervised training data selection improves seizure forecasting in canines with epilepsy, Biomed. Signal Process. Control 57 (2020) 101743.
[21] A.H. Shoeb, Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment (Ph.D. thesis), Massachusetts Institute of Technology, 2009.
[22] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[23] Y. Zhang, Y. Guo, P. Yang, W. Chen, B. Lo, Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network, IEEE J. Biomed. Health Inf. 24 (2) (2019) 465–474.
[24] J.-k. Ge, Y.-H. Qiu, C.-M. Wu, G.-L. Pu, Summary of genetic algorithms research, Appl. Res. Comput. 25 (10) (2008) 2911–2916.
[25] C.-H. Wu, G.-H. Tzeng, Y.-J. Goo, W.-C. Fang, A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy, Expert Syst. Appl. 32 (2) (2007) 397–408.
[26] X. Xiao, M. Yan, S. Basodi, C. Ji, Y. Pan, Efficient hyperparameter optimization in deep learning using a variable length genetic algorithm, 2020, arXiv preprint arXiv:2006.12703.
[27] A.J. Umbarkar, P.D. Sheth, Crossover operators in genetic algorithms: A review, ICTACT J. Soft Comput. 6 (1) (2015).
[28] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[29] B. Direito, C.A. Teixeira, F. Sales, M. Castelo-Branco, A. Dourado, A realistic seizure prediction study based on multiclass SVM, Int. J. Neural Syst. 27 (03) (2017) 1750006.
[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.
[31] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, 2021, arXiv preprint arXiv:2103.14030.
[32] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.
[33] Jazzbin, et al., Geatpy: The genetic and evolutionary algorithm toolbox with high performance in Python, 2020.
[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[35] D. Cho, B. Min, J. Kim, B. Lee, EEG-based prediction of epileptic seizures using phase synchronization elicited from noise-assisted multivariate empirical mode decomposition, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (8) (2016) 1309–1318.
[36] D.E. Snyder, J. Echauz, D.B. Grimes, B. Litt, The statistics of a practical seizure warning system, J. Neural Eng. 5 (4) (2008) 392.
[37] I. Jemal, N. Mezghani, L. Abou-Abbas, A. Mitiche, An interpretable deep learning classifier for epileptic seizure prediction using EEG data, IEEE Access (2022).
[38] Y. Park, L. Luo, K.K. Parhi, T. Netoff, Seizure prediction with spectral power of EEG using cost-sensitive support vector machines, Epilepsia 52 (10) (2011) 1761–1770.
[39] P.D. Emmady, A.C. Anilkumar, EEG abnormal waveforms, 2020.
[40] E. Sitnikova, G. Van Luijtelaar, Electroencephalographic characterization of spike-wave discharges in cortex and thalamus in WAG/Rij rats, Epilepsia 48 (12) (2007) 2296–2311.