UNIVERSITY OF TURKU

Epileptic Seizure Detection and Prediction based on EEG Signal Analysis

Smart Systems
Department of Computing
Master's thesis in Technology

Author(s):
Renjie Zhang

Supervisor(s):
Prof. Tomi Westerlund
Prof. Wei Chen

18.07.2021
Turku

Master's thesis in Technology

Subject: Intelligent medical engineering
Author: Renjie Zhang
Title: Epileptic Seizure Detection and Prediction based on EEG Signal Analysis
Supervisor(s): Prof. Tomi Westerlund, Prof. Wei Chen
Number of pages: 69 pages
Date: 18.07.2021

Epilepsy is a kind of chronic brain disfunction, manifesting as recurrent seizures which is caused by sudden and excessive discharge of neurons. Electroencephalogram (EEG) recordings is regarded as the golden standard for clinical diagnosis of epilepsy disease. The diagnosis of epilepsy disease by professional doctors clinically is time-consuming. With the help artificial intelligence algorithms, the task of automatic epileptic seizure detection and prediction is called a research hotspot.

The thesis mainly contributes to propose a solution to overfitting problem of EEG signal in deep learning and a method of multiple channels fusion for EEG features. The result of proposed method achieves outstanding performance in seizure detection task and seizure prediction task.

In seizure detection task, this paper mainly explores the effect of the deep learning in small data size. This thesis designs a hybrid model of CNN and SVM for epilepsy detection compared with end-to-end classification by deep learning. Another technique for overfitting is new EEG signal generation based on decomposition and recombination of EEG in time-frequency domain. It achieved a classification accuracy of 98.8%, a specificity of 98.9% and a sensitivity of 98.4% on the classic Bonn EEG data.

In seizure prediction task, this paper proposes a feature fusion method for multi-channel EEG signals. We extract a three-order tensor feature in temporal, spectral and spatial domain. UMLDA is a tensor-to-vector projection method, which ensures minimal redundancy between feature dimensions. An excellent experimental result was finally obtained, including an average accuracy of 95%, 94% F1-measure and 90% Kappa index.

Key words: seizure detection, seizure prediction, CNN, UMLDA, tensor

Table of contents

Chapter 1 Introduction

1.1 Research background

Epilepsy is an intractable neurological disease, and almost 1% of world's population is troubled by it, particularly in some developing countries without enough medical resources. The epileptic seizures are the hallmark of epilepsy disease, manifesting as consciousness loss, twitching of limbs, which is related to abnormal neurophysiological activity in the brain, such as sudden excessive discharges. In addition to the damage of epilepsy to brain neurons, the sudden and unknown arrival of seizure onset may lead to traumatism.

EEG signal is regarded as the overall response of brain neuron electric activity, it contains a great amount of explored and unexplored physical and disease information. EEG recordings can reflect the cerebral electric activity in different brain regions, so it is an unreplaceable form of understanding the information processing in brain and nerve disease diagnosis.

EEG plays a very important role in the diagnosis and treatment of epilepsy. The epilepsy disease is known as the symptoms of spasm, consciousness loss. Those symptoms without EEG signal are not enough to diagnose that patients suffer from epilepsy or not. When epileptic seizure occurs, it can be viewed through EEG that the amplitude and frequency of cerebral electricity is quite different. Such kind of abnormal EEG waveform is called epileptiform discharge. Combine the symptoms and EEG recordings, epilepsy disease can be determined clinically and multichannel EEG recordings make it possible to locate the epileptic lesions. And epilepsy disease differs in seizure types, such as tonic seizure, absence seizure and atonic seizure. EEG also aims to classify seizure types of various severity and provide treatment solutions. With the advent of long-term EEG monitor and ambulatory EEG (aEEG), so recurrent and

sudden epileptic seizure can be recorded and observed by EEG monitoring throughout a day.

Besides clinic diagnosis and treatment, automatic seizure detection is of great significance. Biomedical engineering explores epileptic EEG analysis by signal processing method, like time-frequency analysis, spectral analysis and nonlinear method. Especially in recent years, the technology of artificial intelligence (AI) improves rapidly with the development of the theory and computing performance, including machine learning and deep learning. As a result, the research about seizure detection and prediction have already made lots of progress with advanced AI technologies.

▪ 1.2 Epileptic EEG signal

Since the first acquisition of EEG signals in 1924 by H.Berger [1], there are several ways to collect EEG signals, which can be generally divided into invasive and non-invasive method. Invasive EEG acquisition method implants electrodes into the cerebral cortex, so the electrodes are close to neurons and it can accurately reflect the electrophysiological activities in the cerebral cortex with less interference, better stability, higher resolution and signal-to-noise ratio. But the defect is that it will cause trauma to the head and the surgery operation is quite difficult. Therefore, invasive EEG device is mainly used for research purpose with some special requirements. Non-invasive EEG acquisition method attaches electrodes closely to the scalp, and the obtained EEG signal is also called scalp EEG, which has the advantages of no trauma, simple operation and low cost in comparison. Scalp EEG is the most widely used EEG acquisition method because of its obvious advantages. In order to unify the standard of scalp EEG acquisition, the International EEG Society formulated the “10-20 system EEG placement” in 1958 [2]. How and where electrodes are placed are illustrated in fig1.

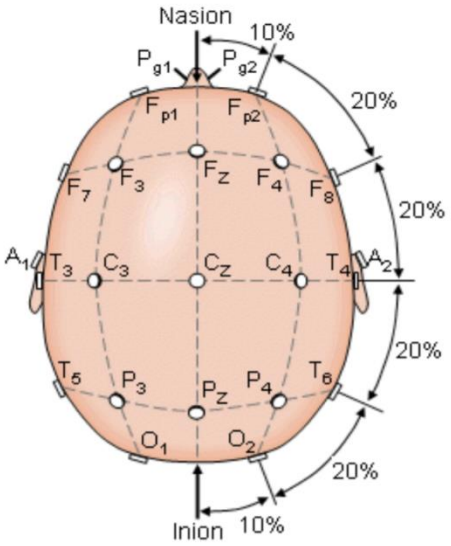


Fig 1 10-20 system EEG placement [2]

There are three elements in brain waves: frequency, amplitude, and waveform. Frequency components present the brain activity in some ways. According to the frequency, EEG can be divided into δ, θ, α, β, γ. Regarding the amplitude, the interpretation of left and right difference and the location of the abnormal side is of great clinical significance.

Collected scalp EEG is a timing signal and its value represents EEG amplitude on each electrode. Human scalp EEG is clinically divided into six frequency bands δ (< 3 Hz), θ (4 – 7 Hz), α (8 −15 Hz), β (16 – 31 Hz) and γ (> 32 Hz), among which the bands greater than 16 Hz can be collectively referred to as β waves or fast waves. These frequency bands are considered to be related to brain activity and consciousness.

The amplitude of the δ wave range from 20 μV to 200 μV. This waveband generally appears in adult slow-wave sleep, infants and young children with incomplete intellectual development may contain the component in their EEG signal. It can be recorded on the temporal lobe and parietal lobe.

The amplitude of θ wave ranges from 5 μV to 20 μV, which is quite low compared with other components. It is the main component of the EEG in infants and teenagers, and the proportion of the EEG in the elderly is higher than that in adults. Adults in

depressed or frustrated mood may generates θ wave. The position of observing θ wave is the temporal lobe, frontal lobe and central area.

The amplitude of α wave ranges from 20 μV to 100 μV. Its amplitude is higher in children's EEG than that in adults' EEG, and the amplitude will gradually decrease to the adult level during growth and development. Alpha wave is considered as the basic rhythm of normal human brain waves. It is very obvious when human is awake or quiet with eyes closed. If the vision is stimulated suddenly (such as light stimulation) or active mental activities are performed, the alpha wave will be greatly reduced. The position where alpha waves appear includes: posterior temporal area, occipital area, parietal area, etc.

The amplitude of β wave ranges from 100 μV to 150 μV, which is the most common EEG waveform for a normal adult in awake state. This wave band generally appears in mental tension, emotional agitation and excitement. The beta wave will increase with age, and the proportion of beta wave in adult women will be slightly higher than that in adult men. Beta waves can be found in a wide range of locations on head, it is symmetrically distributed on both sides of the brain, and the front area is the most obvious.

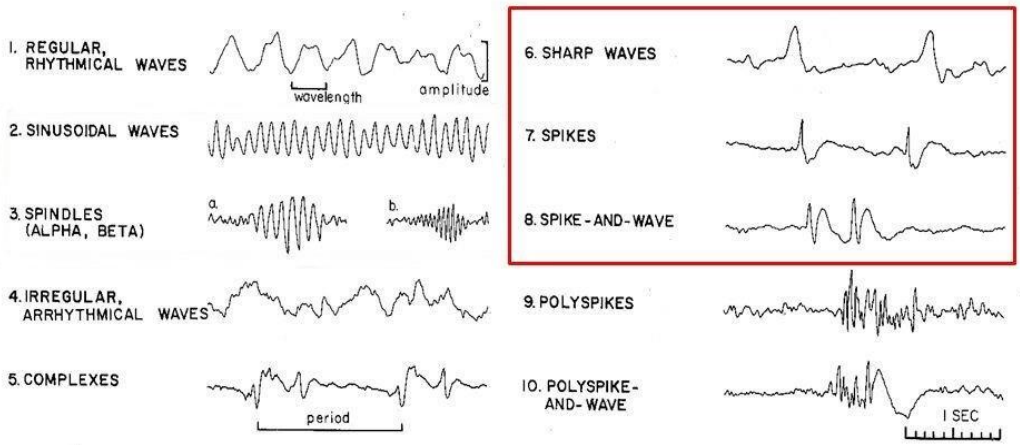


Fig 2. Morphology of EEG [3]

The interpretation of EEG in the diagnosis of epilepsy is marked by epileptic seizures, which appear as sudden and abnormal discharges in shape of epileptiform

patterns, including sharp waves, spikes, spike-and-wave, polyspikes, complexes and polyspike-and-wave. The morphology of EEG is shown in fig 2.

According to the clinical symptoms of epilepsy, seizures can be divided into two categories: partial seizures and generalized seizures. Partial epilepsy is the result of single nidus in brain. It includes a) Simple partial seizures, which is not accompanied by consciousness, and EEG manifests as discharge on single side. b) Complex partial seizures, it is accompanied by consciousness and cannot be recalled afterwards, and it is often bilateral abnormal discharge. Generalized seizures are bilateral and symmetrical abnormal discharges in both hemispheres, including a) Absence seizures, which are more common in children, manifest as sluggishness, activity and language stop, it will return to normal after tens of seconds, but the situation of seizure cannot be recalled. EEG appears as normal background with paroxysmal polyspike-and-wave. b) Myoclonic seizures, it often occurs when falling asleep and waking up with clonic muscle beating, EEG appears as polyspikes and sharp waves. c) Clonic seizures, it manifests as convulsive body, and EEG manifestations of rapid activity above 10 Hz. d) Tonic seizures, the clinical characteristics is strong muscle contraction, and body keeps the fixed position. EEG during seizure period is fast rhythm discharge around 9-10 Hz, the frequency gradually decreases and the amplitude increases after that. e) Tonic-clonic seizure, it is the most common seizure type clinically, clonic period and tonic period occurs alternately. f) Atonic seizure, loss of muscle tone accompanies the seizure onset, which may lead to a sudden fall. EEG is characterized by polyspike-and-wave.

EEG recordings can be categorized by seizure duration, ictal interval represents EEG recording during a seizure, interictal interval represents EEG recordings between seizures and preictal interval represents EEG recordings before seizure onset. The waveform of continuous categorized intervals is plotted in fig 3.

During the interictal phase, the EEG signal is a transient waveform, which is manifested as a sharp wave or spike wave. The EEG signal of an epileptic seizure is

continuous, with a composite waveform of sharp wave and spike wave. Interictal epileptiform discharges and unilateral periodic epileptiform discharges have clear meaning in the diagnosis of epilepsy disease. Spike waves and sharp waves are mainly used to confirm whether the patient has epilepsy or not. Spike waves and sharp waves are usually used to determine whether seizure occurs. Therefore, clarifying the waveform, amplitude, frequency and other characteristics of epilepsy EEG signal has important clinical significance for the research of epilepsy automatic detection and prediction.

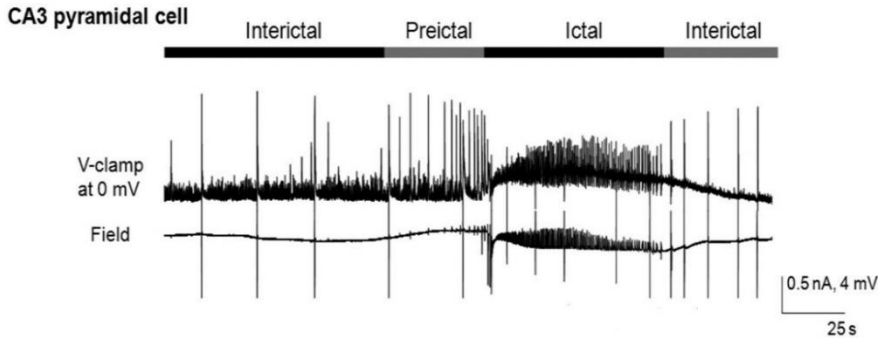


Fig 3. Interictal, preictal and ictal interval of EEG [4]

The object of epileptic EEG signal processing can be specified as classification task by EEG recordings categories, including seizure detection and seizure prediction. Seizure duration only occupies a small part of entire EEG recording, which means interictal intervals are much longer than ictal intervals. So the seizure detection task tries to classify ictal intervals form interictal intervals. The detection technology can relief the diagnostic burden of doctors clinically and avoid the subjective mistakes of EEG interpretation by doctors. In comparison, seizure prediction aims to distinguish interictal intervals and preictal intervals. Once preictal intervals are identified, the following seizure onset is predicted. The time lead of seizure prediction is related to the preictal definition. The earlier preictal intervals are predicted, the sooner medical help can be delivered to patients and the more difficult the task is.

1.3 Thesis outline

In this thesis, we mainly focus on the EEG signal processing of seizure detection and seizure prediction task. The innovation of the thesis includes: CNN for processing EEG in small data size, EEG augmentation method, and discriminant feature extraction and feature fusion for multichannel EEG. The content of each chapter is organized as follows:

In chapter 1, we introduce the background of EEG signal and epileptic disease, then the significance of automatic seizure detection and prediction is explained. And characteristics of epileptic EEG signal is presented, it proves the feasibility of seizure detection and prediction from the perspective of EEG waveform. The final part of the chapter explains the goals and difference of seizure detection and prediction in detail.

In chapter 2, we mainly review the research about seizure detection and prediction, including feature extraction method of tradition EEG signal processing, end-to-end classification method of deep learning, and some other novel techniques. The result of mentioned research is analyzed and strengths and shortcomings are concluded. The challenges faced by seizure detection and prediction are described. Based on the review, this thesis tries to propose corresponding method to improve the existed shortcomings.

In chapter 3, we explored the application of CNN in seizure detection of small EEG data size. In end-to-end model, CNN does not work well as a feature extractor and classifier at the same time. In order to improve the overfitting problem of deep learning model in small samples, we propose two techniques. A hybrid model of CNN and SVM is proposed, CNN is adopted as feature extractor and SVM is adopted as classifier. The structure and training process of the hybrid model is well described. Another technique is the EEG augmentation method based on decomposition and recombination in time-frequency domain. Two proposed techniques improve the performance in seizure detection task compared to end-to-end model.

In chapter 4, we proposed a feature fusion method for multichannel EEG based on UMLDA, which is a tensor-to-vector projection algorithm. Firstly, a tensor object of PSD feature extracted in time-frequency domain and 23 channels is constructed, which includes the information of temporal, spectral and spatial domain. The key point of this part is UMLDA, it projects the three-order tensor into 15-dimension vector with orthogonality between dimensions. And the processed feature is transmitted to KNN for classification. To the best of our knowledge, tensor-based UMLDA has not been applied to predict seizure so far. Based on above algorithm, we perform a 10-minute seizure prediction task on individual subject in CH-MIT dataset.

In chapter 5, a summary of the thesis is presented. We analyzed the result of proposed techniques applied in seizure detection and seizure prediction. And we mention some points needed for improvement in the future.

Chapter 2 Related work

EEG diagnosis has always been a key step in the diagnosis of epilepsy, it was interpreted by experienced doctors in the past. However, this greatly consumes doctors' resources and is prone to errors. With the advancement of signal processing technology, the technology of EEG signal processing developed step by step. In the 1950s, British physician William Grey Walter well described the topography of EEG signal [5], it provides guides both for clinical diagnosis and automatic EEG waveform detection. After 1990, much effort has been made to research the methods of denoising, classification and event detection of EEG signal. Some of current widely-used and effective techniques were emerging from that time, including independent component analysis (ICA) and blind source separation (BSS) [6]. They can extract EEG signal from complex background noise, they are very import because scalp EEG is noisy and limitation of EEG acquisition conditions. And neural network is utilized to classify EEG for seizure activity detection [7]. The rise of artificial intelligence represented by neural networks greatly improves the performance of EEG signal classification, it also promotes the feature engineering of EEG signal. In this chapter, we will review the EEG signal processing method and introduce current research of seizure detection and prediction.

2.1 EEG signal processing method

The general workflow of EEG signal processing can be concluded as signal acquisition, preprocessing, feature engineering and classification.

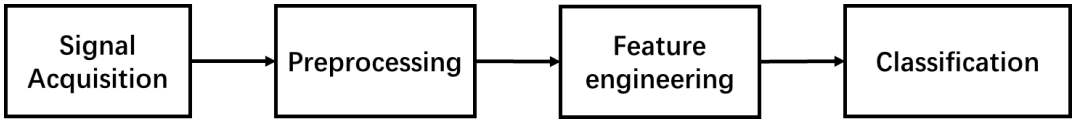


Fig4. Workflow of EEG signal processing

For signal acquisition, scalp EEG is the most common EEG signal because of its advantage of non-invasive, easy collection and low cost. Besides, EEG acquisition of depth electrode implantation is used unless necessary. Because it needs a surgery to implant the electrode inside cerebral cortex and it is quite expensive. So almost all public EEG datasets are scalp EEG, and EEG mentioned in this thesis also refers to scalp EEG.

▪ 2.1.1 EEG preprocessing

Preprocessing work aims to standardize the data into a form which is prepared for following different applications. It tries to preserve as much useful information as possible, while eliminating useless signal component, such as noise and artifacts. This step also specializes the selected signal for certain application.

The artifact in EEG signal can be divided into technical artifacts and biological artifacts. The technical artifacts come from careless signal recording process and noisy environment, such as power line interference, impedance fluctuation and wire movement. The biomedical artifacts are a kind of pollution due to biomedical electrical signal outside brain, so the collected EEG is a superposition of multiple signals. It can be caused by eyeblink, eye-movement, heart beats, muscle contraction.

There are some common steps of preprocessing explained as follows:

a. Filtering

Technical artifacts are much stronger than EEG signal itself, it can be weakened by stringent experimental conditions and equipment. Besides, digital filter can purify the EEG signal.

Scalp EEG signal contains a lot of noise during acquisition, the noise comes from power frequency, thermal effect, acquisition equipment and so on. The doped noise reduces the quality and intelligibility of EEG signals, it makes an unpredictable effect on EEG signal classification. So it is necessary to design an appropriate filter based on requirements.

The frequency band of EEG signal is described above, there is little information contained in high frequency components of EEG clinically. Considering large interface of power frequency, so it is very common to design a band-pass filter which limits the band between 0.5 and 40 Hz.

b. Biomedical Artifact removing

Due to good conductivity of the scalp, biological artifacts from outside the brain can contaminate the observed EEG signal. Electromyography (EMG) artifact caused by facial or neck muscle contraction when swallowing or biting, the frequency distribution of the artifact is 0-200 Hz. The frequency distribution of ECG artifacts caused by heart beat is 0-75 Hz. And the frequency distribution of EOG artifacts caused by eye-movement is 0-13 Hz. The frequency components of biological artifacts and EEG signals may overlap. So it is not possible to totally remove biomedical artifacts by filtering, other methods need to be put forward.

Blind source separation (BSS) refers to the method of separate source signal from observed signal when the mix model of observed signal and source signal is unknown [8]. Independent component analysis (ICA) tries to solve the separation problem by decomposing the observed signal into several independent signal components. The decomposition process is carried out according the principle of statistical independence, and premise is that the source signals are independent non-gaussian signal. In [9], it is proved that ICA outperforms regression-based method in removing artifactual sources. Besides hand-optimized selection of source components in ICA, automatic method based on machine learning are adopted to remove artifact in ICA [10].

c. Other steps

There are some other necessary steps in EEG signal preprocessing besides artifacts removing. The step of segmentation chooses the labeled segment of EEG event. Baseline correction can reduce the impact of baseline drift, it is achieved by subtracting the selected reference mean value. Because of possible poor electrode contact, some EEG channel may miss data or do not satisfy the quality standard. EEG channel interpolation is important in such cases, the data of the channel can be replaced by the

mean value of several around channels. Finally, the abnormal trails which has huge amplitude fluctuation caused by activity outside brain have to be rejected.

2.1.2 Feature extraction method

A good method of feature extraction is very necessary for achieving robust result of signal classification. So the principle of feature extraction is to extract as distinguishable feature as possible. The most commonly used method of EEG feature extraction is introduced as follows:

a. Time domain features

Time domain analysis is intuitive and has a quite relatively clear physical meaning. The frequently used statistical parameters are mean, variance, median, skewness and kurtosis. These amplitude measurements are relevant to discriminate EEG activities. The combination of above features is attempted to achieve good classification performance [11].

b. Frequency domain features

Frequency domain features are very crucial in EEG signal processing, because it involves the spectral analysis. It brings up the parameters of frequency components and energy, which are very distinguishable features since different EEG events vary in these two parameters. For example, the energy of epileptic seizure is distinct form the background due to excessive discharge.

In 2008, intensity weighted mean frequency (IWMF) discussed the frequency distribution of the EEG signal by analyzing the normalized power spectral distribution (PSD), and extracted feature achieves a good performance in seizure detection task [12]. And intensity weighted bandwidth (IWBW) is another measurement of PSD, which is introduced by McDicken and Evans in [13], it compares a large number of proposed energy extraction method and combine the differential energy to achieve better result than single energy extraction. Besides the energy or PSD of certain length EEG segment, instantaneous energy is also explored to achieve more sensitive detection task [14].

c. Time-frequency domain features

Since time feature or frequency feature only take one domain into consideration, its information is limited. And frequency analysis based on Fourier transform is only suitable for signals whose frequency components do not change over time. But EEG signal is a non-stationary random signal, its frequency components changes a lot during various brain activities. In comparison, time-frequency analysis takes both time domain and frequency domain into consideration, it provides a method to observe the information of joint distribution in time and frequency domain.

There are many time-frequency transforms proposed to analyze the signal. Short time Fourier transform (STFT) calculate the Fourier transform in a sliding window, so the frequency components of each time point are obtained if the window length is small enough. Different window length in STFT has different preference for time resolution and frequency resolution. Wavelet transform (WT) offers a solution for the paradox between time resolution and frequency resolution in STFT. It overcomes the problem of resolution in STFT with orthonormal wavelet [16]. And Cohens time-frequency transform is a family of quadratic distribution, its kernel fuzzy function makes it suitable for different conditions. In [18], the difference of various time-frequency transform method is well discussed in seizure detection task.

d. Non-linear features

With the in-depth study of the brain, more and more evidences show that the brain can be approximately regarded as a nonlinear dynamic system, and the EEG signal is its output. Therefore, people continue to try to introduce nonlinear analysis methods into the analysis of EEG signals.

Features that characterize nonlinear complexity or irregularity are widely proposed. In the early 1990s, Pincus proposed and developed the concept of approximation entropy (ApEn) from the perspective of measuring the complexity of time series, and successfully applied it to physiological time series analysis, such as EEG, heart rate signals, blood pressure signals, etc [19]. A large number of entropy derived parameters have been proposed since then, such as sample entropy, permutation

entropy, fuzzy entropy, weighted-permutation entropy, distribution entropy, hurst entropy, etc [20-24]. Hilbert-Huang transform (HHT) is a popular method of extracting features from non-stationary signal, it includes empirical mode decomposition (EMD) and Hilbert transform [25]. The key step of HHT is decompose the signal several intrinsic mode function (IMF), these functions represents the physical meaning of the original signal. HHT has the advantage of clear time and frequency resolution, and it has complete adaptability which does not need the selection of basis function. In [26], information related to IMF is extracted to track the local frequency and amplitude of EEG signal, which helps to distinguish ictal and interictal segments.

2.2 Classification method

For a given set of features, we need select the most appropriate classifiers according the extracted features. Due to the characteristics of EEG signal, there two main problems in EEG classification: the curse-of-dimensionality and the bias-variance tradeoff.

a. The curse of dimensionality

Feature expression ability and separability increase with feature dimension. As we describe above, EEG has a lot of proposed features ready to extract. But if the dimension of signal is close or large than the sample number, the classifiers will suffer from poor results. It has been proved in [27] that the sample numbers of each class in training should be at least five times as many as dimensionality. But it is not easy in EEG signal processing, because data acquisition and annotation are hard to get.

b. Bias-variance tradeoff

The error of classification can be decomposed into three terms: noise, bias and variance. The calculation is listed below:

$$\begin{aligned} ERROR &= E[(y^* - f(x)^2] \\ &= E[(y^* - f^*(x) + f^*(x) - E[f(x)] + E[f(x)] - f(x)^2)] \\ &= E[\left(y^* - f^*(x)\right)^2] + E[(f^*(x) - E[f(x)])^2] \end{aligned}$$

$$+E[\left(E[f(x)]-f(x)\right)^2]$$
$$=noise^2+bias^2+var$$

where y^* is the truth value, $f(x)$ is the predict value and $f^*(x)$ is the label value. And noise here represents the irreducible error between label value and truth value. And bias estimate the fit of the model. Variance reflects the sensitivity of the training data.

There is natural tradeoff between bias and variance [28]. As the model's fitting ability increases, the bias will decrease. At the same time, the variance of the model will gradually increase. In this process, the model goes from underfitting to overfitting. In EEG signal processing, the training data from different sessions are quite dissimilar. In order to achieve generalization ability, a low variance model of EEG classification is preferred in most cases.

▪ 2.2.1 Machine learning

As we introduced above, complex model which have strong expressive ability in can easily lead to overfitting problem. Machine learning is considered as weaker in distinguishing than deep learning, but it can satisfy the EEG classification need in most cases and avoid overfitting problem. Machine learning methods include linear classifiers, decision tree models, Bayesian classifiers, nearest neighbor classifiers, etc.

Linear discriminant analysis is a binary classification method with low computational complexity, which makes it appropriate for EEG activity recognition [29]. But it also shows its drawback of linearity in nonlinear EEG classification task [30]. Support vector machine (SVM) is another linear classifier by locating the optimal hyperplane [31]. And it overcome the drawback of LDA by kernel function, which allows it to have a nonlinear hyperplane, such as RBF SVM and polynomial SVM. So it has been widely used in EEG recognition tasks [32-33]. Hidden Markov model (HMM) is a popular Bayesian classifier in EEG signal recognition. A novel distance coupled HMM is proposed in [34], and it has outstanding advantages in computational complexity. A highly accurate and boosting ensemble method called gradient boosting

decision tree (GBDT) is applied to determine the state of EEG signal [35]. K-nearest neighbor (KNN) is a simple classifier that does not need training, it is used to highlight the effect of feature extraction in [36].

▪ 2.2.2 Deep learning

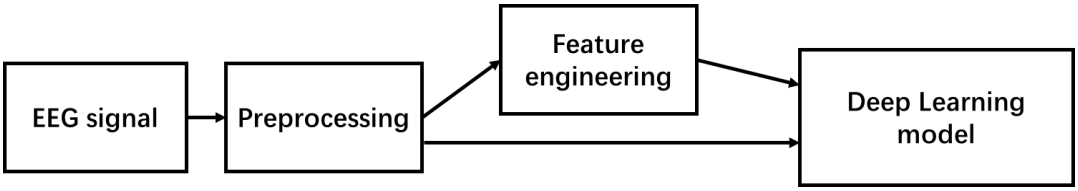


Fig5. EEG signal in deep learning model

Deep learning can be briefly summarized as artificial neural networks with deep depth of hidden layers. Generally, they are flexible under certain network structures, such as convolutional neural networks (CNN) and Recurrent Neural Network (RNN). The increase in the number of layers makes the network model have better fitting capabilities.

With the increasing separability between different types of signals, the signal can be directed fed into deep learning model to perform recognition task without feature engineering. In traditional machine learning, the quality of feature design has a crucial impact on classification performance. Deep learning directly removes the steps of manually designing features, which greatly improves the automation of signal processing. But interpretability of deep learning is insufficient without feature engineering.

Depending on the form of input EEG representation, like time series, dimension features and image representation, various deep learning models have been explored to achieve better recognition performance. In [37], a set of different architectures of end-to-end CNN models are designed to encode and visualize EEG signal. Besides original EEG signal, extracted feature from time and frequency domain still can be the input of deep learning model [38]. Kostas introduces long short-term memory model (LSTM),

which is an improved version of RNN, in seizure prediction task, and it achieves an outstanding result [39]. Generative adversarial network (GAN) is an unsupervised learning method on complex distribution, and it is the current research hotspot of deep learning. It has been applied to generate EEG signal as data augmentation [40].

2.3 Summary of existing research

After the above introduction, it is concluded that there are many methods for brain signal feature extraction and classification, and new ones are constantly emerging. But some innovations in feature combination or classifier ensemble methods are not pioneering.

In our opinion, there are still several problems that need to be improved in EEG signal processing. In feature engineering, the extraction of multi-channel information has not received enough attention. And the fusion method of channel information needs further study. In deep learning, the potential overfitting in small EEG data is not well studied. In the following task of epilepsy detection and prediction, we will put forward our ideas and attempts about these two issues.

Chapter 3 CNN for epileptic seizure detection in small data size

In this chapter, we focus on the deep learning solution for seizure detection task. In order to solve the contradiction between the complex CNN network and the small amount of data, two techniques are mainly introduced and their improvements in classification performance are verified. One of the techniques is a hybrid CNN-SVM model. the hybrid model uses CNN as a feature extractor to avoid overfitting in classification, and the feature vector before the last fully connected layer is input to the SVM to perform classification task. Another technique is generating new EEG signal by segmentation and recombination in time-frequency domain. The model without handcrafted features achieves a competitive result in seizure detection task.

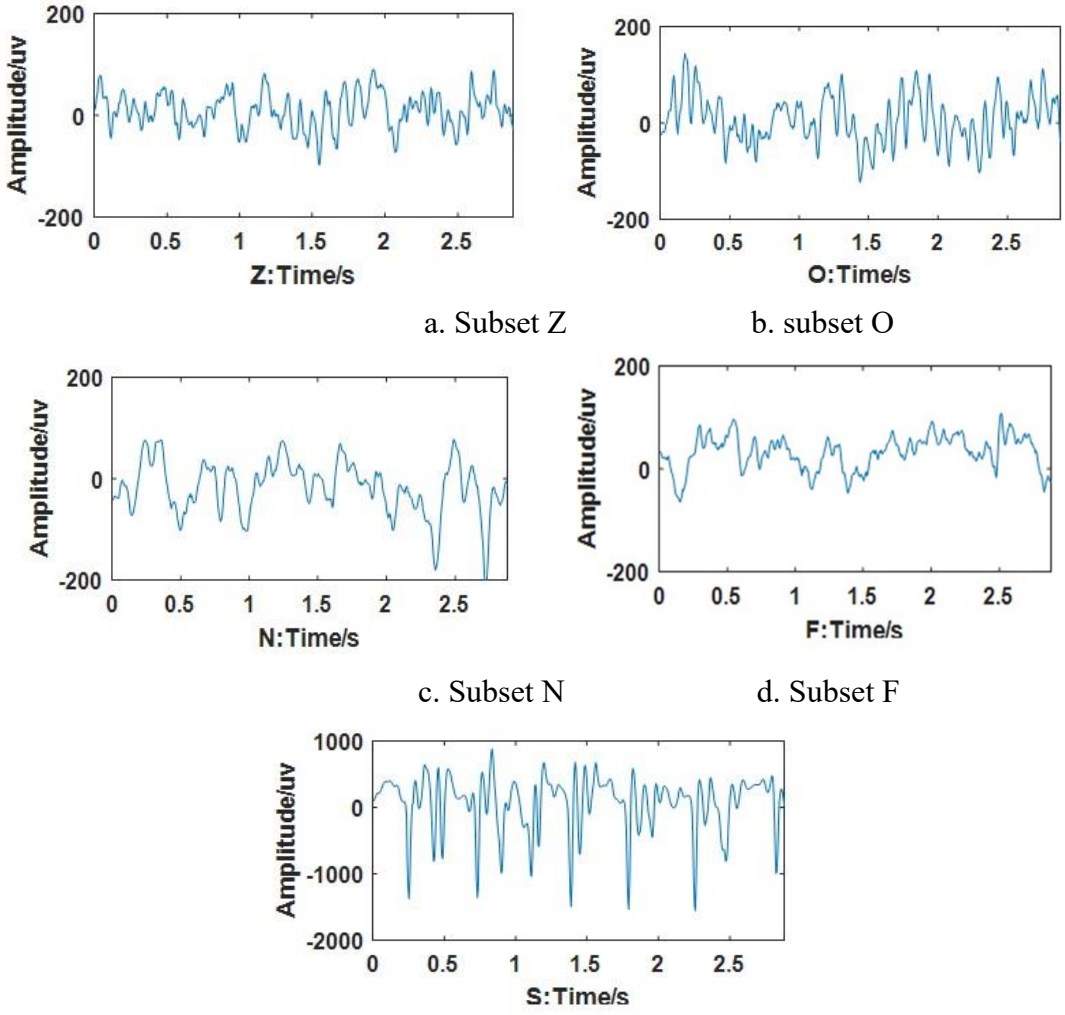
▪ 3.1 Dataset

Table 1. Details of subsets in Bonn dataset

Subsets	Subjects	Subject's state	Segment number	Duration	Frequency
Z	5 healthy volunteers	Eyes open	100	23.6s	173.61Hz
O	5 epileptic volunteers	Eyes closed	100	23.6s	173.61Hz
N	5 epileptic volunteers	Interictal intervals one side of hippocampal formation	100	23.6s	173.61Hz
F	5 epileptic volunteers	Interictal intervals another side of hippocampal formation	100	23.6s	173.61Hz
S	5 epileptic volunteers	Ictal intervals	100	23.6s	173.61Hz

In order to meet the experimental purpose of small data samples, the dataset collected by Bonn University is utilized in the thesis [41]. And it has been widely used

in the research filed of epileptic seizure detection. There are five subsets contained in the dataset, donated as subset Z, subset O, subset N, subset F and subset S. Each subset consists of 100 single channel EEG segment, and the length of all individual segment is 23.6 seconds. The EEG sampling frequency is 173.61 Hz with 12bits resolution, and its spectral bandwidth ranges from 0.5 to 85 Hz which is consistent with the acquisition equipment. The EEG recording is under the standard placement of electrode, and the reason of single channel and short duration is that each segment is selected and cut from long-term multi-channel EEG to remove biomedical artifacts by experienced doctors.



a. Subset Z b. subset O

c. Subset N d. Subset F

e. Subset S

Fig.6 Examples of EEG segment in 5 subsets of Boon dataset

The difference of subsets lies in subjects, subjects' state. Subset Z and O are collected from five healthy subjects without epilepsy, and the difference is whether the eyes are open. And subjects N, F and S are collected from five epileptic patients. Subset N and F only contains interictal intervals, which have no overlap with seizure segments. And EEG of subsets N and F are taken from the symmetrical sides of hippocampal formation in brain where is considered as the seizure generating area clinically. Subsets S only contains pure ictal segments. The detail of each subsets is listed in table 1, and fig 6 give an example of EEG waveform in each subset.

3.2 Data preprocessing

Preprocessing of raw EEG data helps to achieve better result in following classification task. The EEG signal of Bonn dataset is quite clean because the biomedical artifacts are removed by doctors' selection, such as ECG artifacts caused by heart beats, EMG artifacts caused by muscular activities and EOG artifacts caused by eye blinks. The spectral bandwidth of EEG signal in dataset is form 0.5 Hz to 85 Hz, which is determined by the acquisition system. And the frequency of most understandable brain activities is between 3-31 Hz [42]. Considering the other noises, such as power frequency, we design a 6-order Butterworth bandpass filter, which offers a flat passband to preserve desired signal. The cut-off frequency is set as 0.5 HZ and 40 Hz.

3.3 CNN for seizure detection

3.3.1 Artificial neural networks (ANN)

Biological neurons can cooperate with each other to complete information processing. Inspired by the complex system in brain, artificial neural network is created as a digital signal processing structure [43]. Generally speaking, a neural network consists of multiple layers of neurons, and the neurons of two adjacent layers are linked

by weights. Once the structure of the network is determined, that is the number of layers and the number of neurons in each layer, the process of ANN training is to find the appropriate weight to best fit the target output. The weight update is achieved through the gradient back propagation of the objective function, which is the most critical principle and calculation of ANN. And activation function helps neural network to obtain nonlinearity. A neural network with a single hidden layer is illustrated in Fig7, the number of hidden layers and neurons will be larger in general.

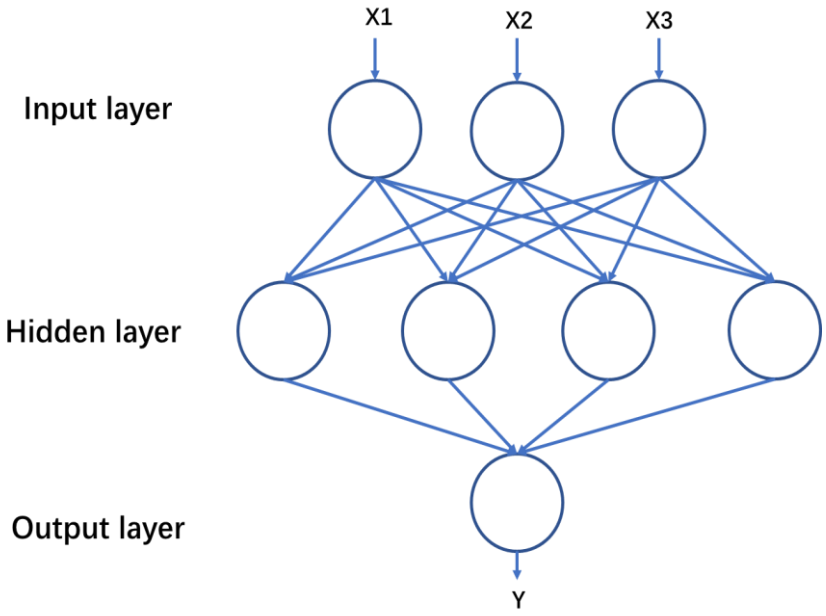


Fig7. Structure of single hidden layer ANN

▪ 3.3.2 1-Dimension CNN structure

CNN is a specialized neural network with certain structure. It achieves shift and translation invariance with the introduction of convolution layer compared with ANN, which is very import in image processing. With the development of deep learning theory and the improvement of computing equipment performance, CNN has caused a research boom in computer vision (CV), natural language processing and other related fields [44].

CNN has three types of layers, convolutional layer, pooling layer, and fully connected layer, which implement feature extraction, feature selection and classification respectively. The operation of dot product between convolution kernel weight and signal in the convolution layer can be regarded as filters, and the result obtained is the destination value of feature extraction. The pooling layer samples the feature maps output by the convolutional layer, which is the process of feature selection. The final fully connected layer converts high-dimensional features into output categories to achieve classification.

CNN has shown excellent performance in the field of image processing, including biomedical signal images. However, it is not widely used in one-dimensional time series signals such as EEG signals. In EEG signal processing, the convolutional operation in CNN is also transformed into one-dimensional, but advantages of convolution layer still exist, so this thesis uses CNN to do the task of epilepsy detection. The following part will explain each layer of one-dimension convolution in detail.

a. 1-D convolution layer

In ANN, the calculation of the output of each neuron requires the output of all neurons in the previous layer and the corresponding weights, which is called fully connected structure. Such a process is computationally expensive and generates a large number of parameters, which is not necessary in some cases.

The convolution operation in the convolutional neural network is a non-fully connected structure. The convolution kernel is a structure smaller than the size of the input signal, and its value represents the weight, which is the target in learning process. The feature map output can be obtained by moving the convolution kernel on the signal. In this way, local features are extracted with the reuse of weight, and computational complexity is reduced. The process of one-dimension convolution illustrated in fig8.

There are some parameters need to be set during the convolution process, including number of input and out channels, stride, padding and convolution mode. The number of output channels determines the number of one-dimension output signal here. Stride specifies the step size of each move of the convolution kernel, and padding

specifies the padding size of input signal. Convolution mode has other two selections in addition to mode ‘full’, mode ‘same’ specifies same size of input and output, and mode ‘valid’ only produces output when kernel is fully used.

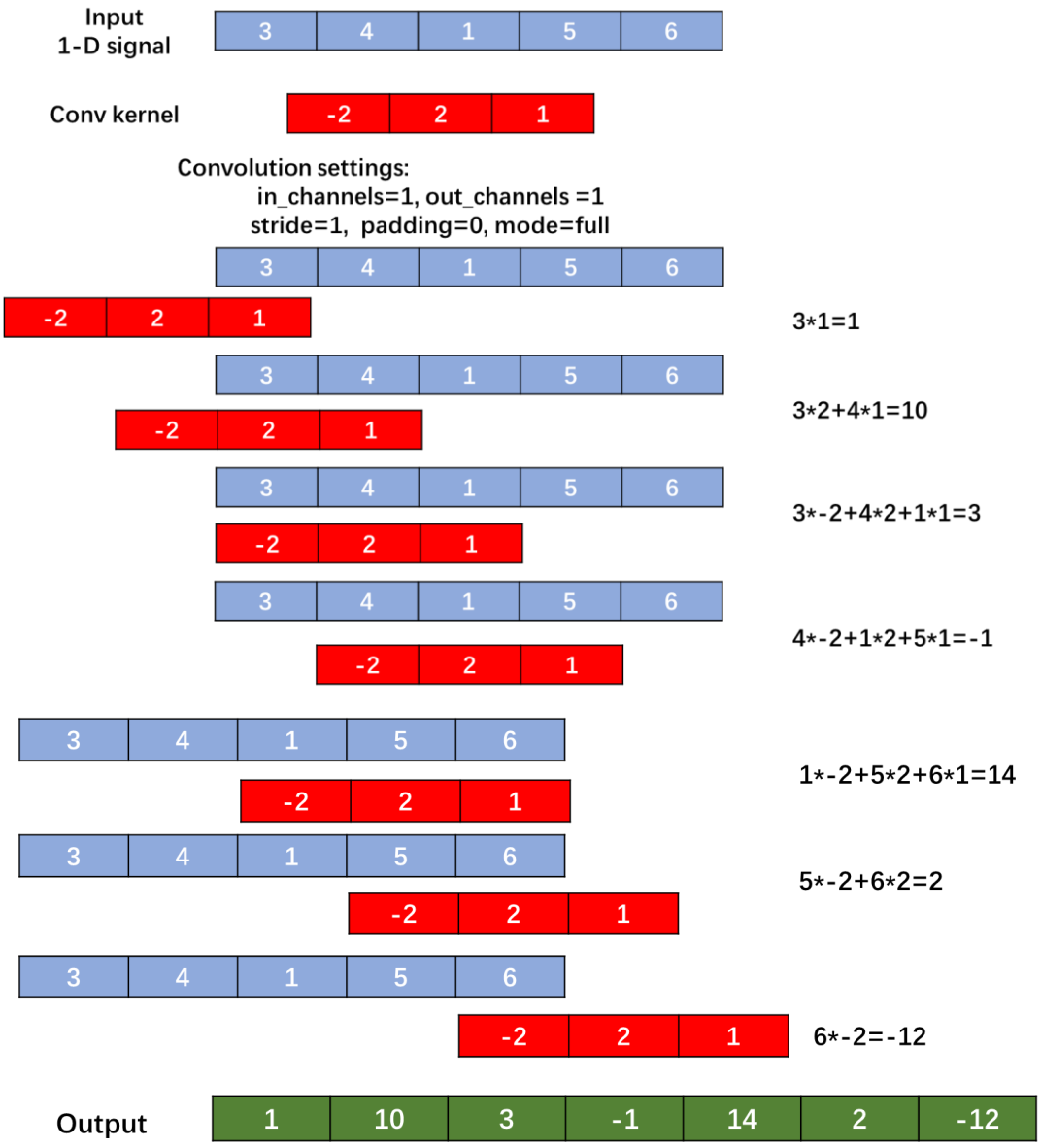


Fig8. An example of 1-dimension convolution

b. 1-D pooling layer

The pooling layer is also called the down-sampling layer. It has two main functions in CNN. The first is to reduce the dimensionality of feature vectors, reduce parameters and calculations, and prevent overfitting to improve model generalization performance. The second function is that it brings feature invariance. For example, in one-

dimensional signal processing, the same time sequence signal can obtain the same feature through pooling operation at different sampling frequencies.

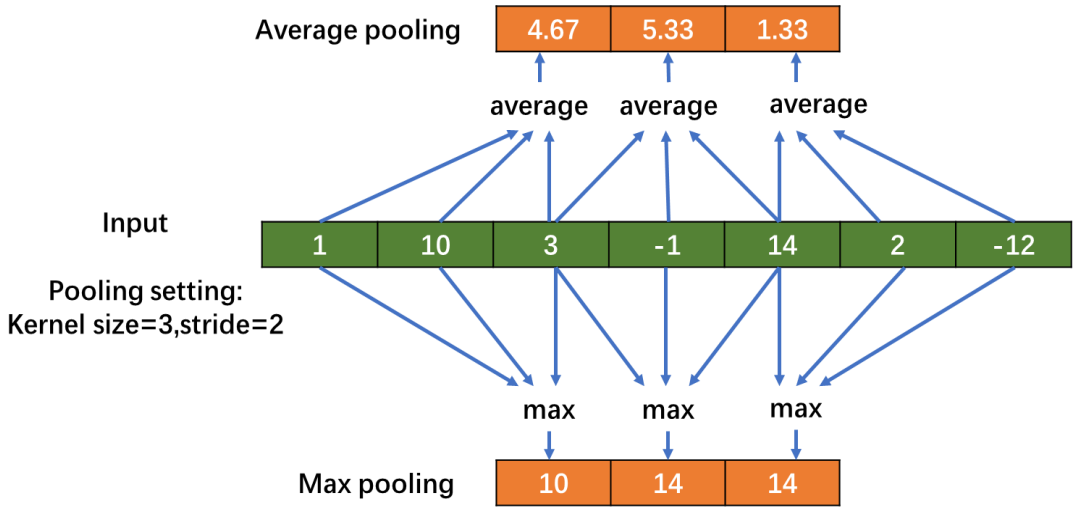


Fig9. An example of one-dimension pooling

The process of one-dimension pooling is plotted in fig9. There are two ways of pooling operation: maximum pooling and average pooling. Maximum pooling saves the texture characteristics of the data, and average pooling emphasizes the overall data characteristics. According to their characteristics, maximum pooling can be used to extract features with obvious response in shallow layers, and average pooling can be used to store more information in deep layers.

c. fully connected layer

In CNN, full connection usually appears in the last few layers, and it is used to do a weighted sum of the previously designed features. Its structure is similar to the above-mentioned ANN. In the classification task, the number of output neurons in the final fully connected layer is equal to the number of categories to realize the classification task.

d. activation function

The output of ANN is always a linear combination of the input without activation function, and the approximation ability of the network is quite limited. Nonlinear function is introduced as the activation function, so that the deep neural network

expression ability is more powerful. There are two activation functions adopted in this work: leaky rectified linear unit (RELU) and softmax [45-46].

(1) Leaky RELU

RELU is a widely used in activation function, especially in deep learning. The equation is listed below.

$$f(x) = \begin{cases} x & when\ x \geq 0 \\ 0 & when\ x < 0 \end{cases}$$

RELU turns all negative values into 0, which means that only neurons with an input greater than 0 will be activated. It makes the network sparse and improves computational efficiency, and the training process converges faster. At the same time, there is no saturation zone like in the S function. The gradient is always 1 when the input is greater than 0, so the continuous multiplication of the number of layers will not cause vanishing gradient and exploding gradient. The disadvantage of RELU is neuron death, which means that the excessive gradient update causes the neuron to always stay in the area less than 0 and no longer activates.

The leaky RELU is proposed to improve the situation. The equation is listed below. The gradient of the negative zone is no longer 0, but a constant close to 0. Constant α is set as 0.01 here.

$$f(x) = \begin{cases} x & when\ x \geq 0 \\ \alpha x & when\ x < 0 \end{cases}$$

(2) Softmax

Softmax is widely used in classification tasks, it generally serves as the last layer of the network. It maps the input to a 0-1 real number, and the sum of the output is 1 through normalization. So the output represents the probability of the corresponding class, and the category corresponding to the maximum value is the predicted output. Combined with the cross-entropy loss function in the training process, the gradient calculation is very simple and convenient. The equation is listed below.

$$S_i = \frac{e^i}{\sum_j e^j} \quad for\ j = 1,2, \dots k$$

where i, j represents the input, k is the number of classes, and S_i is the output probability.

▪ 3.3.3 Proposed CNN in seizure detection

As we described in Bonn dataset, the length of the input signal is 4097 according to the frequency of 173.61 Hz and the duration of 23.6s. And seizure detection is a binary classification task. So the final input dimension of the network is 4097, and the output dimension is 2.

The structure of designed CNN is described in table 2. The size of the data in Table 1 is expressed in tensor form, the first two dimensions represent the signal size, and the last dimension represents the number of channels. And the process of signal transition is illustrated in fig 10.

Table 2. The structure of CNN

layers or activation function	type	input size	kernel size	number of kernel (output channel)	stirde	output size
1	**Conv**	**4097*1*1**	**7*1**	**2**	**2**	**2046*1*2**
2	**max pooling**	**2046*1*2**	**2*1**	**2**	**2**	**1023*1*2**
3	**leaky RELU**	**1023*1*2**	/	/	/	**1023*1*2**
4	**Conv**	**1023*1*3**	**3*1**	**4**	**2**	**511*1*4**
5	**max pooling**	**511*1*4**	**2*1**	**4**	**2**	**256*1*4**
6	**leaky RELU**	**256*1*4**	/	/	/	**256*1*4**
7	**Conv**	**256*1*4**	**3*1**	**8**	**2**	**128*1*8**
8	**average pooling**	**128*1*8**	**2*1**	**8**	**2**	**64*1*8**
9	**leaky RELU**	**64*1*8**	/	/	/	**64*1*8**
10	**fully cnnected**	**64*1*8**	/	/	/	**100*1*1**
11	**fully cnnected**	**100*1*1**	/	/	/	**2*1*1**
12	**softmax**	**2*1*1**	/	/	/	**2*1*1**

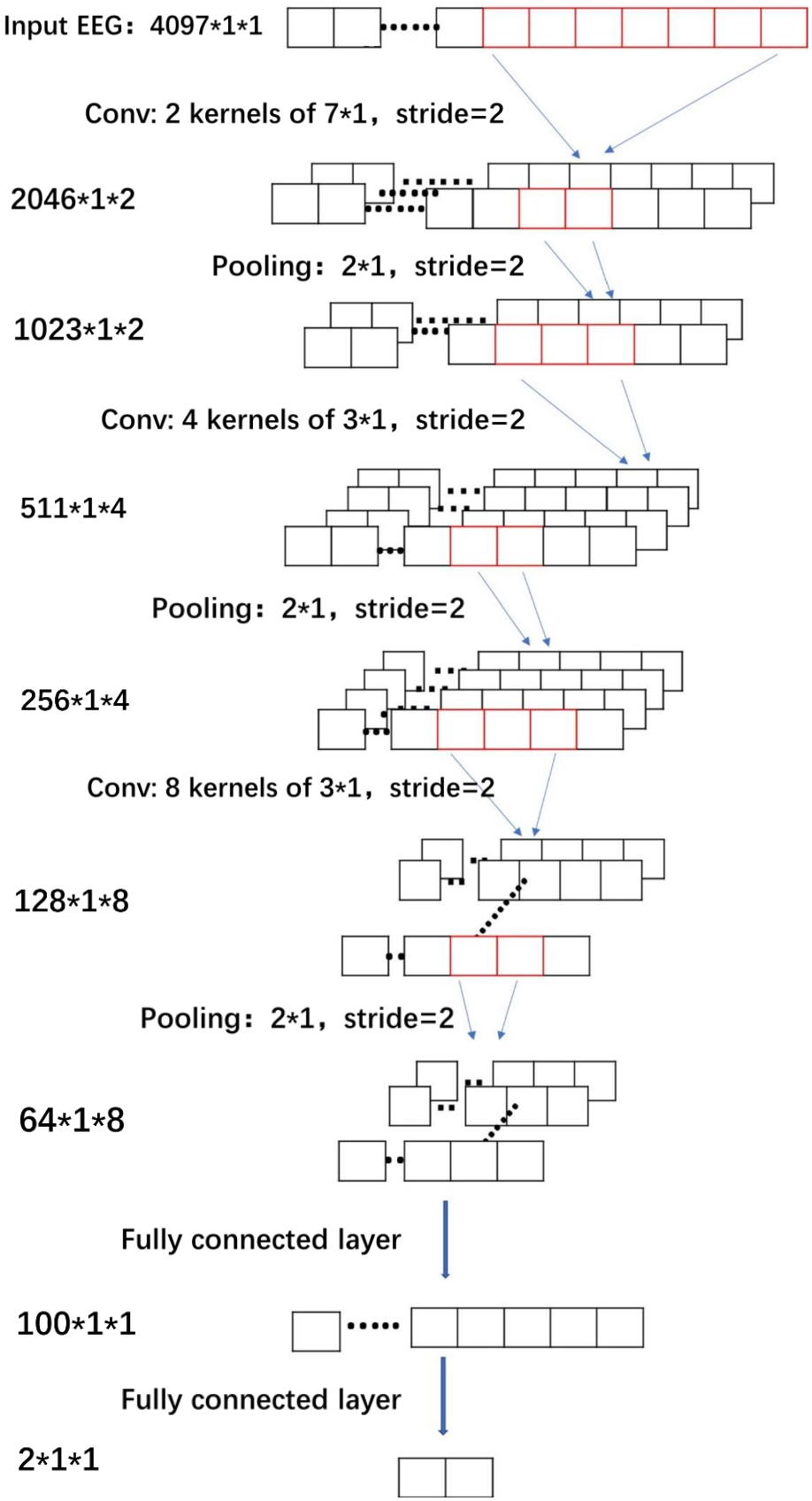


Fig 10. Signal transition in CNN layers

The original input signal is first convolved with a kernel of size 7, and number of signal channels is increased to 2. Under the action of the convolution kernel with stride equal to 2, the size of the feature map is reduced by half. At the same time, the number of channels is double every time. The advantage of the operation is that multi-channel computing and parallel computing are fast during feature extraction. Under the constraint of the same receptive field size, the computation complexity of consecutive small convolution kernels is lower than a single large convolution kernel. And the following pooling layer is used to reduce the feature dimension, and final pooling layer is replaced by average pooling to preserve more information in extracted features. Two fully connected layers were designed instead of one, the design purpose of them is different. The first fully connected layer maps the information learned by convolution layers to feature vectors with larger dimensions, which increases model capacity. The second fully connected layer mainly matches the output category of the epilepsy detection network. Finally, a softmax activation function is deployed to normalize feature vectors and generate probability estimates of each class.

▪ 3.3.4 Results

In seizure detection task, we select subset N and subset F as interictal intervals and subset S as ictal intervals to evaluate our proposed method. Subset Z and subset O are left out to avoid sample imbalance between classes. So totally 300 EEG signal segments are included to perform the classification task.

A 5-fold cross-validation is adopted here [47]. This method can reduce the random error of training on a dataset. Take the 5-fold cross-validation used in this experiment as an example. It divides the data evenly into five parts, 4 parts are used for training, and the remaining part is used for testing. The strategy is repeated by loop selection of training set and testing set. In this way, the performance of the model on 5 datasets is obtained, and the average of the result is taken as final result.

We adopt the metrics of sensitivity, specificity and accuracy to evaluate the performance of the CNN. In the two classification problems of epilepsy detection, we regard the ictal intervals as positive cases and the interictal intervals as negative cases. So true positive (TP) represents the correct classification of the ictal intervals, and false negative (FN) represents that ictal interval is mistaken as interictal interval. And true negative (TN) represents the correct classification of the interictal interval, and false positive (FP) represents that interictal interval is mistaken as ictal interval. The presented metrics is calculated as below. The result of classification is listed in table 3 and the confusion matrix is plotted in fig11.

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

Raw EEG signal classification by CNN has achieved an average sensitivity of 92.0%, specificity of 92.5% and accuracy of 92.3%. Our designed CNN has outperformed the research proposed by U. Rajendra Acharya, which also uses CNN for seizure detection without extra feature extraction [85]. But there are still some other works on the same dataset that have better performance than ours. X. Zhao et al. extract instantaneous energy to perform seizure detection, it is achieved an average accuracy of 99.3% [48]. Diykh et al. designed complex networks features and LS-SVM reaches 97.8% accuracy in same classification of subsets [49]. And Zeng et al. utilizes entropy of visibility graph and LS-SVM achieved a classification accuracy of 98.4% [50]. In comparison, the performance pf result still has room for improvement. Our view is that CNN is not doing well with a small sample size, so two techniques are employed to improve CNN's performance in this situation.

Table 3. Seizure detection result of CNN

		Predicted class		Sensitivity	Specificity	Accuracy
		ictal	interictal			
Actual class	ictal	92	8	92.0%	92.5%	92.3%
	interictal	15	185			

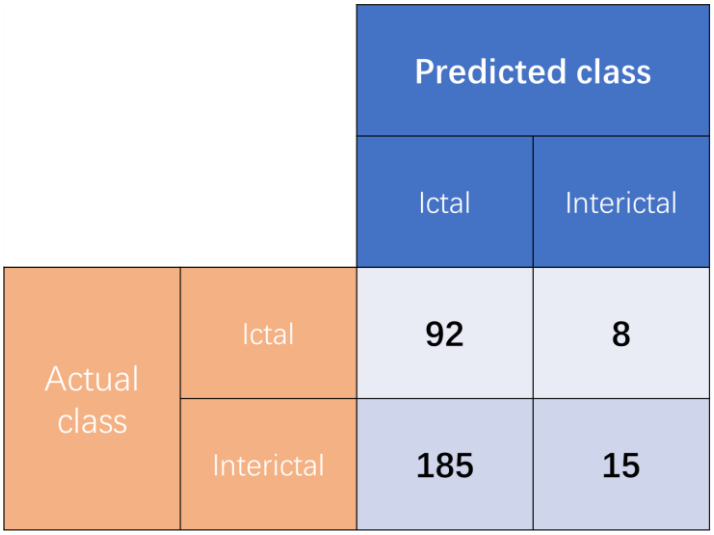


Fig11. Confusion matrix of CNN classification

3.4 Improvements: techniques in small data size

3.4.1 Hybrid model: CNN-SVM

The classification result of direct CNN output is not sufficiently competitive compared with the existed researches. The reason may come from the fact that the data sample is too small and it is not enough for deep network training. However, traditional machine learning methods, such as SVM, can often achieve good results when in small sample classification tasks. Its computational complexity is also very small because of its structure. Considering that the features extracted by CNN have excellent

characteristics of shift and translation invariance. It is still advisable to extract feature by CNN structure. Therefore, we consider using CNN to extract features of EEG, and then input the extracted features into SVM to perform classification task. A hybrid model of CNN and SVM is designed to detect seizure.

a. SVM

SVM is a two-class classification model. SVM maps feature vectors to space, its basic idea of classification is to find the separation hyperplane with the largest interval between classes. SVM is suitable for small and medium data samples, nonlinear, high-dimensional classification problems. Any hyperplane can be described by the following equation:

$$w^t x + b = 0$$

The points closest to the hyperplane in the sample are called support vectors. The distance from the support vector to the hyperplane is assumed as ***d***, so the distance from other points to the hyperplane is greater than ***d***. According to the formula for the distance from a point to a straight line, we can get the following formula:

$$\begin{cases} \dfrac{w^t x + b}{||w||} \geq d & y = 1 \\ \dfrac{w^t x + b}{||w||} \geq d & y = -1 \end{cases}$$

Let $\left|\left|w\right|\right| * d = 1$, we can merge the two equations:

$$y(w^t x + b) \geq 1$$

And the goal of maximizing the distance from the support vector to the hyperplane can be summarized as:

$$\max \frac{1}{2} ||w||^2$$

So SVM is optimization problem mathematically, which is to maximize the objective function under the above constraint function. For the above problem, the Lagrange multiplier method and the Karush-Kuhn-Tucker (KKT) condition of unequal

constrained optimization problems can used to convert and solve the optimization problem [51-53]. A simple illustration of a linear SVM is drawn in Fig 12.

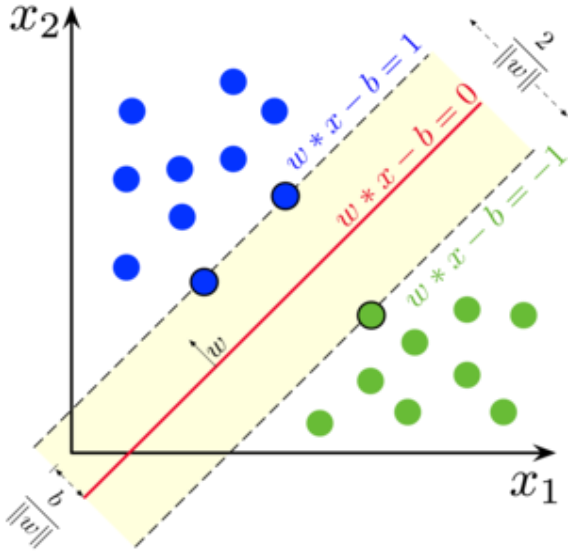


Fig 12. Illustration of linear SVM [78]

The above SVM can only be effective when the sample is linearly separable, kernel techniques is introduced to bring nonlinear ability to SVM [54]. Kernel techniques are what make SVM powerful. The kernel function maps the linearly inseparable samples into the high-dimensional space, where the sample points are linearly separable. Common kernel functions include Gaussian kernel function, linear kernel function, and polynomial kernel function.

b. Hybrid of CNN and SVM

The whole process of feature extraction and classification is as follows: 1) the designed CNN is trained for seizure detection task. 2) the output of the penultimate fully connected layer is the target feature in CNN. 3) the extracted features and corresponding labels are fed into SVM for training process. 4) the remaining dataset is adopted as testing data to evaluate the hybrid model. The workflow is also described in fig 13.

The output of the penultimate fully connected layer of CNN is a 100-dimensional vector. So it is also the input of SVM and the kernel function of SVM selected here is linear kernel function.

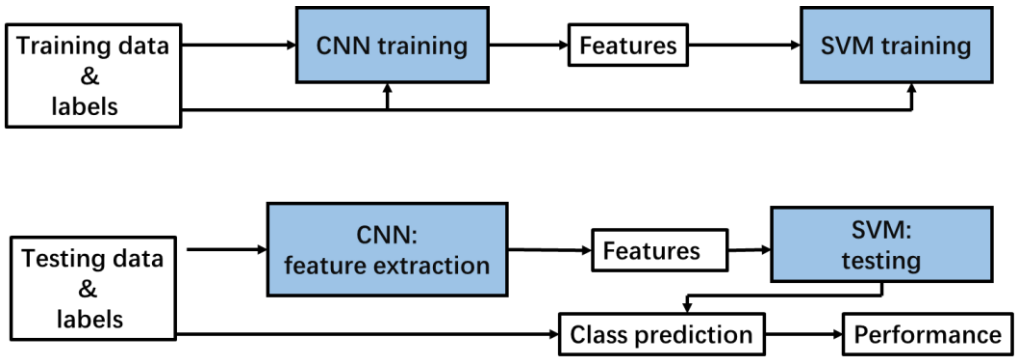


Fig 13. Workflow of hybrid model

c. Result of hybrid model

The result of proposed hybrid model in concluded in table 4. The improved hybrid model achieved better results than an end-to-end CNN. The sensitivity increased by 6%, specificity increased by 4%, and the final accuracy increased by 4.7%.

The experiments can prove that the hybrid model of CNN and SVM is more advantageous when dealing with small data volume EEG, because it combines the advantages of CNN feature extractor and the excellent performance of SVM as a classifier.

The proposed hybrid model avoids the poor classification performance of CNN caused by insufficient training in a small amount of data. Retaining CNN as the feature extractor is more convenient than extracting features manually. Finally, a good classification through was obtained through SVM. The entire hybrid model structure provides a new idea for the deep learning model in the processing of such small amounts of data in biomedical signals.

Table 4. Seizure detection result of hybrid model

		Predicted class		Sensitivity	Specificity	Accuracy
		ictal	interictal			
Actual class	ictal	98	2	98.0%	96.5%	97.0%
	interictal	7	193			
compared with end-to-end CNN		/	/	6.0% increase	4% increase	4.7% increase

▪ 3.4.2 EEG data generation

To further improve the performance of seizure detection, we considered the impact of the amount of data on the results. Not only in EEG processing, all machine learning models desire large amounts of data. The large amount of data will include almost complete data distribution, improving the robustness and generalization ability of training model. Therefore, artificial data generation is proposed to augment training dataset. It has proved its performance increase on testing set in image processing area and speech processing area [55-56].

There are many data enhancement operations in image processing, such as flipping, rotating, scaling, cropping, adding random noise [57]. But those operations are not appropriate for time series signal like EEG. We apply the perspective of component analysis to create new EEG data.

A simple idea is to divide the preprocessed EEG trail into smaller segments, and then recombine the segments between different trials to generate a new EEG trial. Fig 14 demonstrate the process. An EEG trial is divided into 3 equal duration, two trials from the same class can form a new EEG trial by exchanging segments in the same position. In this way, every two EEG trial can generate new six trials. However, this method is too brutal. Although the newly generated EEG trials maintains continuity in time, there may be discontinuities in amplitude. And it results in conflicts in frequency components between different segments, so it is no longer suitable for feature extraction.

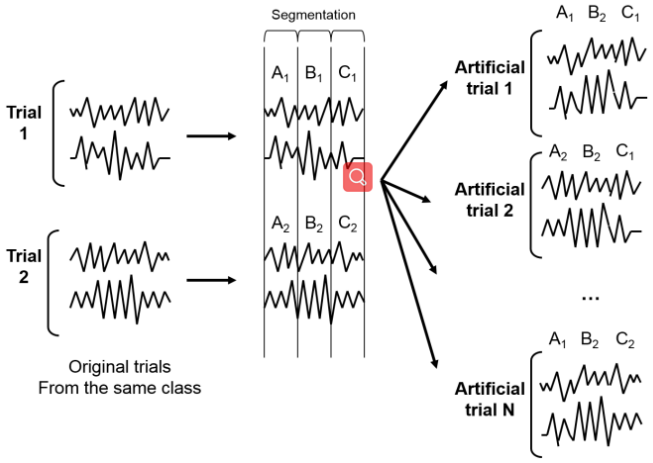


Fig 14. EEG data augment in time domain [58]

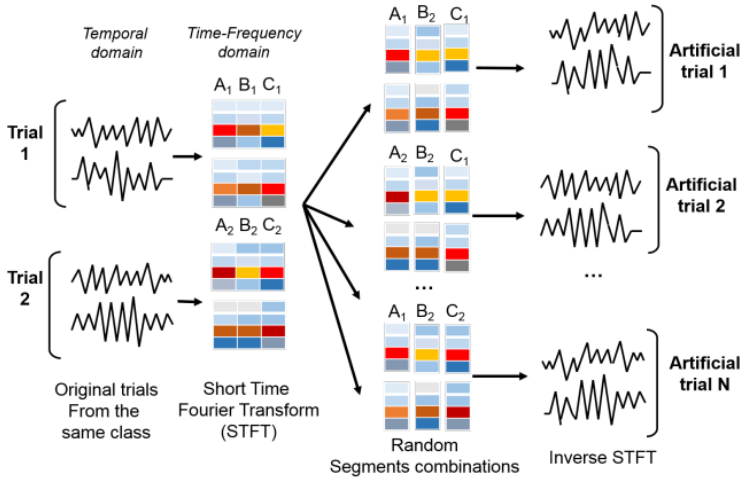


Fig 15. EEG data augment in time-frequency domain [58]

In order to avoid above defects, the segmentation and combination process is performed in time-frequency domain. We obtain the two-dimensional time-frequency distribution of the signal through STFT. Similarly, the two-dimension distribution is divided into three parts on the time scale and then combined. The generated EGG is obtained by inversed STFT. The calculation of STFT and inverse STFT are listed below. The process is plotted fig 15.

$$\mathrm{STFT(t},\omega) = \int x(\tau)h(t-\tau)e^{-j\omega\tau}d\tau$$

$$\text{inverse STFT} \quad \mathrm{x(t)} = \frac{1}{2\pi}\int X(\tau,\omega)e^{-j\omega\tau}d\omega\, d\tau$$

After time-frequency transformation, each time segment has its own independent frequency component, so the exchange between segments will not destroy the continuity of time and frequency.

All possible combinations within each class are enumerated, the augmented EEG data contains 29800 ictal trials and 119600 interictal trials. We retrain the hybrid model of CNN and SVM, the testing result is listed in table 5. Based on the hybrid model, the performance is further improved. The sensitivity increased by 0.4%, specificity increased by 2.4%, and the final accuracy increased by 1.8%.

Table 5. Seizure detection result of augmented EEG

		Predicted class		total	Sensitivity	Specificity	Accuracy
		ictal	interictal				
Actual class	ictal	29323	477	29800	98.4%	98.9%	98.8%
	interictal	1330	118270	119600			
compared with end-to-end CNN in small data size		/	/	/	6.4% increase	6.4% increase	6.5% increase
compared with CNN-SVM in small data size		/	/	/	0.4% increase	2.4% increase	1.8% increase

3.5 Discussion

Deep learning methods represented by CNN have made more and more achievements in the field of epilepsy detection. But few of them paid attention to the dilemma of small data size, which is the target of this chapter.

In [85], CNN is used to process one-dimension EEG signal for seizure detection for the first time, it achieves an average accuracy of 88.7%, which is even lower than our initially proposed end-to-end model. The reason for its structure design is not clearly explained. CNN model design in this thesis has higher interpretability, the

difference between our CNN structure lies in the design of the convolutional layer, the choice of pooling layer, and the number of fully connected layers. Consecutive small convolution kernels are adopted in this thesis to achieve same receptive field size as single large convolution kernels, this action improves the non-linearity while reducing the computational complexity. And the last pooling layer in this thesis is average pooling layer, it can preserve extracted features in deep layers compared to max pooling layer in [85]. And the three fully connected layers is unnecessary, since two fully connected layers increase model capacity and output category probability respectively. These differences are the reason why our model performs better in seizure detection task on same dataset.

Most CNN methods applied in seizure detection task after [85] handle with 2-dimension images, where EEG signal is first converted into time-frequency distribution by time-frequency transform [86-87]. It can not be regarded as a strict end-to-end model. This thesis explores the performance of CNN in processing one-dimension EEG signal, it reduces the steps and complexity of EEG signal processing and it is also the direction of future research.

To the best of my knowledge, the problem of insufficient EEG data in seizure detection based on deep learning has not be well studied before. This thesis proposed two techniques to solve the problem. And they are proved to be effective. Although the result of seizure detection is not the best on Bonn dataset, it provides some thoughts for applications of CNN in EEG signal processing.

3.6 Summary

In this chapter, we explored the CNN model for seizure detection task in small sample size. The end-to-end CNN structure does not show good superiority with 92.3% accuracy, insufficient training affects the classification performance. We proposed two techniques to improve the detection of seizure. A hybrid model of CNN and SVM is constructed, features are exacted from fully connected layers in CNN and classification

task is performed in SVM. Another technique is EEG data augmentation based on decomposition and combination of time-frequency components. The accuracy of seizure detection is increased to 98.8% by these two techniques.

Biomedical signal processing is often plagued by insufficient data, the work of this chapter has made some attempts for the above problem in the seizure detection task. And the proposed techniques are proved to be effective.

Chapter 4 UMLDA for multichannel EEG feature fusion in epileptic seizure prediction

Epileptic Seizures may cause sudden faint, convulsions, and consciousness loss. Due to the sudden and unpredictable nature of the attack, it may bring great harm to the patient's body when engaged in dangerous operations, like driving. Therefore, if the onset of epilepsy can be predicted, even in a short period of time, it can help patients protect themselves and receive timely medical assistance. In addition, this can also help people understand the characteristics of epilepsy signals more clearly and provide ideas for new treatment options.

Epileptic seizures were believed as unpredictable at first, and the corresponding EEG signals only had two states: ictal intervals and interictal intervals. However, more and more studies have shown that the physiological manifestation and signals during seizures change over time. In 1997, 42% of 562 epileptic patients in a study conducted in the United States had symptoms before the onset [59]. Baumgartner et al. proved that the blood flow around the brain changed before the seizure [60]. Heart activity has also been shown to have an early warning effect on seizures [61]. So preictal intervals are proved and presented clinically to mark the duration before seizure onset.

Inspired by clinical evidence, seizure prediction becomes a research hotspot in EEG signal processing. The method of epilepsy prediction is to detect the preictal EEG signal from the long interictal intervals, so that it can warn the arrival of seizure.

In this chapter, we focus on the task of seizure prediction. We present a power extraction algorithm by wavelet transform in time-frequency domain, the feature of two-dimension power spectral density (PSD) is obtained by segment in time and frequency axis based on physiological characteristics. In order to extract more information from the original signal, 23 channels of CHB-MIT dataset are all implemented for PSD extraction. We construct a three-order tensor feature from

multichannel. UMLDA is applied to project the tensor into vector under the constraint of orthogonal dimension in feature vector, which means information is well preserved during the projection. The projection from tensor to vector is also the process of multichannel feature fusion. And final feature vectors are fed into KNN to do seizure prediction. The process of seizure prediction task is plotted in fig 16.

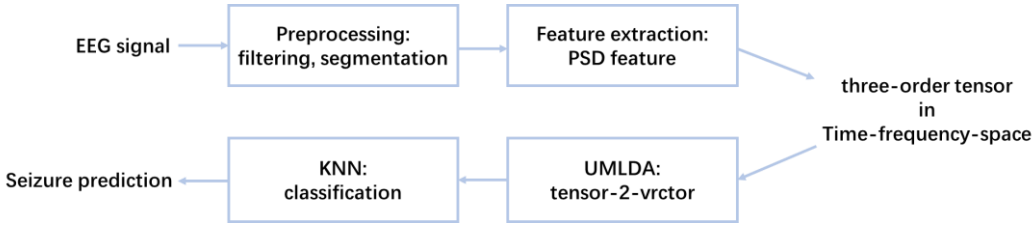


Fig 16. The process of proposed seizure prediction

4.1 Dataset

The EEG dataset used in this task is from CHB-MIT [62]. In the previous chapter, Boon dataset is adopted to apply deep learning model in small data size. Unlike that purpose, we want the continuous EEG signal to segment the preictal intervals, and a large amount of data is preferred. The CHB-MIT EEG dataset meets our needs very well.

The scalp EEG data is collected at the Children's Hospital Boston, it contains 24 child cases suffering from intractable seizures. Subjects are monitored continuously for several days to judge their qualification for surgery. There are 23 subjects and 24 cases in total, case chb23 and case chb01 are from the same subjects with 1.5 years interval. Since the information of case24 is added later and is missing, so the data set includes 5 males and 17 females, ranging in age from 1.5 to 22. Each case has at least 23 channel and the placement of electrodes is in accordance with the 10-20 international standard. And the EEG signal is recorded at 256 Hz with 16-bit resolution. And each case contains different number of seizures, varying from 3 to 40. And each case contains 19

to 40 continuous EEG recording in EDF format. And the time of seizure start and end is described in the additional document. The detail of each case is described in table 6.

Table 6. The detailed description of CHB-MIT dataset

Gender includes: Female (F) and Male (M). Seizure type includes: Complex partial seizure (CP), Simple partial seizure (SP) and Generalized tonic-clonic seizure (GTC).

ID	Gender	Age	Seizure type	Number of EDF	Number of seizures
chb01	F	11	SP,CP	44	7
chb02	M	11	SP,CP,GTC	36	3
chb03	F	14	SP,CP	38	7
chb04	M	22	SP,CP,GTC	42	4
chb05	F	7	CP,GTC	39	5
chb06	F	1.5	CP,GTC	18	10
chb07	F	14.5	SP,CP,GTC	19	3
chb08	M	3.5	SP,CP,GTC	21	5
chb09	F	10	CP,GTC	19	4
chb10	M	3	SP,CP,GTC	25	7
chb11	F	12	SP,CP,GTC	35	3
chb12	F	2	SP,CP,GTC	24	40
chb13	F	3	SP,CP,GTC	33	12
chb14	F	9	CP,GTC	26	8
chb15	M	16	SP,CP,GTC	40	20
chb16	F	7	SP,CP,GTC	19	10
chb17	F	12	SP,CP,GTC	21	3
chb18	F	18	SP,CP	36	6
chb19	F	19	SP,CP,GTC	30	3
chb20	F	6	SP,CP,GTC	29	8
chb21	F	13	SP,CP	33	4
chb22	F	9	\	30	3
chb23	F	6	\	9	7
chb24	\	\	\	22	16

4.2 Data preprocessing

4.2.1 Filtering

We design a finite impulse response (FIR) bandpass filter to remove the noise, artifacts and unwanted high frequency components, and its passband ranges from 0.5 to 40 HZ. Compared with infinite impulse response (IIR) filter, FIR is easier to be optimized and it has a linear phase [63]. The cut-off frequencies of the designed band-pass filter are 0.5 Hz and 40 Hz, the cut-off frequencies of the stop-band are 0.1 Hz and 40.5 Hz, and their corresponding attenuation coefficients are 60db and 80db. The frequency and phase response of designed FIR filter are plotted in fig 17.

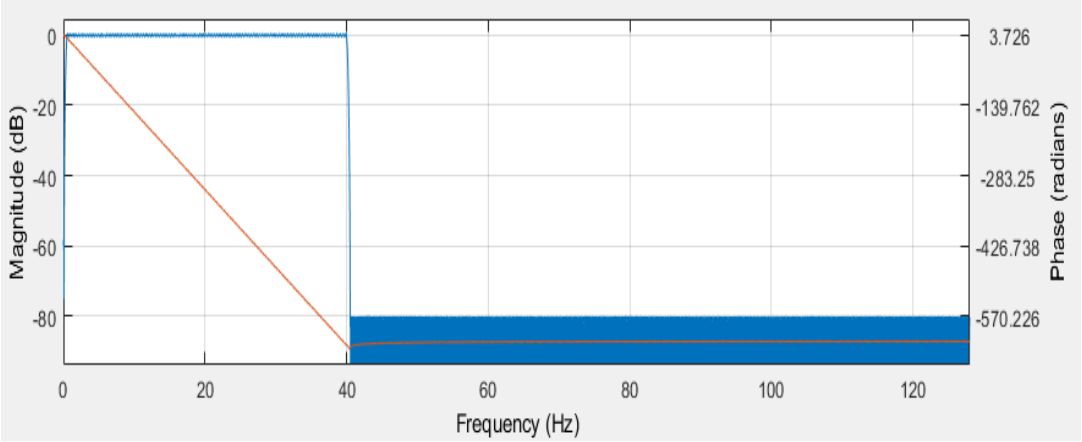


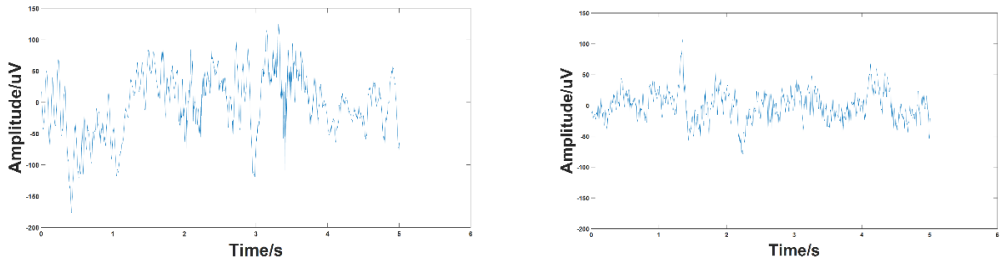
Fig 17. The frequency and phase response of designed FIR filter

▪ 4.2.2 Segmentation and selection

The duration of each EDF file in raw CHB-MIT dataset is a continuous long-term signal around one hour, and we need to slice it to obtain the trials we want according to the needs of seizure prediction, which is the interictal intervals and preictal intervals. The segmentation is also related to the duration of preictal length.

Seizure preictal horizon (SPH) and seizure interictal horizon (SIH) are important parameters to be determined before algorithm implementation [66]. And they stay controversial in different work. Epileptic prediction several hours in advance, like presented in [64-65], seems attractive. But it lacks medical evidence and leads to less available data. So we take 10 minutes duration before seizure onset as SPH, and SIH is determined as more than an hour from seizure onset. Such duration is reasonable for alerting the patients in time and achieving good prediction results. For each patient, we collect 20 minutes of interictal intervals and preictal intervals separately.

Seizures within 10 minutes will be abandoned to keep the integrity of each preictal intervals, so less data is available. In order to augment EEG data, each selected EEG segment is divided into a duration of 4 seconds with 50% overlapping. An example of selected segments is plotted in fig 18.



a. Preictal segment b. Interictal segment

Fig 18. An example of selected preictal and interictal segment

The number of channels varies in the cases. Most EDF files contain 23 channels, and some cases have 24 or 26 channels. In order to maintain consistency between different cases, 23 channels shared by all cases are selected, which are FP1-F7, FP1-F3, FP2-F4, FP2-F8, F3-C3, F4-C4, F7-T7, F8-T8, T7-P7, P3-O1, P4-O2, P7-O1, P7-T7, P8-O2, C3-P3, C4-P4, T8-P8, T8-P8, FZ-CZ, CZ-PZ, T7-FT9, FT9-FT10 and FT10-T8.

4.3 Feature extraction

The electrophysiological performance of epileptic seizures is abnormal excessive discharge of neurons, so the energy features of brain electricity are widely used in epilepsy detection and prediction tasks, and good results have been achieved. Bruno et al. proposed an energy relative measures based on wavelet transform to achieve seizure prediction task [67]. In [68], spectral power and cost-sensitive SVM are combined to predict seizure. Low spectral power and ratios of spectral power are extracted at low complexity to discriminant interictal intervals and preictal intervals [69].

In this section, we proposed an energy extraction algorithm in time-frequency distribution of wavelet transform. First, we calculated the PSD in the two-dimensional distribution after wavelet transform. By segmenting the time and frequency domain, we get a two-dimensional feature distribution, which is basic element of constructed tensor.

▪ 4.3.1 Wavelet transform

The classical Fourier transform can only analyze signals with fixed frequency components, while the time-frequency distribution is to analyze signals whose frequency changes with time. For example, the idea of the short-time Fourier transform is conducting Fourier transform through a sliding window in the time domain, the result represents the instantaneous frequency component when the window is small enough. However, the choice of window width is a problem. If the window is too wide, the time resolution is poor and the spatial resolution is good. If the window is too narrow, the situation is the opposite.

Wavelet transform solves the conflict of time and frequency resolution in STFT by introducing wavelet basis function [70]. Different from the trigonometric function in the Fourier transform, the wavelet function is a finite-length attenuation function, so the length of the sliding window in the wavelet transform can be changed adaptively. It is very suitable for non-stationary random signals like EEG. The equation of wavelet transform is listed below:

$$WT(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \times \varphi(\frac{t-\tau}{a}) dt$$

where $f(t)$ represents the input signal and φ is the mother wavelet. There are two variants in wavelet transform, scale a and translation τ. The scale controls the expansion and contraction of the wavelet function, and the translation amount τ controls the translation of the wavelet function. The scale corresponds to frequency in inverse proportional, and the amount of translation corresponds to time. There are families of mother wavelet for selection. We adopt Daubechies wavelet 4 (Db4) here, because it is proved that it can conserve the energy of signals and redistributes energy into a compact form [71].

▪ 4.3.2 PSD feature extraction

Power spectral density expresses the distribution of signal power at each frequency point [72], it is calculated by following equation:

$$\mathrm{PSD(t},\omega) = \frac{|WV(t,\omega)|^2}{N}$$

where $WV(t,f)$ represents the time-frequency distribution after wavelet transform, and N is the signal length. The distribution of PSD is plotted in fig 19.

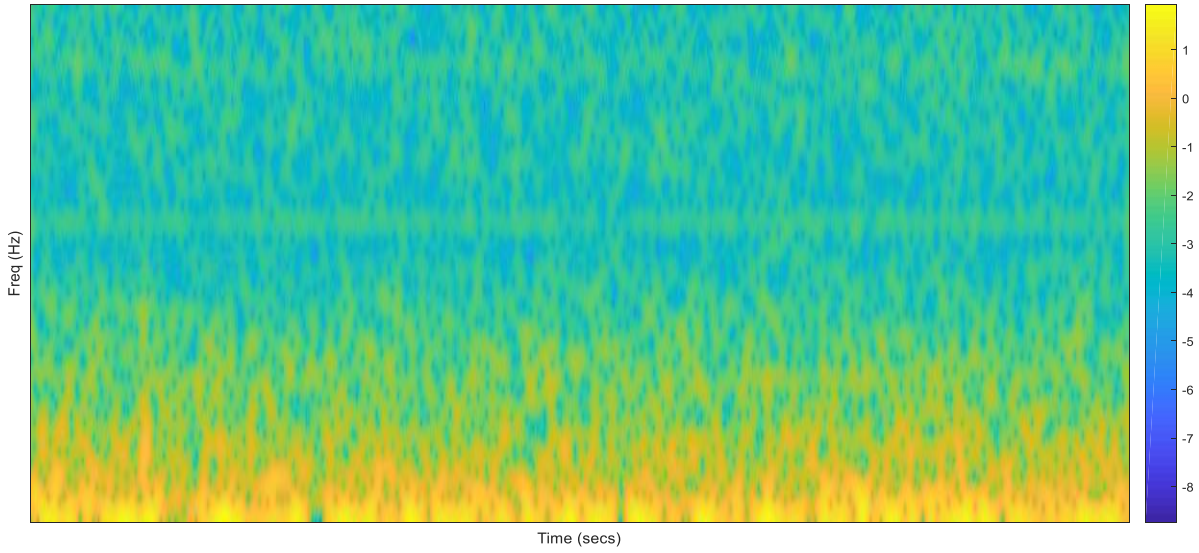


Fig 19. PSD distribution in time-frequency domain

The size of the PSD feature map obtained by wavelet transform is too large to constructed as a tensor, it will lead to large computational complexity and low computational speed. In order to reduce the size of feature map, the 4 second duration is equally divided into 32 segments, and the frequency band ranging from 0 to 40 Hz is divided into 40 segments. So the sum of PSD in each small time-frequency window is calculated as:

$$f(\mathrm{i,j}) = \iint PSD_X d\omega_j dt_i$$

where ω_j represents j_{th} frequency band, t_i represents i_{th} time band, and $f(\mathrm{i,j})$

is the calculation result of PSD sum in each time-frequency window. Because of the discrete characteristic of digital signal, the integral can be modified to accumulate:

$$f(\mathrm{i},\mathrm{j}) = \Sigma_{t\epsilon t_i}\Sigma_{\omega\epsilon\omega_j} PSD_x(t,\omega)$$

where $PSD_x(t,\omega)$ represents the energy at time t and frequency ω in the PSD distribution.

According to the feature map obtained by the above method, the energy difference in grids is relatively large. Because the energy is mainly distributed in the low frequency part. The magnitude difference of the data in the feature map will affect the convergence speed of the subsequent classification model and the accuracy of the result. So normalization of feature is applied here:

$$\mathrm{y} = \frac{x - min}{max - min}$$

where min represents minimum value of points at the same position in feature map of different trials, max represents the maximum value. As a result, the value of output y is normalized to between 0 and 1 [73].

The PSD feature map is extracted in 23 channels, so each trial produces a 23*40*32 three-order tensor.

▪ 4.4 UMLDA for multichannel feature fusion and tensor projection

Channel information is ignored in most EEG signal processing researches, because medical researcher indicates that there are several specific channels related to seizures. However, human cognition of the brain is at a very preliminary stage, and there is still a lot of unknown information to be explored. We take channel information into consideration, it is considered as spatial distribution of EEG signal. So our extracted feature is three-order tensor in temporal, spectral and spatial domain.

It is difficult to classify tensor object directly, because it is computationally complex. And it is proved that tensor recognition suffers from so-called curse of dimensionality [74]. So dimensionality reduction is needed with maximum preservation of information in the original tensor structure. Traditional feature reduction method only handles with vector object which can be also regarded as one-order tensor, such as principle component analysis (PCA) and linear discriminant analysis (LDA). The tensor object can be reshaped to vector to perform these algorithms, but the operation of reshape destroys the spatial structure of the tensor. Multilinear subspace feature extraction methods are introduced to preserve the underlying structure of tensor object. Multilinear principle component analysis (MPCA) and multilinear discriminant analysis (MLDA) are improved version for feature extraction in multichannel signal [75-76]. But they also receive the inherent shortcomings of the method itself. MPCA is an unsupervised method, which cannot be used for classification. And MLDA is a supervised method compared to MPCA, its constraint is weak that dimension independence after projection is not promised. For the reasons mentioned above, UMLDA is proposed to achieve tensor-to-vector (TVP) projection with minimum redundancy.

UMLDA is introduced to is proposed to ensure orthogonal dimensionality of the vectors projected from tensor [77]. The constraint of correlation minimizes the redundancy of the projected vector, which means the original information is well preserved. The calculation of such TVP is an optimization problem, and it is achieved by an algorithm called alternating projection method (APM). Next, we will introduce the principle of the UMLDA and its application in seizure prediction task.

▪ 4.4.1 Linear discriminant analysis (LDA)

LDA is a supervised learning dimensionality reduction technique. The idea of LDA can be summarized as maximizing the variance between classes and minimizing the variance within classes. Specifically, the projection points of same class should be

as close as possible, and the center of projection points of different classes should be as far as possible. The aim of LDA is to find optimal projection vector w, and the LDA projection for two-dimension vector is plotted in fig 20.

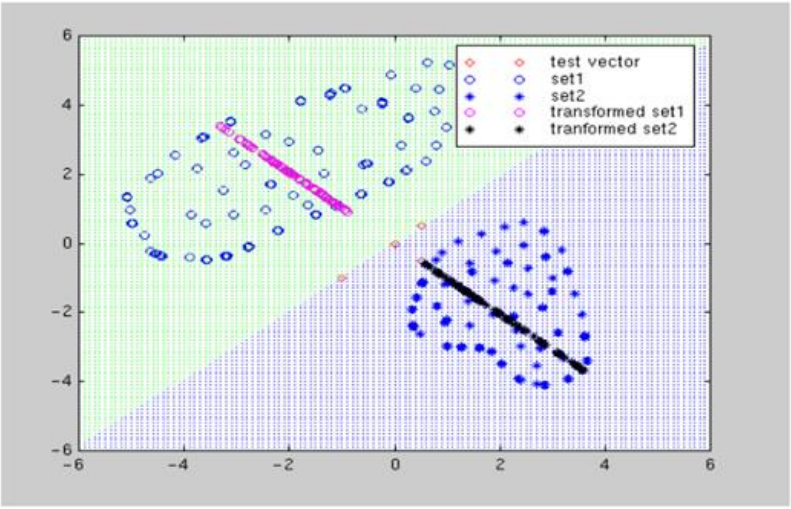


Fig 20. LDA for two-dimension vector [78]

For a given dataset $\{(\boldsymbol{x_i}, y_i)\}_{i=1}^{m}$, we denote X_i $\{i = 0,1\}$ as the set of class i, and N_i is the sample number of class i, then the mean vector μ_i for X_i is calculated as:

$$\mu_i = \frac{1}{N_i} \sum_{x \in X_i} x \ \ (i = 0,1)$$

And the covariance matrix $\sum_i$ for X_i is calculated as:

$$\sum_i = \sum_{x \in X_i} (x - \mu_i)(x - \mu_i)^T \ \ (i = 0,1)$$

After projection by w^T, the mean vector $\widetilde{\mu_i}$ and the covariance matrix $\widetilde{\sum_i}$ is calculated as:

$$\widetilde{\mu_i} = \frac{1}{N_i} \sum_{x \in X_i} w^T x = w^T \mu_i \ \ (i = 0,1)$$

$$\widetilde{\sum_i} = \sum_{x \in X_i} (w^T x - \widetilde{\mu_i})(w^T x - \widetilde{\mu_i})^T = w^T \sum_i w$$

the optimization target J is summarized as below, which is also the ratio of variance between classes and variance within classes.

$$J = \frac{||w^T\mu_0 - w^T\mu_1||^2}{w^T\sum_0 w + w^T\sum_1 w} = \frac{w^T(\mu_0 - \mu_1)(\mu_0 - \mu_1)^T w}{w^T(\sum_0 + \sum_1)w}$$

the within-class scatter matrix S_w is defined as:

$$S_w = \sum_0 + \sum_1 = \sum_{x \in X_0}(x - \mu_0)(x - \mu_0)^T + \sum_{x \in X_1}(x - \mu_1)(x - \mu_1)^T$$

and the between-class scatter matrix S_b is defined as:

$$S_b = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T$$

so the target of LDA is converted into the generalized Rayleigh quotient:

$$argmax_w \ J = \frac{w^T S_b w}{w^T S_w w}$$

The above optimization problem can be solved using Lagrange multiplier method.

▪ 4.4.2 UMLDA

UMLDA is developed based on the concept of LDA and MLDA. It not only maximizes generalized Rayleigh quotient like LDA, but also realizes the goal of uncorrelated feature dimension after tensor-to-vector projection. Orthogonal dimensions indicate that there is no information redundancy, the information contained in tensor object is preserved to the greatest extent during projection.

The process of tensor-to vector is explained in fig 21. The size of extracted tensor feature is 23*40*32, it can be projected into a 40*32*1 matrix by a 1*1*23 vector, then the matrix can be transformed into a 1*32*1 vector through a 40*1*1 vector, and finally a scalar is obtained by projection of a 1*32*1 vector. Such a set of three projection vector is called as elementary multilinear projections (EMP). The tensor can be projected into a 15-dimension vector by 15 sets of EMPs.

In a more general situation, we assume X is a N-order tensor object ($X \in R^{I_1 \mathrm{x} I_2 \mathrm{x} \ldots\ldots I_N}$). According to the above tensor-to-vector projection process, we can project it into a vector object y of P-dimension by P EMPs $\{\boldsymbol{u}_p^{(1)^T}, \boldsymbol{u}_p^{(2)^T}, \ldots, \boldsymbol{u}_p^{(N)^T}\}$, where norm of each projection is normalized as 1 ($|| \boldsymbol{u}_p^{(n)^T} || = 1$ for $p=1,\ldots,P$ and $n=1,\ldots,N$). We denote the training data of tensor features as $\boldsymbol{X}\{X_m, m = 1, \ldots, M\}$, then the result of tensor-to-vector projection $\boldsymbol{y}\{\boldsymbol{y}_m,\ m = 1, \ldots, M\}$ is calculated as $\boldsymbol{y} = \boldsymbol{X} \times \{\boldsymbol{u}_p^{(n)^T}, n = 1, \ldots, N\}_{p=1}^{P}$, among which the p_{th} dimension of projected vector $\boldsymbol{y}$ is calculated by p_{th} EMP, $\boldsymbol{y}(p) = \boldsymbol{X} \times \boldsymbol{u}_p^{(1)^T} \times \boldsymbol{u}_p^{(2)^T} \ldots \times \boldsymbol{u}_p^{(N)^T}$, where $\times$ is mode product between tensor object.

The optimization target in LDA for vector is concluded as maximization of generalized Rayleigh quotient, which is developed as ratio of between-class scatter $S_{B_p}^{y}$ and within-class scatter $S_{W_p}^{y}$ here. The calculation is illustrated below.

$$y_{m_p} = X_m \times \{\boldsymbol{u}_p^{(n)^T}, n = 1, \ldots, N\}$$

$$S_{B_p}^{y} = \sum_{c=1}^{C} N_c \times (\overline{y_{c_p}} - \overline{y_p})^2$$

$$S_{W_p}^{y} = \sum_{m=1}^{M} (y_{m_p} - \overline{y_{c_{m_p}}})^2$$

$$F_p^{y} = \frac{S_{B_p}^{y}}{S_{W_p}^{y}}$$

where C is the category number in samples, N_c is the total number of sample belonged to category c, $\overline{y_{c_p}}$ represents mean output value corresponding to samples belonged to category c, $\overline{y_p}$ is the average of all provided samples and $\overline{y_{c_{m_p}}}$ is the same as $\overline{y_{c_p}}$ if

y_{m_p} is in class c. At last, the modified optimization target for each elementary projection set in UMLDA is listed above.

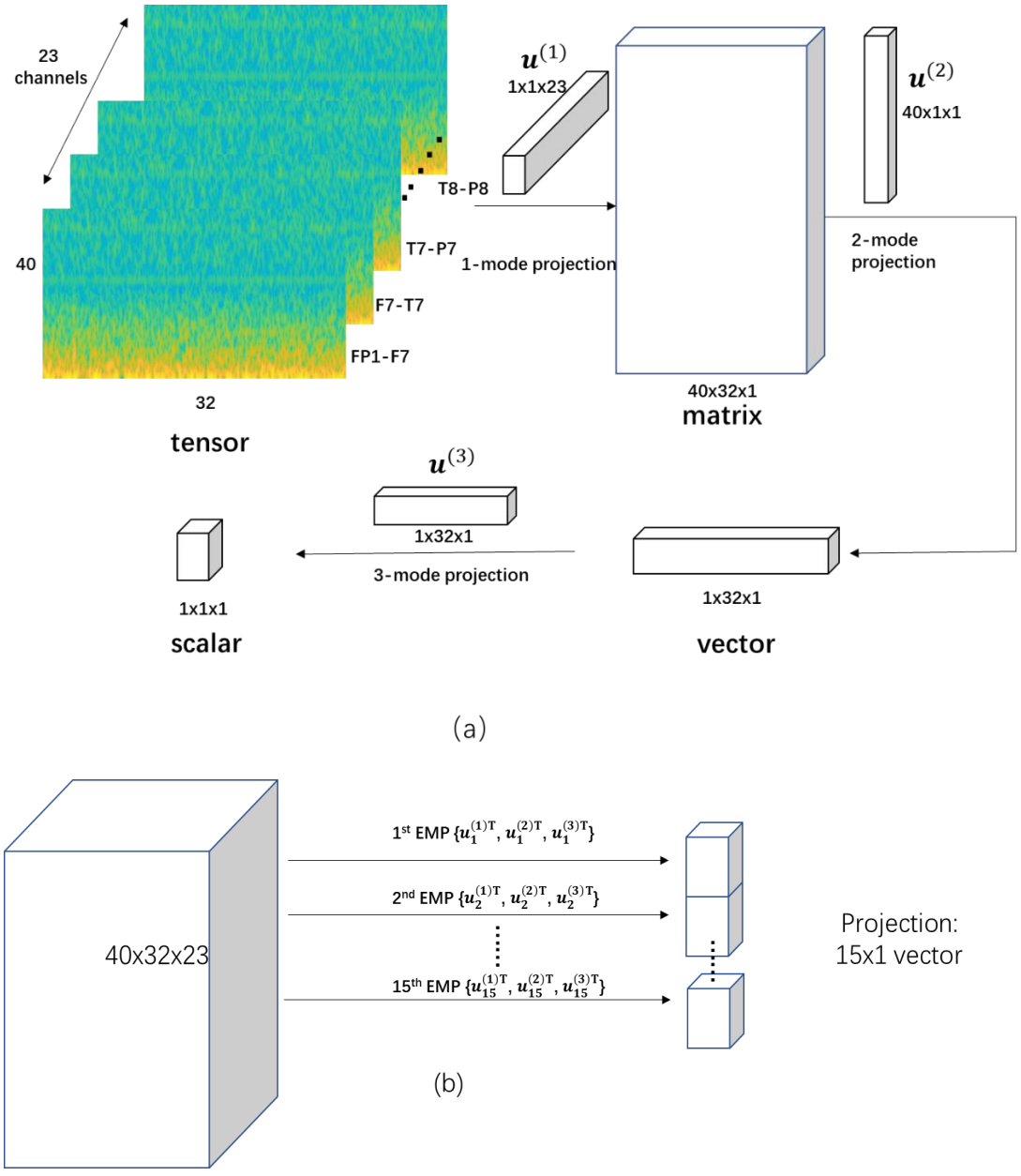


Fig 21. Tensor-to-vector projection in UMLDA

The goal of UMLDA can be summarized as to find a tensor-to-vector projection following the rules in fig 21. We hope it can maximize generalized Rayleigh quotient in each EMP, and the uncorrelated constraint of projected dimensions is satisfied at the same time. So the optimization problem is presented in below mathematical form:

$$p_{th}\ \text{EMP}\ \left\{ \boldsymbol{u}_p^{(n)^T},\ n = 1, \dots, N \right\} = \arg\max F_p^y$$

$$\text{subject to } \frac{\boldsymbol{y}(p)^T \boldsymbol{y}(q)}{||\boldsymbol{y}(p)|| \times ||\boldsymbol{y}(q)||} = 0 \quad when\ p \neq q,\ \ p, q = 1, \dots, P$$

In this nonlinear optimization problem, multiple sets of EMPs projection need to be solved. The calculation complexity of finding the target directly in the solution space is very large. We try to solve the above problems by finding the local optimal solution, and global optimal solution can be approached in loop. Thus, the problem is reduced to a linear optimization problem, such method is called alternating projection method (APM).

For the calculation of p_{th} EMP $\{\boldsymbol{u}_p^{(n)^T}, n = 1, \dots, N\}$, we determine every projection vector $\boldsymbol{u}_p^{(n^*)}$ one by one. We assume that $\boldsymbol{u}_p^{(n^*)}$ is the desired element currently, the rest projection vectors in p_{th} EMP $\{\boldsymbol{u}_p^{(n)^T}, n = 1, \dots, N\ and\ n \neq n^*\}$ project original tensor object into a vector object instead of a scalar object: $\boldsymbol{y}_{m_p}{}^{(n^*)} = X_m \times \boldsymbol{u}_p^{(1)^T} \dots \times \boldsymbol{u}_p^{(n^*-1)^T} \times \boldsymbol{u}_p^{(n^*+1)^T} \dots \times \boldsymbol{u}_p^{(N)^T}$. Then the problem is transformed into a vector optimization problem of between-class scatter $S_{B_p}^y$ and within-class scatter $S_{W_p}^y$, which is the same as LDA. Another constraint is orthogonal between dimensions, so the optimization of p_{th} EMP is converted to:

$$\boldsymbol{u}_p^{n^*} = argmax\ \frac{\boldsymbol{u}_p^{(n^*)^T} \boldsymbol{S}_{B_p}^{(n^*)} \boldsymbol{u}_p^{(n^*)}}{\boldsymbol{u}_p^{(n^*)^T} \boldsymbol{S}_{W_p}^{(n^*)} \boldsymbol{u}_p^{(n^*)}}$$

$$\text{subject to } \boldsymbol{y}(p)^T \boldsymbol{y}(q) = \boldsymbol{u}_p^{n^*} \boldsymbol{Y}_p^{(n^*)} \boldsymbol{y}(q) = 0,\ \ q = 1, \dots, P\ and\ q \neq p$$

The projection result of samples is expressed as $\boldsymbol{Y}_p^{(n^*)} = [\boldsymbol{y}_{1_p}{}^{(n^*)}, \boldsymbol{y}_{2_p}{}^{(n^*)}, \dots, \boldsymbol{y}_{M_p}{}^{(n^*)}]$. If $\boldsymbol{S}_{W_p}^{(n^*)}$ is nonsingular, the solution of this optimization problem happens to be the eigenvector corresponding to the largest eigenvalue of the following equation [77]:

$$\boldsymbol{R}_p^{(n^*)}\boldsymbol{S}_{B_p}^{(n^*)}\boldsymbol{u} \;=\; \lambda \boldsymbol{S}_{W_p}^{(n^*)}\boldsymbol{u}$$

where

$$\boldsymbol{R}_p^{(n^*)} \;=\; \boldsymbol{I}_{n^*} - \boldsymbol{Y}_p^{(n^*)}\boldsymbol{G}_{p-1}\times \left(\boldsymbol{G}_{p-1}^T\,\boldsymbol{Y}_p^{(n^*)^T}\boldsymbol{S}_{W_p}^{(n^*)^{-1}}\,\boldsymbol{Y}_p^{(n^*)}\,\boldsymbol{G}_{p-1}\right)^{-1} \times \boldsymbol{G}_{p-1}^T\,\boldsymbol{Y}_p^{(n^*)^T}\boldsymbol{S}_{W_p}^{(n^*)^{-1}}$$

$$\boldsymbol{G}_{p-1} \;=\; [\boldsymbol{y}(1), \boldsymbol{y}(2), \dots, \boldsymbol{y}(p-1)\,] \;\epsilon\; R^{M\times(\mathrm{p}-1)}$$

Iterative calculation in loop is performed to meet the loss requirement in stopping criterion.

In order to achieve a compromise between computational complexity and information preservation of projection, the dimension of final feature is determined as 15.

▪ 4.5 Principle component analysis (PCA)

We conduct an experiment without UMLDA to verify the performance of proposed tensor-to-vector projection method. The tensor object is flattened into vector object crudely. Since the obtained dimension of feature vector is too large, we applied PCA to achieve dimensionality reduction.

PCA Transform a set of potentially correlated variables into a set of linearly uncorrelated variables through orthogonal transformation, it is an unsupervised dimensionality reduction method.

The parameter goal of PCA optimization is variance. The retention of data information in the dimensionality reduction process is interpreted as the data scattered in the new direction as much as possible. In the case of high dimensions, it is hoped to achieve large variance within the feature dimension and small the covariance between feature dimensions, which corresponds to the diagonal matrix of the covariance matrix.

The process of calculation is PCA is concluded as: 1) normalization: the mean value of samples is shifted to 0, it is necessary because it leads to mean value of 0 in

calculation of covariance matrix. 2) the covariance matrix after projection is calculated as:

$$D = \frac{1}{m} YY^T = \frac{1}{m}(PX)(PX)^T = P\left(\frac{1}{m} XX^T\right)P^T$$
$$= PCP^T$$

where X represents the samples after normalization, P is the projected vector, and C is the covariance matrix for original samples. The eigenvalue decomposition of C is listed:

$$C = E \Lambda E^T$$

where E is the matrix of eigenvectors, and Λ is diagonal matrix of eigenvalues. The equation can be transformed into:

$$E^T C\ E = \Lambda$$

the right side of equation is a diagonal matrix, which is the goal of D. So the projection solution for PCA is the eigenvectors of original covariance matrix.

The elements in the eigenvalue matrix are arranged from large to small, reflecting the importance of the corresponding eigenvectors. The selection of the number of eigenvectors determines the dimension of the output features.

▪ 4.6 KNN for seizure prediction

As for the feature projected by UMLDA, the redundancy is removed maximumly. So in order to highlight the performance of our proposed feature extraction method, we utilize KNN to achieve the seizure prediction task [79]. KNN is known as a supervised learning strategy without training. The idea of the KNN algorithm is selecting the K samples in training set that are most similar to the testing sample, and the category of the testing sample is determined by voting of K samples.

The measure of similarity in KNN algorithm is distance, and the commonly used distance metrics includes Minkowski distance and cosine distance [80].

$$Minkowski\ distance:\ d_{xy} = \sqrt[p]{\sum_{k=1}^{n}(x_k - y_k)^p}$$

Minkowski distance is a generalization of a type of distance measure with variant p. If p is 2, it is the representation of well-known Euclidean distance:

$$\text{Euclidean } distance:\ d_{xy} = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$$

The form of Minkowski distance is called Manhattan distance when p is 1, and it is called Chebyshev distance when p approximates infinite.

And cosine distance is defined as:

$$consine\ distance:\ d(A,B) = 1 - \cos(A,B) = 1 - \frac{A \cdot B}{||A||^2||B||^2}$$

Compared with the numerical measurement of Euclidean distance, cosine distance reflects the difference in direction.

The optimization goal of UMLDA is variance, which is a measure of numerical difference, so Euclidean distance is adopted here and value of K is set as 3.

▪ 4.7 Results and discussion

The subject-independent experiment is performed on 24 cases, and each case contains 20 minutes preictal intervals and interictal intervals. And 70% samples are used as training set for UMLDA, and 30% samples are used as testing set. In addition to the accuracy, we also calculated two other evaluating metrics: Kappa and F1-measure.

$$Kappa:\ \ k = \frac{p_o - p_e}{1 - p_e}$$

the calculation of p_o is the same as accuracy, and p_e is calculated as:

$$p_e = \frac{a_1 * b_1 + a_2 * b_2 + \cdots + a_i * b_i}{n * n}$$

where a_1 is the number of samples in class i, and b_i is the number of samples predicted to be class i, and n is the number of all samples. And F1-measure can be viewed as a tradeoff between precision and recall. It is calculated as:

$$F1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2 * P * R}{P + R}$$

$$Precision: \quad P = \frac{TP}{TP + FP} \qquad Recall: \quad R = \frac{TP}{TP + FN}$$

The result of 24 cases is presented in table 7.

The proposed method of UMLDA achieves an excellent result for seizure prediction task, its average accuracy reaches 95%, Kappa of 0.94 and F1-measure of 0.90. Most of the cases have achieved relatively good results. But case 6 is relatively poor. It is found that the age of the subject is only 1.5 years old, which is the youngest of all subjects. The poor performance of case 6 is also common in other works [84].

It is a quite competitive result of seizure prediction on CHB-MIT dataset. In [81], spatiotemporal relationship of EEG signals based on phase correlation is used to predict seizure, and it achieves an average accuracy of 91.9%. Zhen et al. utilized approximate entropy to achieve 94.5% accuracy in seizure prediction task [82]. And a self-adaption strategy of combining reinforcement learning, online monitoring and adaptive control theory is proposed for seizure prediction task, and it achieves an average accuracy of 71.3% [83].

To test the effectiveness of UMLDA, we repeat the seizure prediction task without UMLDA. The three-order tensor of size 23*40*32 is flattened into a 29440-dimension feature vector. The dimension of flattened feature is too large, even higher than the number of samples, so it is very troublesome to avoid the problem of overfitting, and the calculation for classification is too complex. So we applied PCA to reduce the feature dimension to 15, which equals the number of the output of UMLDA. Take case

1 as an example, the variance explained of top 15 components is plotted in fig 22. And the feature vector is also fed into KNN to do seizure prediction task. The subject-independent experiment is repeated, and the average metrics is listed in table 8.

Table 7. Seizure prediction performance of 24 cases

Case ID	Accuracy	F1-Measure	Kappa
01	0.98	0.98	0.97
02	0.97	0.96	0.93
03	0.96	0.96	0.92
04	0.91	0.91	0.82
05	0.97	0.97	0.94
06	0.82	0.80	0.64
07	0.97	0.97	0.95
08	0.98	0.98	0.97
09	0.95	0.95	0.90
10	0.95	0.95	0.91
11	0.98	0.98	0.96
12	0.91	0.90	0.83
13	0.89	0.89	0.79
14	0.97	0.97	0.95
15	0.97	0.97	0.94
16	0.94	0.94	0.88
17	0.97	0.97	0.94
18	0.85	0.85	0.71
19	0.97	0.97	0.95
20	0.97	0.97	0.94
21	0.91	0.91	0.82
22	0.98	0.98	0.97
23	0.99	0.99	0.98
24	0.94	0.94	0.89
Average	**0.95**	**0.94**	**0.90**

The seizure prediction achieves a poor result when the feature extraction does not have UMLDA. The average accuracy of 23 cases dropped by 16%, F1-measure dropped by 18% and Kappa dropped by 29%. The necessity of UMLDA is illustrated by the

comparative experiments. There may be two reasons for the poor result, one is the structure information loss during flattening. And it can be found in fig 22 that top 15 components only present 46% variance explained. So another reason is that PCA cannot well extract variance feature in limited dimension compared to UMLDA.

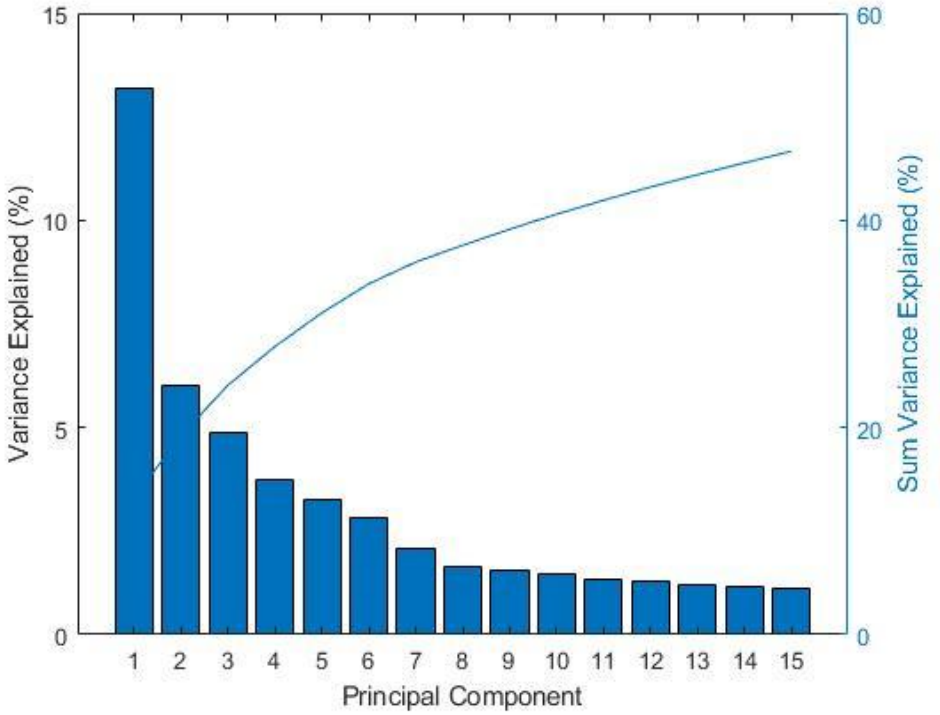


Fig 22. Principal components in PCA of case 1

Table 8. Seizure prediction result without UMLDA

Case ID	Accuracy	F1-Measure	Kappa
Average	0.79	0.76	0.61

4.8 Summary

In this chapter, we propose a seizure prediction algorithm based on time-frequency-space tensor feature construction and tensor-to-vector projection method. We make full use of the time-frequency information and channel information of the EEG signal. By segmentation of time-frequency distribution after wavelet transform, a three-order tensor of PSD feature in temporal, spectral and spatial domain is built. UMLDA is

applied to project the tensor into vector, the information of the tensor structure is well preserved by minimizing the redundancy between dimensions. At the same time, multichannel feature fusion is implemented during the projection. And KNN performs the final classification task for interictal intervals and preictal intervals.

The comparative experiment results exhibit the excellent performance of UMLDA. This chapter provides a new idea for tensor object processing in addition to CNN. And we introduce a new strategy for feature fusion in multichannel EEG, which is tensor-to-vector projection. Compared with the existing multi-channel fusion methods, such as MPCA, MLDA and fully connected layers in CNN, it has its own unique advantages of correlation constraint.

Chapter 5 Summary

The mechanism of the brain's operation has not been fully understood by humans. Humans try to reveal the mysteries of the brain not only through medical research but also through signal processing. Among them, decades of effort have been made to study epileptic EEG signal. Because epilepsy disease troubles a lot of people all over the world. Seizure detection and seizure prediction have attracted a lot of attention. Seizure detection provides aim for automatic diagnosis of epilepsy disease, and seizure prediction forecasts the upcoming seizure onset and provide a time lead for patients to get timely medical aid.

There are many researches for seizure detection and seizure prediction tasks, which processes non-stationary random epileptic EEG signals. And we found that there is still potential for deep learning methods and multi-channel features. But they all face their own challenges in EEG processing.

The difficulty of deep learning is the model overfitting caused by insufficient EEG samples. This is a congenital defect of biomedical signals, because their collection requires strict specifications and they require professional doctors to annotate, so the amount of data is very small in most cases. However, deep learning has powerful fitting capabilities and complex networks, it requires a lot of data to train. We take epileptic seizure detection as an example to try to provide solutions for this contradiction. Different from end-to-end CNN model for classification, we proposed a hybrid model of CNN and SVM. We design a CNN structure with three convolutional layers. The output of the penultimate fully connected layer of CNN is used as feature extraction to avoid overfitting caused by direct classification. The output features of CNN are classified by SVM, which is good at handling signal in small data size. Another proposed technique is new EEG trial generation through decomposition and recombination in the time-frequency domain. Both techniques further improve the accuracy of seizure detection based on initial end-to-end model.

The difficulty for multi-channel feature extraction and fusion is how to preserve spatial information contained in channels during signal processing. Channel information increases the dimensionality of the data, so a combination method of dimensionality reduction and feature fusion is desired. UMLDA is introduced in the thesis to perform seizure prediction task. We construct a three-order tensor of feature by PSD extraction after wavelet transform, and it is projected into vector by UMLDA with correlation constraint. We compared the epilepsy prediction results with or without UMLDA on 24 cases of the CHBMIT data set. The results proved the superior performance of UMLDA in processing tensor signals and feature fusion for multichannel signals.

This thesis aims at the problem of deep learning for seizure detection in small data size and feature fusion for multichannel EEG. The proposed methods are new attempts in EEG signal processing. They provide new ideas for epileptic EEG signal processing.

The thesis still has some content to be improved in the future. Firstly, we want to verify our models and algorithms between subjects, because generalization ability for different individuals is important in practical applications. Besides, the CNN structure adopted in this thesis is determined by empirical approach. The design of the network should be more logical and provides enough theoretical support. This also needs to be well demonstrated in future work.

Reference

[1] Berger H. On the electroencephalogram of man[J]. Electroencephalography and clinical neurophysiology, 1969: Suppl 28: 37+.
[2] Dattola S, Morabito F C, Mammone N, et al. Findings about LORETA Applied to High-Density EEG—A Review[J]. Electronics, 2020, 9(4): 660.
[3] Blume W T, Young G B, Lemieux J F. EEG morphology of partial epileptic seizures[J]. Electroencephalography and clinical neurophysiology, 1984, 57(4): 295-302.
[4] Zhang Z J, Koifman J, Shin D S, et al. Transition to seizure: ictal discharge is preceded by exhausted presynaptic GABA release in the hippocampal CA3 region[J]. Journal of Neuroscience, 2012, 32(7): 2499-2512.
[5] Bladin P F. W. Grey Walter, pioneer in the electroencephalogram, robotics, cybernetics, artificial intelligence[J]. Journal of clinical neuroscience, 2006, 13(2): 170-177.
[6] Hyvärinen A. Survey on independent component analysis[J]. 1999.
[7] Pradhan N, Sadasivan P K, Arunodaya G R. Detection of seizure activity in EEG by an artificial neural network: A preliminary study[J]. Computers and Biomedical Research, 1996, 29(4): 303-313.
[8] Cardoso J F. Infomax and maximum likelihood for blind source separation[J]. IEEE Signal processing letters, 1997, 4(4): 112-114.
[9] Jung T P, Humphries C, Lee T W, et al. Extended ICA removes artifacts from electroencephalographic recordings[C]//Advances in neural information processing systems. 1998: 894-900.
[10] Winkler I, Haufe S, Tangermann M. Automatic classification of artifactual ICA-components for artifact removal in EEG signals[J]. Behavioral and brain functions, 2011, 7(1): 30.
[11] Geethanjali P, Mohan Y K, Sen J. Time domain feature extraction and classification of EEG data for brain computer interface[C]//2012 9th International Conference on Fuzzy Systems and Knowledge Discovery. IEEE, 2012: 1136-1139.
[12] Greene B R, Faul S, Marnane W P, et al. A comparison of quantitative EEG features for neonatal seizure detection[J]. Clinical Neurophysiology, 2008, 119(6): 1248-1261.
[13] Harati A, Golmohammadi M, Lopez S, et al. Improved EEG event classification using differential energy[C]//2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE, 2015: 1-4.
[14] O'Toole J M, Temko A, Stevenson N. Assessing instantaneous energy in the EEG: a non-negative, frequency-weighted energy operator[C]//2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2014: 3288-3291.
[15] Griffin D, Lim J. Signal estimation from modified short-time Fourier transform[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(2): 236-243.
[16] Shensa M J. The discrete wavelet transform: wedding the a trous and Mallat algorithms[J]. IEEE Transactions on signal processing, 1992, 40(10): 2464-2482.
[17] Zhang B, Sato S. A time-frequency distribution of Cohen's class with a compound kernel and its application to speech signal processing[J]. IEEE transactions on signal processing, 1994, 42(1): 54-64.
[18] Tzallas A T, Tsipouras M G, Fotiadis D I. Epileptic seizure detection in EEGs using time–frequency analysis[J]. IEEE transactions on information technology in biomedicine, 2009, 13(5): 703-710.
[19] Pincus S M. Approximate entropy as a measure of system complexity[J]. Proceedings of the National Academy of Sciences, 1991, 88(6): 2297-2301.
[20] Bandt C, Pompe B. Permutation entropy: a natural complexity measure for time series[J]. Physical review letters, 2002, 88(17): 174102.
[21] Kosko B. Fuzzy entropy and conditioning[J]. Information sciences, 1986, 40(2): 165-174.

[22] Fadlallah B, Chen B, Keil A, et al. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information[J]. Physical Review E, 2013, 87(2): 022911.
[23] Li P, Liu C, Li K, et al. Assessing the complexity of short-term heartbeat interval series by distribution entropy[J]. Medical & biological engineering & computing, 2015, 53(1): 77-87.
[24] Ferraz M S A, Kihara A H. Hurst entropy: A method to determine predictability in a binary series based on a fractal-related process[J]. Physical Review E, 2019, 99(6): 062115.
[25] Huang N E, Wu M L, Qu W, et al. Applications of Hilbert–Huang transform to non-stationary financial time series analysis[J]. Applied stochastic models in business and industry, 2003, 19(3): 245-268.
[26] Oweis R J, Abdulhay E W. Seizure classification in EEG signals utilizing Hilbert-Huang transform[J]. Biomedical engineering online, 2011, 10(1): 38.
[27] Raudys S J, Jain A K. Small sample size effects in statistical pattern recognition: Recommendations for practitioners[J]. IEEE Transactions on pattern analysis and machine intelligence, 1991, 13(3): 252-264.
[28] James G M. Variance and bias for general loss functions[J]. Machine learning, 2003, 51(2): 115-135.
[29] Zhou W, Liu Y, Yuan Q, et al. Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG[J]. IEEE Transactions on Biomedical Engineering, 2013, 60(12): 3375-3381.
[30] Garcia G N, Ebrahimi T, Vesin J M. Support vector EEG classification in the Fourier and time-frequency correlation domains[C]//First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings. IEEE, 2003: 591-594.
[31] Noble W S. What is a support vector machine?[J]. Nature biotechnology, 2006, 24(12): 1565-1567.
[32] Li S, Zhou W, Yuan Q, et al. Feature extraction and recognition of ictal EEG using EMD and SVM[J]. Computers in biology and medicine, 2013, 43(7): 807-816.
[33] Guler I, Ubeyli E D. Multiclass support vector machines for EEG-signals classification[J]. IEEE transactions on information technology in biomedicine, 2007, 11(2): 117-126.
[34] Zhong S, Ghosh J. HMMs and coupled HMMs for multi-channel EEG classification[C]//Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290). IEEE, 2002, 2: 1154-1159.
[35] Hu J, Min J. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model[J]. Cognitive neurodynamics, 2018, 12(4): 431-440.
[36] Mohammadi Z, Frounchi J, Amiri M. Wavelet-based emotion recognition system using EEG signal[J]. Neural Computing and Applications, 2017, 28(8): 1985-1990.
[37] Schirrmeister R T, Springenberg J T, Fiederer L D J, et al. Deep learning with convolutional neural networks for EEG decoding and visualization[J]. Human brain mapping, 2017, 38(11): 5391-5420.
[38] Tabar Y R, Halici U. A novel deep learning approach for classification of EEG motor imagery signals[J]. Journal of neural engineering, 2016, 14(1): 016003.
[39] Tsiouris K M, Pezoulas V C, Zervakis M, et al. A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals[J]. Computers in biology and medicine, 2018, 99: 24-37.
[40] Hartmann K G, Schirrmeister R T, Ball T. EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals[J]. arXiv preprint arXiv:1806.01875, 2018.
[41] Andrzejak R G, Lehnertz K, Mormann F, et al. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state[J]. Physical Review E, 2001, 64(6): 061907.
[42] Saab M E, Gotman J. A system to detect the onset of epileptic seizures in scalp EEG[J]. Clinical Neurophysiology, 2005, 116(2): 427-442.
[43] Sarle W S. Neural networks and statistical models[J]. 1994.
[44] Chua L O, Roska T. The CNN paradigm[J]. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 1993, 40(3): 147-156.

[45] Xu B, Wang N, Chen T, et al. Empirical evaluation of rectified activations in convolutional network[J]. arXiv preprint arXiv:1505.00853, 2015.
[46] Dunne R A, Campbell N A. On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function[C]//Proc. 8th Aust. Conf. on the Neural Networks, Melbourne. Citeseer, 1997, 181: 185.
[47] Browne M W. Cross-validation methods[J]. Journal of mathematical psychology, 2000, 44(1): 108-132.
[48] Zhao X, Zhang R, Mei Z, et al. Identification of Epileptic Seizures by Characterizing Instantaneous Energy Behavior of EEG[J]. IEEE Access, 2019, 7: 70059-70076.
[49] Diykh M, Li Y, Wen P. Classify epileptic EEG signals using weighted complex networks based community structure detection[J]. Expert Systems with Applications, 2017, 90: 87-100.
[50] Zeng M, Zhao C, Meng Q H. Detecting seizures from EEG signals using the entropy of visibility heights of hierarchical neighbors[J]. IEEE Access, 2019, 7: 7889-7896.
[51] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural processing letters, 1999, 9(3): 293-300.
[52] Everett III H. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources[J]. Operations research, 1963, 11(3): 399-417.
[53] Wu H C. The Karush–Kuhn–Tucker optimality conditions in an optimization problem with interval-valued objective function[J]. European Journal of Operational Research, 2007, 176(1): 46-59.
[54] Soman K P, Loganathan R, Ajay V. Machine learning with SVM and other kernel methods[M]. PHI Learning Pvt. Ltd., 2009.
[55] Chakraborty R, Garain U. Role of synthetically generated samples on speech recognition in a resource-scarce language[C]//2010 20th International Conference on Pattern Recognition. IEEE, 2010: 1618-1621.
[56] Mouchère H, Bayoudh S, Anquetil E, et al. Synthetic on-line handwriting generation by distortions and analogy[C]. 2007.
[57] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 60.
[58] Lotte F. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces[J]. Proceedings of the IEEE, 2015, 103(6): 871-890.
[59] Rajna P, Clemens B, Csibri E, et al. Hungarian multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures[J]. Seizure, 1997, 6(5): 361-368.
[60] Baumgartner C, Serles W, Leutmezer F, et al. Preictal SPECT in temporal lobe epilepsy: regional cerebral blood flow is increased prior to electroencephalography-seizure onset[J]. Journal of Nuclear Medicine, 1998, 39(6): 978-982.
[61] Novak V, Reeves A L, Novak P, et al. Time-frequency mapping of R–R interval during complex partial seizures of temporal lobe origin[J]. Journal of the autonomic nervous system, 1999, 77(2-3): 195-202.
[62] Shoeb A H. Application of machine learning to epileptic seizure onset detection and treatment[D]. Massachusetts Institute of Technology, 2009.
[63] Rabiner L R, Kaiser J F, Herrmann O, et al. Some comparisons between FIR and IIR digital filters[J]. Bell System Technical Journal, 1974, 53(2): 305-331.
[64] Alotaiby T N, Alshebeili S A, Alotaibi F M, et al. Epileptic seizure prediction using CSP and LDA for scalp EEG signals[J]. Computational intelligence and neuroscience, 2017, 2017.
[65] Chu H, Chung C K, Jeong W, et al. Predicting epileptic seizures from scalp EEG based on attractor state analysis[J]. Computer methods and programs in biomedicine, 2017, 143: 75-87.
[66] Khan H, Marcuse L, Fields M, et al. Focal onset seizure prediction using convolutional networks[J]. IEEE Transactions on Biomedical Engineering, 2017, 65(9): 2109-2118.
[67] Direito B, Dourado A, Vieira M, et al. Combining energy and wavelet transform for epileptic seizure prediction in an advanced computational system[C]//2008 International Conference on BioMedical Engineering and Informatics. IEEE, 2008, 2: 380-385.

[68] Park Y, Luo L, Parhi K K, et al. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines[J]. Epilepsia, 2011, 52(10): 1761-1770.
[69] Zhang Z, Parhi K K. Low-complexity seizure prediction from iEEG/sEEG using spectral power and ratios of spectral power[J]. IEEE transactions on biomedical circuits and systems, 2015, 10(3): 693-706.
[70] Daubechies I. The wavelet transform, time-frequency localization and signal analysis[J]. IEEE transactions on information theory, 1990, 36(5): 961-1005.
[71] Mohamed M, Deriche M. An approach for ECG feature extraction using daubechies 4 (DB4) wavelet[J]. International Journal of Computer Applications, 2014, 96(12): 36-41.
[72] CusidÓCusido J, Romeral L, Ortega J A, et al. Fault detection in induction machines using power spectral density in wavelet decomposition[J]. IEEE Transactions on Industrial Electronics, 2008, 55(2): 633-643.
[73] Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems[J]. Pattern recognition, 2005, 38(12): 2270-2285.
[74] Friedman J H. On bias, variance, 0/1—loss, and the curse-of-dimensionality[J]. Data mining and knowledge discovery, 1997, 1(1): 55-77.
[75] Lu H, Plataniotis K N, Venetsanopoulos A N. MPCA: Multilinear principal component analysis of tensor objects[J]. IEEE transactions on Neural Networks, 2008, 19(1): 18-39.
[76] Ye J, Janardan R, Li Q. Two-dimensional linear discriminant analysis [C]//Advances in neural information processing systems. 2005: 1569-1576.
[77] Lu H, Plataniotis K N, Venetsanopoulos A N. Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition[J]. IEEE Transactions on Neural Networks, 2008, 20(1): 103-123.
[78] Balakrishnama S, Ganapathiraju A. Linear discriminant analysis-a brief tutorial[C]//Institute for Signal and information Processing. 1998, 18(1998): 1-8.
[79] Dudani S A. The distance-weighted k-nearest-neighbor rule[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1976 (4): 325-327.
[80] Singh A, Yadav A, Rana A. K-means with Three different Distance Metrics[J]. International Journal of Computer Applications, 2013, 67(10).
[81] Parvez M Z, Paul M. Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2015, 24(1): 158-168.
[82] Zhang Z, Chen Z, Zhou Y, et al. Construction of rules for seizure prediction based on approximate entropy[J]. Clinical Neurophysiology, 2014, 125(10): 1959-1966.
[83] Wang S, Chaovalitwongse W A, Wong S. A novel reinforcement learning framework for online adaptive seizure prediction[C]//2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2010: 499-504.
[84] Zhang Y, Guo Y, Yang P, et al. Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network[J]. IEEE Journal of Biomedical and Health Informatics, 2019, 24(2): 465-474.
[85] Acharya U R, Oh S L, Hagiwara Y, et al. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals[J]. Computers in biology and medicine, 2018, 100: 270-278.