



OPEN ACCESS

EDITED BY

Xi Jiang,
University of Electronic Science and
Technology of China, China

REVIEWED BY

Lu Zhang,
University of Texas at Arlington,
United States
Lin Zhao,
University of Georgia, United States

*CORRESPONDENCE

Tinglin Zhang
ztlyoyo@163.com
Limei Zhang
zhanglimei@luc.edu.cn

SPECIALTY SECTION

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

RECEIVED 30 June 2022

ACCEPTED 24 August 2022

PUBLISHED 26 September 2022

CITATION

Wu X, Zhang T, Zhang L and Qiao L
(2022) Epileptic seizure prediction
using successive variational mode
decomposition and transformers deep
learning network.
Front. Neurosci. 16:982541.
doi: 10.3389/fnins.2022.982541

COPYRIGHT

© 2022 Wu, Zhang, Zhang and Qiao.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Epileptic seizure prediction using successive variational mode decomposition and transformers deep learning network

Xiao Wu¹, Tinglin Zhang^{1*}, Limei Zhang^{2*} and Lishan Qiao¹

¹School of Mathematics Science, Liaocheng University, Liaocheng, China, ²School of Computer Science and Technology, Shandong Jianzhu University, Jinan, China

As one of the most common neurological disorders, epilepsy causes great physical and psychological damage to the patients. The long-term recurrent and unprovoked seizures make the prediction necessary. In this paper, a novel approach for epileptic seizure prediction based on successive variational mode decomposition (SVMD) and transformers is proposed. SVMD is extended to multidimensional form for time-frequency analysis of multi-channel signals. It could adaptively extract common band-limited intrinsic modes among all channels on different time scales by solving a variational optimization problem. In the proposed seizure prediction method, data are first decomposed into multiple modes on different time scales by multivariate SVMD, and then, irrelevant modes are removed for preprocessing. Finally, power spectrum of denoised data is input to a pre-trained bidirectional encoder representations from transformers (BERTs) for prediction. The BERT could identify the mode information related to epileptic seizures in time-frequency domain. It shows fair prediction performance on an intracranial EEG dataset with the average sensitivity of 0.86 and FPR of 0.18/h.

KEYWORDS

seizure prediction, successive variational mode decomposition, multiscale time-frequency analysis, BERT, intracranial EEG

Introduction

Epilepsy is one of the most common brain diseases that affect people of all ages. The long-term recurrent and unprovoked seizures could cause great damage to physical and mental health of patients (Schulze-Bonhage and Kühn, 2008). An incoming seizure may be inhibited by some interventions such as medication and electrical or magnetic stimulation of the brain, if it is predicted in advance (Elger, 2001). Therefore, accurate prediction of epileptic seizure could not only significantly improve the quality of life for patients, but also provide a basis for the development of more effective methods of prevention and treatment of epilepsy. There are four phases of brain activity for patients: interictal phase (between seizures), preictal phase (prior to seizure), ictal phase (seizure), and postictal phase (after seizure). If the preictal state could be identified from other

states, an imminent seizure will be predicted. The primary challenge in seizure prediction is the classification of preictal and interictal states (baseline). Electroencephalogram (EEG) is a method commonly used to diagnose epilepsy and evaluate its therapeutic effect (Fisher et al., 2005). In the recent years, an increasing number of literature demonstrates that there is a pattern in preictal EEG (Usman et al., 2019), and prediction of epileptic seizure by EEG is feasible.

Recently, the methods of seizure prediction have focused on time-frequency analysis, non-linear dynamics, and deep learning network. Common time-frequency analysis methods such as wavelet transform and empirical mode decomposition (EMD) have been applied to obtain EEG modes on different scales for seizure detection and prediction (Zahra et al., 2017; Zhang et al., 2018; Hassan et al., 2020; Savadkoobi et al., 2020). However, wavelet transform is not adaptive and the problems of EMD on low robustness and limited mathematical interpretation need to be improved (Dragomiretskiy and Zosso, 2013). Recently proposed variational mode decomposition (VMD) could separate the non-stationary signal into intrinsic modes with narrow band as well as EMD, but the advantages in complete mathematical theory framework and greater robustness (Dragomiretskiy and Zosso, 2013; Lahmiri, 2015) make it applied increasingly in various fields (Upadhyay and Pachori, 2015; Xue et al., 2016; Zhang et al., 2017; Li et al., 2018; Taran and Bajaj, 2018; Wang et al., 2019; Dora and Biswal, 2020; Guo et al., 2020), including epileptic seizure classification (Rout and Biswal, 2020; Peng et al., 2021). In addition to some statistical features in time domain and power spectral estimation in frequency domain, some non-linear dynamical parameters such as fractal dimension (Aarabi and He, 2017), largest Lyapunov exponent (Fei et al., 2017), fuzzy distribution entropy (Zhang et al., 2018), and Hjorth parameters (Teixeira et al., 2014) were also selected as features. Because it was difficult to describe preictal state with just a few features, many tedious feature engineering techniques were involved in the previous studies. However, some features were a lack of reproducibility and reliability (Mormann et al., 2005, 2007; Assi et al., 2017b). Recently, deep learning networks, including convolutional neural networks (CNN) and long short-term memory (LSTM) networks, have attracted most interest in seizure prediction, as their classification performance of preictal state and interictal state is superior to traditional machine learning techniques (Tsiouris et al., 2018; Usman et al., 2019). The latest bidirectional encoder representations from transformer (BERT) (Lee and Toutanova, 2018) is a very attractive deep learning network, which has made a great progress in the field of natural language processing (NLP). It has demonstrated superior performance over LSTM on many NLP tasks. Its application potential in other time series analysis is worth further exploring.

In this paper, a multidimensional extension of SVM is proposed to adaptively extract common intrinsic modes

among all channels on different time scales. After decomposed by multivariate SVM, task-independent modes of the data could be removed for preprocessing or denoising. Then, the power spectrum of denoised iEEG data is input to a pre-trained BERT model for seizure prediction. The proposed seizure prediction method works well on two iEEG datasets.

The work is organized as follows. In Materials and methods, we introduce the information of database used in this paper and the proposed scheme, respectively. In addition, method of performance evaluation and seizure prediction are shown in this section. In Results, we present the experiments' results. In Discussion, we discuss the preprocessing method of SVM and different seizure prediction methods used on the iEEG dataset. Finally, we conclude this paper in Conclusion.

Materials and methods

EEG dataset

The first dataset was obtained from Kaggle American Epilepsy Society Seizure Prediction Challenge (<https://www.kaggle.com/competitions/seizure-prediction/>). It is comprised of long-term intracranial EEG (iEEG) recordings from five dogs and two patients. Another dataset used in this study is comprised of continuous iEEG recordings from three dogs (Dog_6, Dog_7, and Dog_8), which could be obtained from NIH-sponsored International Epilepsy Electrophysiology portal (<https://www.ieeg.org>). The Canine iEEG data were sampled from 16 or 15 electrodes at 400 Hz. iEEG data of two patients were sampled at 5,000 Hz and recorded with 15 (Patient_1) and 24 (Patient_2) implanted electrodes, respectively. The type of seizures is focal epilepsy. More details were described in reference (Brinkmann et al., 2016). In this dataset, 1 h before seizure with a 5-min horizon (i.e., 66–5 min before seizure onset) was chosen as preictal phase (Brinkmann et al., 2016; Assi et al., 2017a; Gagliano et al., 2019; Nejedly et al., 2019; Yu et al., 2021). Each consecutive interictal sequence lasted for 1 h, which were randomly chosen from iEEG recordings more than 1 week (dogs) and 4 h (patients) before or after any seizure. The iEEG portal dataset is comprised of continuous iEEG recordings, which are all labeled. The Kaggle dataset consists of training data and testing data. Each labeled iEEG sequence of training data lasts for 1 h, and unlabeled testing data are 10-min iEEG segment (the contest website does not have labels for test data, and the score could only be obtained by uploading the predicted results of all test data to the website). The description of the data used in this work is shown in Table 1.

TABLE 1 Description of the Kaggle dataset.

Participant	No. of channels	No. of seizures	Interictal hours	No. of testing segments
Dog_1	16	4	80	502
Dog_2	16	7	83	1,000
Dog_3	16	12	240	907
Dog_4	16	16	134	990
Dog_5	15	5	75	191
Dog_6	16	41	998	-
Dog_7	16	38	936	-
Dog_8	16	15	286	-
Patient_1	15	2	8.3	195
Patient_2	24	3	7	150

Preprocessing methods

The multidimensional extension of successive variational mode decomposition (SVMD) is proposed for time-frequency analysis of non-stationary multi-channel signals in this section. Multivariate SVMD is used to remove irrelevant modes for denoising in the presented seizure prediction method.

Successive variational mode decomposition is established under the similar theoretical framework as VMD, which requires each extracted mode to be compact around its center frequency and original data to be reconstructed by all modes. However, different from VMD, SVMD could successively decompose each intrinsic mode from a signal without specifying the number of modes in advance. Therefore, there is no complex multi-parameter optimization problem for SVMD. Details of the algorithm could be found in the reference (Nazari and Sakhaei, 2020). As there is also a lot of demand for analyzing multi-channel signals in real-world applications, a simple multivariate extension of SVMD is presented.

Multidimensional SVMD aimed to adaptively extract common intrinsic modes $\mathbf{u}_i(t)$ with limited bandwidth from multivariate signal $\mathbf{f}(t)$ containing C channels, i.e., $\mathbf{f}(t)=[f_1(t), f_2(t), \dots, f_C(t)]$.

$$\mathbf{f}(t) = \sum_{i=1}^L \mathbf{u}_i(t) \tag{1}$$

where $\mathbf{u}_i(t) = [u_{i1}(t), u_{i2}(t), \dots, u_{iC}(t)]$, C is the number of channels and L is the number of common modes decomposed by multivariate SVMD.

It is noteworthy that intrinsic modes on the l th scale $u_l(t)$ are set to the same central frequency ω_l in our model for the purpose of getting common modes of C channels on the same time scale. According to the definition of intrinsic mode function,

$u_l(t)$ should be limited bandwidth signals, which is the central assumption for mode separation in SVMD. Therefore, the average bandwidth of all modes on the l th time scale should be minimized. Equivalently, the total bandwidth of C modes forms cost function L_1 in multivariate SVMD optimization problem and is given by

$$L_1 = \sum_{k=1}^C \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_{lk}(t) \right] e^{-j\omega_l t} \right\|_2^2 \tag{2}$$

To obtain the complete modes on the l th scale and avoid mode mixing with other scales, neither the previously extracted $l - 1$ modes nor undecomposed part $f_{uk}(t)$ of the k th channel ($k = 1, 2, \dots, C$) should contain any information of the l th mode. Meanwhile, there should be no spectral overlap between the l th mode and previously decomposed $l - 1$ modes. Accordingly, criteria L_2 , namely, the total frequency response of residual signals ($\{u_{ik}(t)\}_{i=1}^{l-1}$ and $f_{uk}(t)$) of all channels after passing through the filter $\hat{\beta}_l(\omega)$ (frequency response of the l th filter), should be minimized. Furthermore, for the k th channel, the total energy of filtered $u_{lk}(t)$ by each filter $\hat{\beta}_i(\omega)$ ($i = 1, 2, \dots, l - 1$) requires as less as possible. This constraint is shown in the cost function L_3 .

$$L_2 = \sum_{k=1}^C \left\| \hat{\beta}_l(t) * \left(f_{uk}(t) + \sum_{i=1}^{l-1} u_{ik}(t) \right) \right\|_2^2 \tag{3}$$

$$L_3 = \sum_{k=1}^C \sum_{i=1}^{l-1} \left\| \hat{\beta}_i(t) * u_{lk}(t) \right\|_2^2 \tag{4}$$

$$\hat{\beta}_i(\omega) = \frac{1}{\alpha(\omega - \omega_i)^2} \quad i = 1, 2, \dots, L \tag{5}$$

The constrained variational optimization problem for multivariate SVMD is represented as follows:

$$\left. \begin{aligned} \min_{u_{lk}, \omega_l, f_{uk}(t)} & \alpha L_1 + L_2 + L_3 \\ \text{s.t. } & u_{lk}(t) + f_{uk}(t) + \sum_{i=1}^{l-1} u_{ik}(t) = f_k(t), \\ & k = 1, 2, \dots, C \end{aligned} \right\} \tag{6}$$

The augmented Lagrange function shown in (7) is used to transform this problem into unconstrained optimization problem, which could be solved iteratively by ADMM approach (Bertsekas, 1982)

$$\left. \begin{aligned} L(u_{lk}, \omega_l, \lambda_k) &= \alpha L_1 + L_2 + L_3 \\ &+ \sum_{k=1}^C \left\| f_k(t) - \left(u_{lk}(t) + f_{uk}(t) + \sum_{i=1}^{l-1} u_{ik}(t) \right) \right\|_2^2 \\ &+ \sum_{k=1}^C \left\langle \lambda_k(t), f_k(t) - \left(u_{lk}(t) + f_{uk}(t) + \sum_{i=1}^{l-1} u_{ik}(t) \right) \right\rangle \end{aligned} \right\} \tag{7}$$

The first subproblem is focused on updating the modes u_{lk} iteratively by channel. The $(\mathbf{n} + 1)$ th iteration of the k th channel

could be rewritten as the following equivalent problem, which is actually reduced to a univariate mode update problem in original SVMd.

$$\begin{aligned} \hat{u}_{lk}^{n+1}(t) = \arg \min_{lk} \left\{ \alpha \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_{lk}(t) \right] e^{-j\omega_l t} \right\|_2^2 \right. \\ + \left\| \beta_l(t) * \left(f_{uk}(t) + \sum_{i=1}^{l-1} u_{ik}(t) \right) \right\|_2^2 + \sum_{i=1}^{l-1} \left\| \beta_i(t) * u_{lk}(t) \right\|_2^2 \\ + \sum_{k=1}^C \left\| f_k(t) - \left(u_{lk}(t) + f_{uk}(t) + \sum_{i=1}^{l-1} u_{ik}(t) \right) \right\|_2^2 \\ \left. + \sum_{k=1}^C \left\langle \lambda_k(t), f_k(t) - \left(u_{lk}(t) + f_{uk}(t) + \sum_{i=1}^{l-1} u_{ik}(t) \right) \right\rangle \right\} \end{aligned} \tag{8}$$

Therefore, as same as SVMd, it could be solved in spectral domain based on the Parseval's equality. u_{lk} is updated by (9). Details could be found in reference (Nazari and Sakhaei, 2020).

$$\hat{u}_{lk}^{n+1}(\omega) = \frac{\hat{f}_k(\omega) + \alpha^2 (\omega - \omega_l^n)^4 \hat{u}_{lk}^n(\omega) + \frac{\hat{\lambda}_k}{2}}{\left[1 + \alpha^2 (\omega - \omega_l^n)^4 \right] \left[1 + 2\alpha (\omega - \omega_l^n)^2 + \sum_{i=1}^{l-1} \frac{1}{\alpha^2 (\omega - \omega_i)^4} \right]} \tag{9}$$

The second subproblem is related to updating the center frequency ω_l . The $(n + 1)$ th iteration of each channel is the minimization problem shown in (10), which could be solved with the method and equation applied in SVMd. According to the principle of linear superposition, ω_l could be updated by Equation (11).

$$\omega_l^{n+1} = \arg_{\omega_l} \min \{ \alpha L_1 + L_2 \} \tag{10}$$

$$\omega_l^{n+1} = \frac{\sum_{k=1}^C \int_0^\infty \omega \left| \hat{u}_{lk}^{n+1}(\omega) \right|^2 d\omega}{\sum_{k=1}^C \int_0^\infty \left| \hat{u}_{lk}^{n+1}(\omega) \right|^2 d\omega} \tag{11}$$

The updating equation of Lagrange multiplier λ is the same as SVMd, as long as replace \hat{u}_i by \hat{u}_{ik} .

The result of decomposition is affected by the penalty factor α , which determines the bandwidth of intrinsic modes (Dragomiretskiy and Zosso, 2013; Nazari and Sakhaei, 2020). Furthermore, the optimal α differs obviously when decomposing different types of signals. Consequently, a heuristic method similar to SVMd is introduced to obviate optimization of α . In the iteration of extracting modes of the l th scale, α is set to grow exponentially from a small value α_{\min} to a maximum allowable value α_{\max} , which is actually a process of finding the strongest modes in the residual signals from coarse to fine tuning.

The algorithm terminates search until total energy of all the l th modes is less than the given threshold ε_2 ; namely, the modes extracted could be regarded as noise. Finally, all the obtained modes are sorted by their center frequency from low to high.

The complete algorithm for multivariate SVMd is described in Table 2.

Classification and evaluation

The human iEEG data were down-sampled to 500 Hz to be comparable to canine iEEG. To reduce computational burden of SVMd, both preictal and interictal iEEG data were first divided into 2-s clips without overlap. Then, all iEEG clips were decomposed by multivariate SVMd. Irrelevant modes of raw iEEG data were removed and the remaining ones were added up for reconstruction. Subsequently, the reconstructed data were concatenated into a new time series in chronological order. The denoised iEEG data were split into 30-s-long samples with 28-s overlap. To use modal information in time-frequency domain for prediction, power spectrum was extracted by the short-time Fourier transform (STFT). Each iEEG sample was segmented by a 1-s time window with 75% overlap to compute

the power spectrum by the function *spectrum* in MATLAB. Only the power spectrum from 0 to 140 Hz is selected in this study, and the average of the power per 2 Hz is calculated as the final spectrum. The power spectrum of iEEG samples was input to a deep learning network based on BERT for seizure prediction. To compare the performance of preprocessing, the power spectrum of raw iEEG was also input to BERT for classification.

BERT model architecture

The classic BERT's model architecture is based on a multi-layer bidirectional transformer encoder (Vaswani et al., 2017) and it uses bidirectional self-attention mechanism. After being pre-trained with two unsupervised tasks, all parameters of BERT could be fine-tuned using labeled data from the downstream tasks. The code and pre-trained models are available at <https://github.com/matlab-deep-learning/transformer-models>. In this study, the classification of preictal and interictal iEEG could be considered as a downstream task to finetune a pre-trained BERT model with an additional output layer. Our model architecture consists of input layer, encoder layer (transformer blocks), fully connected layer, and Softmax classification layer, as shown in Figure 1.

It is worth noting that BERT is originally designed to solve NLP tasks, and the input representation is a token sequence transformed from a sentence (Wu et al., 2016). However, the input data are essentially a digital time series, which is unnecessary to convert to tokens and then use word embedding in the input layer. Therefore, a more suitable embedding method

TABLE 2 The complete algorithm of multivariate SVM.

The algorithm of SVM

Initialize: $l \leftarrow 0, \varepsilon_1 \leftarrow 10^{-6}, \varepsilon_2 \leftarrow 10^{-3}$

repeat

$l \leftarrow l + 1$

Set $\hat{u}_{lk}^1(\omega), \hat{\lambda}_k^1, \omega_l^1, n \leftarrow 0, m \leftarrow 0, \alpha_1 \leftarrow \alpha_{\min}$

repeat

$m \leftarrow m + 1$

repeat

$n \leftarrow n + 1$

for $k = 1 : C$ **do**

Update u_{lk} for all $\omega \geq 0$:

$$\hat{u}_{lk}^{n+1}(\omega) = \frac{\hat{f}_k(\omega) + \alpha^2(\omega - \omega_l^n)^4 u_{lk}^n(\omega) + \frac{\hat{\lambda}_k^n}{2}}{\left[1 + \alpha^2(\omega - \omega_l^n)^4\right] \left[1 + 2\alpha(\omega - \omega_l^n)^2 + \sum_{i=1}^{l-1} \frac{1}{\alpha^2(\omega - \omega_i)^4}\right]}$$

end for

for $k = 1 : C$ **do**

Update ω_l

$$\omega_l^{n+1} = \frac{\sum_{k=1}^C \int_0^\infty \omega |u_{lk}^{n+1}(\omega)|^2 d\omega}{\sum_{k=1}^C \int_0^\infty |u_{lk}^{n+1}(\omega)|^2 d\omega}$$

end for

for $k = 1 : C$ **do**

Dual Ascent for all $\omega \geq 0$:

$$\hat{\lambda}_k^{n+1} = \hat{\lambda}_k^n + \tau [\hat{f}_k(\omega) - (\hat{u}_{lk}^{n+1}(\omega) + \frac{\alpha^2(\omega - \omega_l^n)^4 (\hat{f}_k(\omega) - \hat{u}_{lk}^{n+1}(\omega) - \sum_{i=1}^{l-1} \hat{u}_{ik}(\omega) + \frac{\hat{\lambda}_k^n}{2}) - \sum_{i=1}^{l-1} \hat{u}_{ik}(\omega)}{1 + \alpha^2(\omega - \omega_l^n)^4} + \sum_{i=1}^{l-1} \hat{u}_{ik}(\omega))]$$

end for

Until convergence: $\frac{\| \hat{u}_{lk}^{n+1}(\omega) - \hat{u}_{lk}^n(\omega) \|_2}{\| \hat{u}_{lk}^n(\omega) \|_2} < \varepsilon_1, k = 1, 2, \dots, C$

Set $\hat{\lambda}_k^1, \hat{u}_{lk}^1(\omega) \leftarrow \hat{u}_{lk}^{n+1}(\omega), \omega_l^1 \leftarrow \omega_l^{n+1}, \alpha_m \leftarrow \alpha_{\min} + e^m, n \leftarrow 0$

Until $\alpha_m \leq \alpha_{\max}$

Until $\sum_{k=1}^C \left[\frac{1}{T} \| u_{lk}(t) \|_2^2 \right] < \varepsilon_2$

for digital sequence needs to be designed. The input data of all channels are concatenated and weighted as a kind of data embedding [refer to Equations (12) and (13)], which could be considered as a kind of data fusion.

$$x_j = \begin{bmatrix} \tilde{x}_{1j} \\ \tilde{x}_{2j} \\ \vdots \\ \tilde{x}_{Ncj} \end{bmatrix}, X = [x_1, x_2, \dots, x_N], j = 1, 2, \dots, N \quad (12)$$

$$E_d = X \odot W \quad (13)$$

where \tilde{x}_{ij} is power spectrum of the i th channel in the j th time window (each 1-s time window is set as a time step, and N is the number of time steps), and all channels are cascaded to construct a $(N_c \times N_p) \times 1$ vector (N_c is the number of channels, and N_p is the number of spectrum frequencies). The Hadamard product of the power spectrum X and weight matrix W is the data embedding E_d .

In the input layer, X is converted to a matrix E by summing the position embedding E_p (the embedding method is the

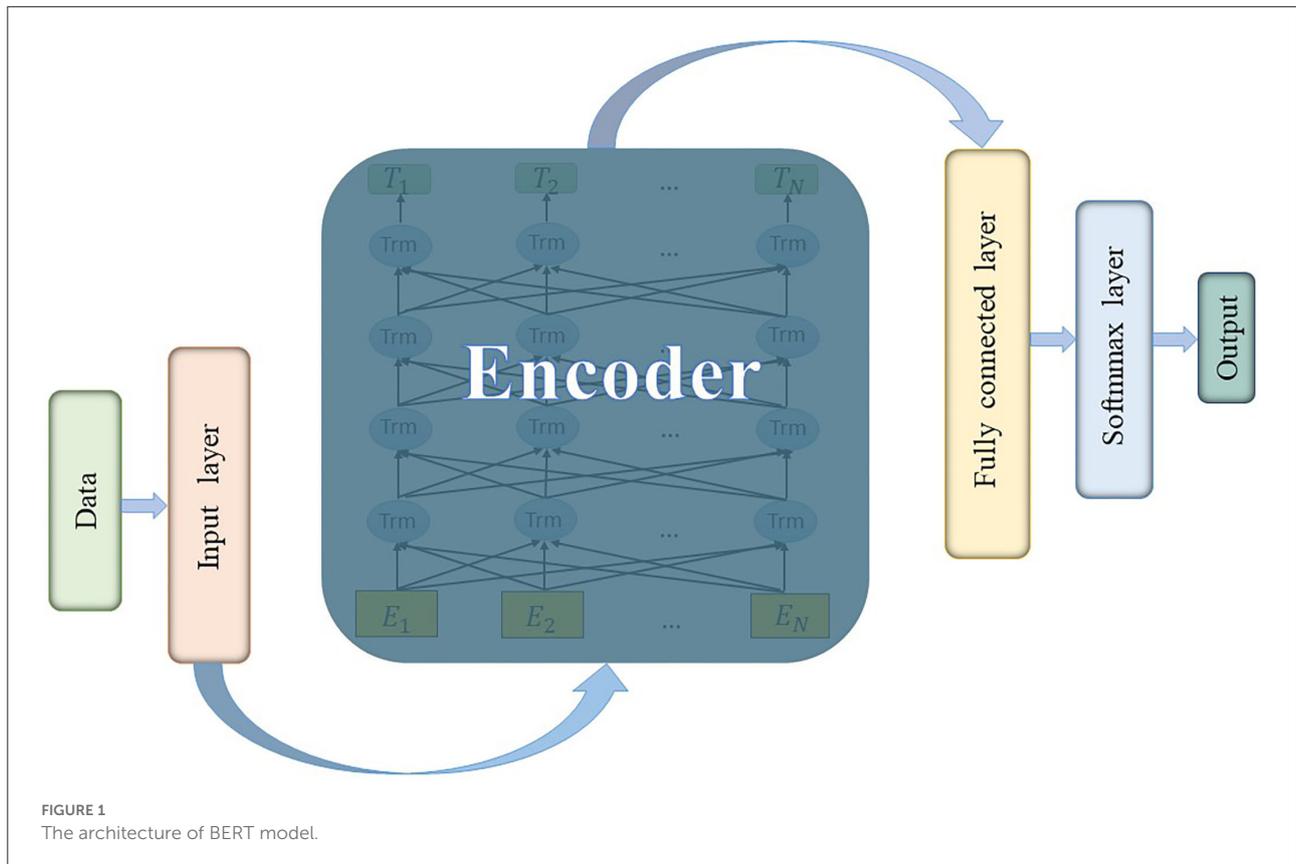
same as BERT) and data embedding E_d . Dynamic coding is applied and all weights are automatically learned by training. Weights are first initialized as random numbers that obey normal distribution. After embedding and normalization, the $(N_c \times N_p) \times N$ (i.e., number of features \times number of time steps) matrix is input to the encoder layer.

In the encoder layer, the number of layers (i.e., transformer blocks) is 12 and the hidden size is 768. The number of self-attention heads is 12. Batch size is set to 32 and the number of epochs in training loop is 10. The BERT model is built with MATLAB R2022a.

Evaluation

To test the predictive ability of this approach for unknown seizures, limited seizures were used for training, whereas the remaining ones were for testing. All the data of iEEG portal dataset and labeled training data in Kaggle dataset could be used. Because there were relatively few seizures for each subject in the training data, a leave-one-out cross-validation method was applied. Namely, $M-1$ seizures were used for training and one for validation if there were M seizures for a subject. The amount of interictal iEEG is much larger than preictal iEEG. Therefore, to avoid the problem of class imbalance, a number of preictal and interictal iEEG sequences were the same in the training set. Each interictal iEEG sequence was randomly selected from the dataset. All remaining interictal sequences were used for validation. We run ten trials and train 10 models for each subject (refer to Lian et al., 2020). The average performance was considered as final prediction performance when using training data. We could also use unlabeled testing data of Kaggle dataset to test the prediction method. Similarly, we trained multiple models to avoid the problem of class imbalance. For each subject, all preictal iEEG and the same amount of randomly selected interictal iEEG were used to train the model, and we run 10 trials. A testing segment in Kaggle dataset would be predicted as preictal iEEG, if more than 6 models identified it as preictal. No labels are given for the testing data in Kaggle dataset, but the score (an index related to classification accuracy that used by the organizer) could be calculated on the competition website. Therefore, the score we achieved on testing data is a key indicator of predictor performance.

To improve the reliability of the prediction, a prediction window of 10 min was applied. According to experiential knowledge [refer to (Truong et al., 2018; Wei et al., 2019)], if more than 60% of EEG samples during 10-min continuous recordings are identified as preictal, the warning alarm would be raised. To evaluate the performance of the prediction method, there are four commonly used measures including sensitivity, false prediction rate (FPR), seizure occurrence period (SOP), and seizure prediction horizon (SPH). Sensitivity is the number of correctly predicted seizures divided by the total number of seizures. FPR is defined as the number of false alarms per hour.



SPH is a predefined interval between the first alarm and the incoming seizure, which is also a period reserved for patients to take intervention measures. SOP is the period during which a seizure is expected to occur (Maiwald et al., 2004). Therefore, for a correct prediction, seizure would not occur during the SPH and must occur within the SOP. There are no common criteria for the length of SOP and SPH, but the SPH should be long enough for intervention and the SOP should not be too long in case of patient’s anxiety. Based on prior knowledge of other studies, we use the SPH of 30 min and the SOP of 20 min here.

To evaluate the statistical significance of the seizure prediction performance, a random predictor is used for comparison. For a given FPR, the probability to raise an alarm during the SOP can be approximated as follows: (Schelter et al., 2006).

$$P \approx 1 - e^{-FPR \cdot SOP} \tag{14}$$

Therefore, the probability of predicting at least m of M independent seizures by chance is given by

$$p = \sum_{i \geq m} \binom{M}{i} P^i (1 - P)^{M-i} \tag{15}$$

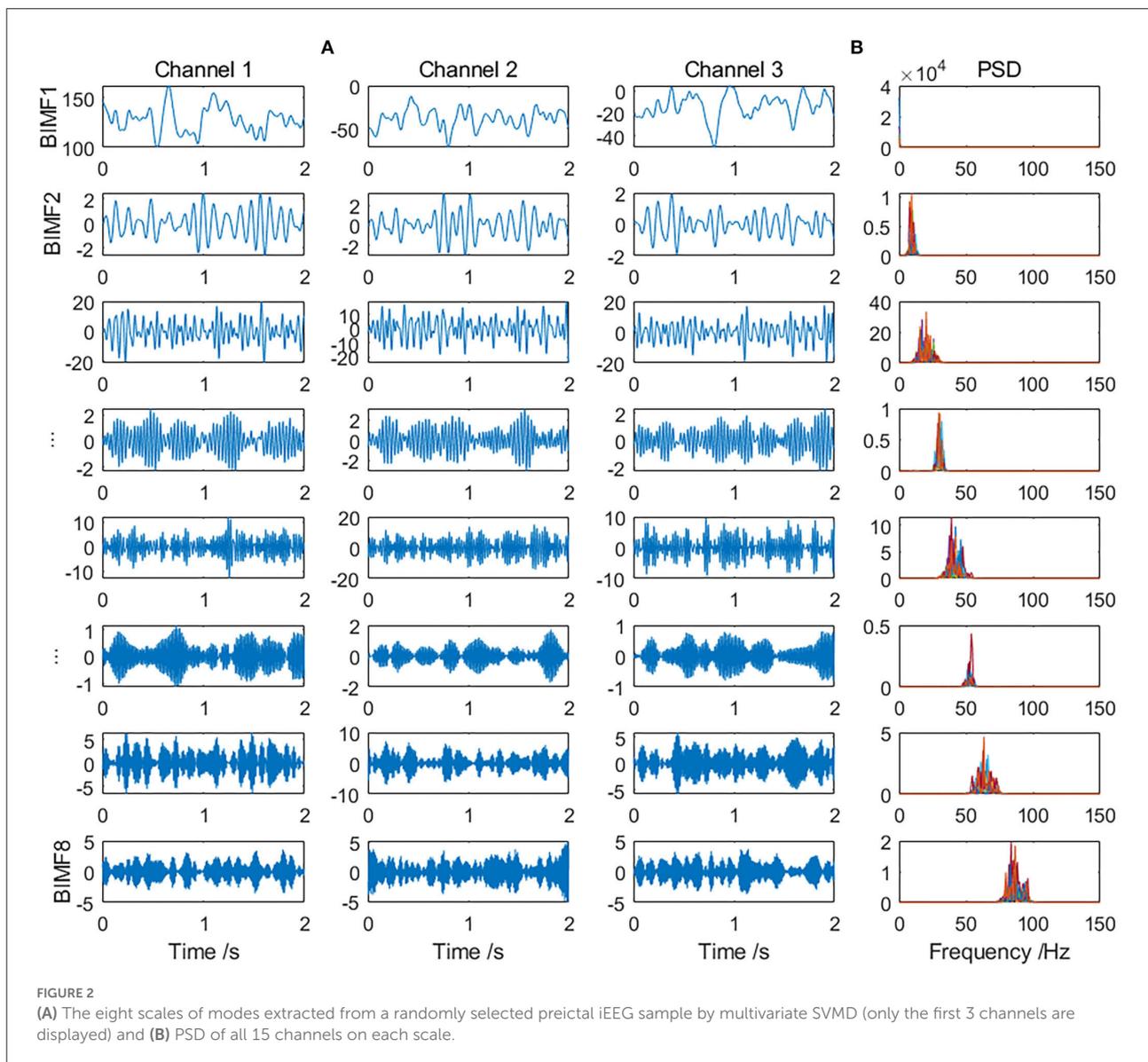
For each patient, p is calculated using the FPR and the number of correctly predicted seizures m . If p is

< 0.05 , the prediction method is considered significantly better than a random predictor at a significance level of 0.05.

Results

Preprocessing results

Multivariate SVM was applied for preprocessing. The range of parameter α in SVM was set to [200, 800] for canine iEEG and [200, 2000] for patient iEEG. The eight scales of common intrinsic modes extracted from a randomly selected 2-s preictal iEEG of Dog_5 are shown in Figure 2A (only the first 3 channels are displayed in the figure due to space limitations). Corresponding power spectrum density (PSD) of all 15 channels on each scale is indicated in Figure 2B. The frequency bands of modes on the same scale were similar, which illustrated the mode-alignment ability of multivariate SVM across multiple channels. The modes which could obtain the highest classification accuracy were considered as effective modes and the others were irrelevant. Irrelevant modes were removed and the remaining modes were added up for reconstruction. It could be considered a kind of denoising.



Prediction results

The power spectrum of reconstructed data was input to BERT for deep learning and classification. It is shown in Table 3 that this prediction algorithm achieves mean sensitivity of 0.86 and the average FPR of 0.18/h. The p -value indicated that the prediction method was significantly superior to a random predictor for all subjects. The mean score of our method on testing data of Kaggle dataset was 0.84125, which was about 0.03 below the competition leader of 0.87154. The power spectrum of raw iEEG was also input to BERT for classification, to compare the preprocessing algorithms. The mean score on testing data was 0.69153, which is, however, much lower than the proposed method.

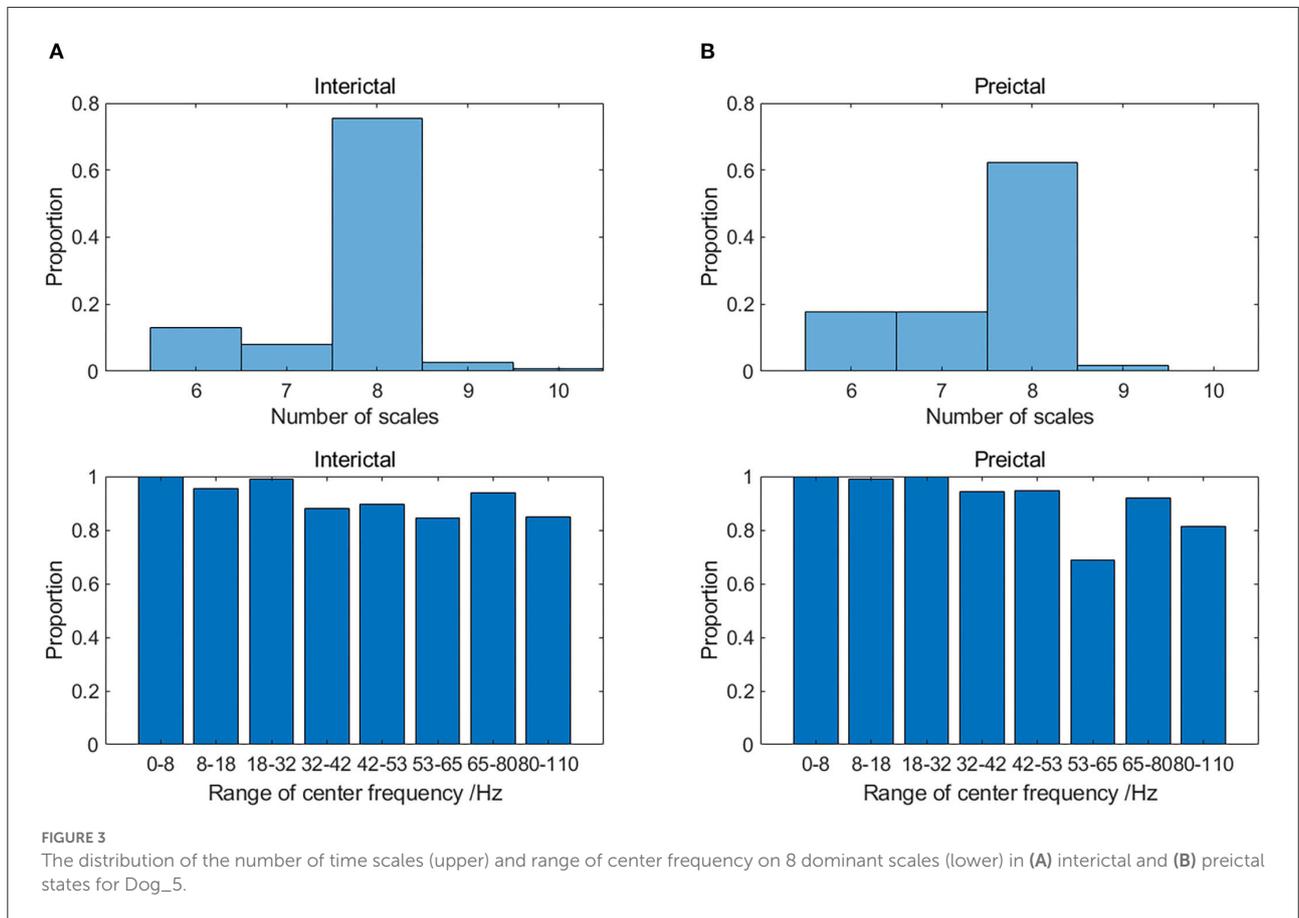
Discussion

Multivariate SVM inherits the advantages of SVM including less parameters, resistance to mode mixing and adaptability. Meanwhile, it could be seen from Figure 3 that the frequency bands of modes on the same scale were similar, which illustrated the mode-alignment ability of multivariate SVM across multiple channels. Furthermore, modes in different scales were in distinctive frequency bands, which demonstrated that SVM might have filter bank property, which is not the focus of this study, but could be further proofed in the future.

The number of the intrinsic modes extracted by multivariate SVM for some samples was not consistent, because of wideband iEEG signals with the effect of ocular artifacts, electromyogram, and other background noise. Take the data

TABLE 3 The performance of the proposed method on 10 subjects.

Participant	No. of seizures	Center frequency of removed modes (Hz)	Sensitivity	FPR (/h)	p
Dog_1	4	$8 < \omega < 15$	0.65	0.25	0.0019
Dog_2	7	$12 < \omega < 20$	0.87	0.06	<0.001
Dog_3	12	$12 < \omega < 20$	0.92	0.23	<0.001
Dog_4	16	$\omega < 60$	0.94	0.07	<0.001
Dog_5	5	$\omega < 55$	0.90	0.16	<0.001
Dog_6	41	$\omega < 30$	0.90	0.15	<0.001
Dog_7	38	$\omega < 30$	0.86	0.18	<0.001
Dog_8	15	$\omega < 20$	0.88	0.09	<0.001
Patient_1	2	$\omega < 55$	1	0.36	0.0128
Patient_2	3	$\omega < 30$	0.67	0.25	0.0182
Mean			0.86	0.18	



of Dog_5 for example, the distribution of the number of time scales and center frequency on 8 dominant scales are displayed in Figure 3. Most of the interictal samples (75.6%) were decomposed into 8 scales of band-limited intrinsic mode function (BIMF), whereas there was less consistency for preictal samples on the number of modes. For both states, the center frequencies of 8 dominant time scales were in the range of [0, 8],

[8, 18], [18, 32], [32, 42], [42, 53], [53, 65], [65, 80], and [80, 110] respectively. However, the proportion of samples containing certain time scales of modes (modes in high gamma band) is significantly reduced in the preictal state, as shown in Figure 3. The reason might be that some modes were interfered by the new modes generated by an impending epileptic seizure, which needs to be proved by exploiting more physiological evidence.

It can be seen from Table 3 that the difference of preictal and interictal modes shows specificity among all subjects. There is a certain consistency for Dog_1, Dog_2, and Dog_3, because all the irrelevant modes are in alpha and beta bands. However, modes that are associated with seizures are in gamma band for other subjects. Therefore, the seizure prediction method is patient-dependent due to the specificity of patients.

As summarized in the reference (Usman et al., 2019), support vector machine (SVM) was widely used in studies before 2019 with good predictive performance. LSTM was the most commonly used model among deep learning models to solve NLP problems and other time series pattern recognition before the emergence of BERT. Therefore, we compare the prediction ability of these two classifiers with BERT. SVM with Gaussian radial basis function (RBF) kernel is used by reference to the literature (Bandarabadi et al., 2015; Xiang et al., 2015; Sharif and Jafari, 2017). The LSTM network is consisted of a sequence input layer, a LSTM layer, a dropout layer, a fully connected layer using the “relu” activation function, and a classification layer using the “softmax” activation. The size of input layer is dependent on the number of power spectrum features. The dropout probability is 0.5. The number of memory units on the LSTM layer is set to 128 (Tsiouris et al., 2018). Although the mean sensitivity of SVM could reach 0.83, the score on testing data is only 0.65839. The

prediction performance of both LSTM and BERT on testing data is much better than that of SVM, which may due to the stronger learning ability of the two deep learning models for temporal information. Moreover, BERT could achieve better prediction results than LSTM, as shown in Table 4. It illustrates that BERT shows better performance in epileptic seizure prediction than LSTM.

As is shown in Table 5, for the canine iEEG dataset, the sensitivity of this method is higher than that of other methods, and the FPR of 0.20 is relatively low. It represents the high prediction performance of this method. The mean score achieved on testing data was 0.84125 with preprocessed data, while only 0.69153 with raw iEEG data, which illustrates that SVM could screen out valid modes for seizure prediction. Meanwhile, it proved again that the difference between preictal state and interictal state of brain exists in the power spectrum of iEEG in time-frequency domain. The self-attention learning mechanism of BERT could extract the information effectively. Although the result is comparable with the work of Assi et al. there were only three subjects and complex feature extraction, and feature selection and channel selection were used to predict seizures in that study. However, there are 10 subjects in the two datasets we used, and our method is relatively simple. Only the power spectrum of denoised iEEG by SVM is used as features for prediction.

The mean score obtained by the first team is 0.87154 in the Kaggle competition. Although it is about 0.03 higher than our method, their result is based on the numerous features and elaborate feature selection. The features include energy in different frequency bands, correlation of energy between channels, square root of each feature, and so on (they only briefly introduced the features in the following websites: <https://www.kaggle.com/competitions/seizure-prediction/discussion/11024>). However, features are learned adaptively in our method. The score we achieved indicates that the proposed approach could be a candidate or auxiliary method for seizure prediction.

TABLE 4 The prediction performance of three classifiers (SVM, LSTM, and BERT).

	Training data		Testing data
	Sensitivity	FPR	score
SVM	0.83	0.24	0.65839
LSTM	0.84	0.21	0.77930
BERT	0.86	0.18	0.84125

The sensitivity and FPR in the table are the average of 10 subjects.

TABLE 5 Comparison of seizures prediction methods using iEEG dataset.

Authors	No. of subjects	Features	Classifier	Sensitivity	FPR
Assi et al. (2017a)	3	Spectral band power, Hjorth mobility and complexity, spectral edge frequency and power, and decorrelation time	SVM	0.85	-
Truong et al. (2018)	7	Short-time Fourier transform	CNN	0.75	0.21
Nejedly et al. (2019)	4	Raw iEEG and spectrogram images	CNN	0.79	-
Gagliano et al. (2019)	3	Higher-order spectral features	LSTM	0.78	-
Yu et al. (2021)	7	Autoregressive (AR) model coefficients and Laguerre–Volterra AR model coefficients	Sparse lasso logistic regression classifier	0.78	-
This work	10	Power spectrum of reconstructed data by multivariate SVM	BERT	0.86	0.18

Conclusion

In this paper, we proposed a seizure prediction method based on SVM and BERT. The simple extension of SVM could decompose multivariate data into its common inherent modes on different scales. The iEEG signals were preprocessed by removing irrelevant modes after decomposition by SVM. The prediction score on Kaggle competition indicated that BERT could learn the difference of preictal and interictal state in time-frequency domain using the self-attention learning mechanism. Therefore, it could be a candidate method for seizure prediction.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

XW and TZ designed the study. XW downloaded and analyzed the data, performed experiments, and drafted the manuscript. LZ, LQ, and TZ revised the manuscript. All authors read and approved the final manuscript.

References

- Aarabi, A., and He, B. (2017). Seizure prediction in patients with focal hippocampal epilepsy. *Clin. Neurophysiol.* 128, 1299–1307. doi: 10.1016/j.clinph.2017.04.026
- Assi, E. B., Nguyen, D. K., Rihana, S., and Sawan, M. (2017a). A functional-genetic scheme for seizure forecasting in canine epilepsy. *IEEE Trans. Biomed. Eng.* 65, 1339–1348. doi: 10.1109/TBME.2017.2752081
- Assi, E. B., Nguyen, D. K., Rihana, S., and Sawan, M. (2017b). Towards accurate prediction of epileptic seizures: a review. *Biomed. Signal Process. Control.* 34, 144–157. doi: 10.1016/j.bspc.2017.02.001
- Bandarabadi, M., Teixeira, C. A., Rasekhi, J., and Dourado, A. (2015). Epileptic seizure prediction using relative spectral power features. *Clin. Neurophysiol.* 126, 237–248. doi: 10.1016/j.clinph.2014.05.022
- Bertsekas, D. P. (1982). *Constrained Optimization and Lagrange Multiplier Methods (Constrained Optimization and Lagrange Multiplier Methods)* (Athena Scientific). Available online at: <https://www.amazon.com/Constrained-Optimization-Lagrange-Multiplier-computation/dp/1886529043>
- Brinkmann, B. H., Wagenaar, J., Abbot, D., Adkins, P., Bosshard, S. C., Chen, M., et al. (2016). Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. *Brain* 139, 1713–1722. doi: 10.1093/brain/aww045
- Dora, C., and Biswal, P. K. (2020). An improved algorithm for efficient ocular artifact suppression from frontal EEG electrodes using VMD. *Biocybern. Biomed. Eng.* 40, 148–161. doi: 10.1016/j.bbe.2019.03.002
- Dragomiretskiy, K., and Zosso, D. (2013). Variational mode decomposition. *IEEE Trans. Signal Process.* 62, 531–544. doi: 10.1109/TSP.2013.2288675
- Elger, C. E. (2001). Future trends in epileptology. *Curr. Opin. Neurol.* 14, 185–186. doi: 10.1097/00019052-200104000-00008
- Fei, K., Wang, W., Yang, Q., and Tang, S. (2017). Chaos feature study in fractional fourier domain for preictal prediction of epileptic seizure. *Neurocomputing* 249, 290–298. doi: 10.1016/j.neucom.2017.04.019

Funding

This work was partly supported by the National Natural Science Foundation of China (Nos. 61976110, 62176112, and 11931008), the Natural Science Foundation of Shandong Province (No. ZR202102270451), and The Open Project of Liaocheng University Animal Husbandry Discipline (No. 319312101-01).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Fisher, R. S., Boas, W. V. E., Blume, W., Elger, C., Genton, P., Lee, P., et al. (2005). Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE). *Epilepsia* 46, 470–472. doi: 10.1111/j.0013-9580.2005.66104.x

- Gagliano, L., Bou Assi, E., Nguyen, D. K., and Sawan, M. (2019). Bispectrum and recurrent neural networks: improved classification of interictal and preictal states. *Sci. Rep.* 9, 1–9. doi: 10.1038/s41598-019-52152-2

- Guo, Z., Liu, M., Wang, Y., and Qin, H. (2020). A new fault diagnosis classifier for rolling bearing united multi-scale permutation entropy optimize VMD and cuckoo search SVM. *IEEE Access.* 8, 153610–153629. doi: 10.1109/ACCESS.2020.3018320

- Hassan, A. R., Subasi, A., and Zhang, Y. (2020). Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise. *Knowl. Based Syst.* 191, 105333. doi: 10.1016/j.knsys.2019.105333

- Lahmiri, S. (2015). Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices. *IEEE Syst. J.* 11, 1907–1910. doi: 10.1109/JSYST.2015.2487339

- Lee, J. D. M. C. K., and Toutanova, K. (2018). *Pre-Training of Deep Bidirectional Transformers for Language Understanding*. Available online at: https://arxiv.org/abs/1810.04805?_hsrc=p2ANqtz--n7PUYWznWMz86GjLjA-LJx8Oyt7ZwXl1kdSGc1BMUWkEnTdj39QK1wTM4ynwo4sZqObOi

- Li, F., Zhang, B., Verma, S., and Marfurt, K. J. (2018). Seismic signal denoising using thresholded variational mode decomposition. *Explor. Geophys.* 49, 450–461. doi: 10.1071/EG17004

- Lian, Q., Qi, Y., Pan, G., and Wang, Y. (2020). Learning graph in graph convolutional neural networks for robust seizure prediction. *J. Neural Eng.* 17, 035004. doi: 10.1088/1741-2552/ab909d

- Maiwald, T., Winterhalder, M., Aschenbrenner-Scheibe, R., Voss, H. U., Schulze-Bonhage, A., and Timmer, J. (2004). Comparison of three nonlinear seizure

- prediction methods by means of the seizure prediction characteristic. *Physica D* 194, 357–368. doi: 10.1016/j.physd.2004.02.013
- Mormann, F., Andrzejak, R. G., Elger, C. E., and Lehnertz, K. (2007). Seizure prediction: the long and winding road. *Brain* 130, 314–333. doi: 10.1093/brain/awl241
- Mormann, F., Kreuz, T., Rieke, C., Andrzejak, R. G., Kraskov, A., David, P., et al. (2005). On the predictability of epileptic seizures. *Clin. Neurophysiol.* 116, 569–587. doi: 10.1016/j.clinph.2004.08.025
- Nazari, M., and Sakhaei, S. M. (2020). Successive variational mode decomposition. *Signal Process.* 174, 107610. doi: 10.1016/j.sigpro.2020.107610
- Nejedly, P., Kremen, V., Sladky, V., Nasser, M., Guragain, H., Klimes, P., et al. (2019). Deep-learning for seizure forecasting in canines with epilepsy. *J. Neural Eng.* 16, 036031. doi: 10.1088/1741-2552/ab172d
- Peng, J., Xue-Jun, Z., and Zhi-Xin, S. (2021). eEpileptic electroencephalogram signal classification method based on elastic variational mode decomposition. *Acta Physica Sinica.* 70:018702-018702. doi: 10.7498/aps.70.20200904
- Rout, S. K., and Biswal, P. K. (2020). An efficient error-minimized random vector functional link network for epileptic seizure classification using VMD. *Biomed. Signal Process. Control.* 57, 101787. doi: 10.1016/j.bspc.2019.101787
- Savadkoobi, M., Oladunni, T., and Thompson, L. (2020). A machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) Signal. *Biocybern Biomed. Eng.* 40, 1328–1341. doi: 10.1016/j.bbe.2020.07.004
- Schelter, B. R., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Schulze-Bonhage, A., et al. (2006). Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. *Chaos* 16, 1–321. doi: 10.1063/1.2137623
- Schulze-Bonhage, A., and Kühn, A. (2008). Unpredictability of seizures and the burden of epilepsy. *Seizure Prediction Epilepsy.* 1–10. doi: 10.1002/9783527625192.ch1
- Sharif, B., and Jafari, A. H. (2017). Prediction of epileptic seizures from EEG using analysis of ictal rules on poincaré plane. *Comput. Methods Programs Biomed.* 145, 11–22. doi: 10.1016/j.cmpb.2017.04.001
- Taran, S., and Bajaj, V. (2018). Clustering variational mode decomposition for identification of focal EEG signals. *IEEE Sens. Lett.* 2, 1–4. doi: 10.1109/LENS.2018.2872415
- Teixeira, C. A., Direito, B., Bandarabadi, M., Le Van Quyen, M., Valderrama, M., Schelter, B., et al. (2014). Epileptic seizure predictors based on computational intelligence techniques: a comparative study with 278 patients. *Comput. Methods Programs Biomed.* 114, 324–336. doi: 10.1016/j.cmpb.2014.02.007
- Truong, N. D., Nguyen, A. D., Kuhlmann, L., Bonyadi, M. R., Yang, J., Ippolito, S., et al. (2018). Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Net.* 105, 104–111. doi: 10.1016/j.neunet.2018.04.018
- Tsiouris, K. M., Pezoulas, V. C., Zervakis, M., Konitsiotis, S., Koutsouris, D. D., and Fotiadis, D. I. (2018). A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput. Biol. Med.* 99, 24–37. doi: 10.1016/j.combiomed.2018.05.019
- Upadhyay, A., and Pachori, R. B. (2015). Instantaneous voiced/non-voiced detection in speech signals based on variational mode decomposition. *J. Franklin Inst.* 352, 2679–2707. doi: 10.1016/j.jfranklin.2015.04.001
- Usman, S. M., Khalid, S., Akhtar, R., Bortolotto, Z., Bashir, Z., and Qiu, H. (2019). Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: review of available methodologies. *Seizure* 71, 258–269. doi: 10.1016/j.seizure.2019.08.006
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* doi: 10.48550/arXiv.1810.04805. Available online at: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
- Wang, C., Li, H., Huang, G., and Ou, J. (2019). Early fault diagnosis for planetary gearbox based on adaptive parameter optimized VMD and singular kurtosis difference spectrum. *IEEE Access.* 7, 31501–31516. doi: 10.1109/ACCESS.2019.2903204
- Wei, X., Zhou, L., Zhang, Z., Chen, Z., and Zhou, Y. (2019). Early prediction of epileptic seizures using a long-term recurrent convolutional network. *J. Neurosci. Methods* 327, 108395. doi: 10.1016/j.jneumeth.2019.108395
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., et al. (2016). Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation. Available online at: <https://arxiv.org/abs/1609.08144>
- Xiang, J., Li, C., Li, H., Cao, R., Wang, B., Han, X., et al. (2015). The detection of epileptic seizure signals based on fuzzy entropy. *J. Neurosci. Methods.* 243, 18–25. doi: 10.1016/j.jneumeth.2015.01.015
- Xue, Y.-J., Cao, J.-X., Wang, D.-X., Du, H.-K., and Yao, Y. (2016). Application of the variational-mode decomposition for seismic time–frequency analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 9, 3821–3831. doi: 10.1109/JSTARS.2016.2529702
- Yu, P.-g., Liu, C. Y., Heck, C. N., Berger, T. W., and Song, D. (2021). A sparse multiscale nonlinear autoregressive model for seizure prediction. *J. Neural Eng.* 18, 026012. doi: 10.1088/1741-2552/abdd43
- Zahra, A., Kanwal, N., ur Rehman, N., Ehsan, S., and McDonald-Maier, K. D. (2017). Seizure detection from EEG signals using multivariate empirical mode decomposition. *Comput. Biol. Med.* 88, 132–141. doi: 10.1016/j.combiomed.2017.07.010
- Zhang, M., Jiang, Z., and Feng, K. (2017). Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump. *Mech. Syst. Signal Process.* 93, 460–493. doi: 10.1016/j.ymssp.2017.02.013
- Zhang, T., Chen, W., and Li, M. (2018). Fuzzy distribution entropy and its application in automated seizure detection technique. *Biomed. Signal Process. Control.* 39, 360–377. doi: 10.1016/j.bspc.2017.08.013