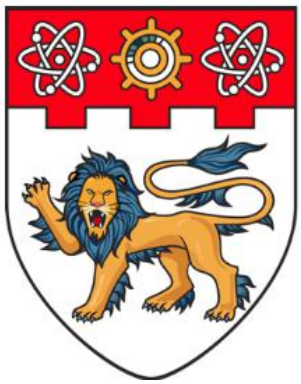

NANYANG TECHNOLOGICAL UNIVERSITY

SINGAPORE

Cascade of Classifiers to Classify Interictal Electroencephalograms of Patients with Epilepsy

Submitted by: Jiang Zhubo

Matriculation Number: G1601840F

Supervisor: Assoc. Prof. Justin Dauwels

SCHOOL OF ELECTRICAL AND ELECTRONIC

ENGINEERING

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN EEE

2017

TABLE OF CONTENTS

LIST OF FIGURES

\

LIST OF TABLES

ABSTRACT

Epilepsy is a chronic disease influencing many people's health worldwide. According to the study of the WHO, there are over 50 million epilepsy patients around the world.

Now, electroencephalogram (EEG) is still a primary method to analyze epilepsy. Experts can detect epilepsy by visual analysis of EEGs, which records the electrical signals of the human brain. Epileptiform transients (ET) or spikes usually appear in the EEG of epileptic patients. The spikes are the main indicatosr for epilepsy. However, detecting epilepsy by only visual inspection may need much time, and there is a lack of experts who can read EEGs. Moreover, there is no standard definition for spikes, which makes the spike detection based diagnosis of epilepsy, tedious and expert-centered. Experts do not always agree on which waveforms are spikes and which ones are not. Hence, an automated method for analysis of epileptic patients' EEG data is of importance for management and diagnosis of epilepsy.

Many methods have been applied to detect the spikes such as template matching, neural network, SVM or random forest. In this thesis, we develop an efficient classification method to eliminate most background waveforms through an effective cascade of classifiers. A cascade of winning classifiers is designed to reject most background waveform for EEG data in several consecutive stages, while prereserving most spikes.

Validating a classification method needs sufficiently large data. We have used 93 epileptic patients' EEG data from Massachusetts General Hospital, which include 18164 spikes in total. We apply the 10-step cascade of decision tree, random forest and (support vector machine) SVM separately to the data by applying cross validation. In the numerical tests of this study, on average, the cascade of decision tree rejected 98.94%% of all background in the EEG dataset while preserving

86.22% of the spikes. The cascade of SVM rejected 98.89% of all background in the EEG dataset while preserving 86.97% of the spikes. The cascade of random forest rejected 98.84% of all background in the EEG dataset while preserving 87.32% of the spikes.

ACKNOWLEDGEMENT

The project of master coursework program provided the author with a quite great opportunity to research the machine learning ways and apply what have learned into practice. Furthermore, it offered the author a chance to master the skills of using machine learning ways to do classification and prediction.

Hereby, the author sincerely wants to say thanks to Prof. Justin Dauwels, who is Associate Professor of School of Electrical & Electronic Engineering in Nanyang Technological University, for his supervision, continuous support, and valuable suggestions.

The author would also like to extend gratitude and appreciation towards Elham Bagheri, Ph.D. student of Associate Professor Justin Dauwels, and Dr. Jing Jin, research fellow, for their generous help and insightful discussions.

The author would also like to thank School of Electrical & Electronic Engineering for providing this chance of joining the program and offering support throughout the project.

Lastly, the author would like to extend great appreciation to all fellow students for their encouragement and help during the project.

Chapter 1

Introduction

1.1 Background of Epilepsy

Epilepsy is characterized by recurrent and sudden seizures, which is born from fiercely electrical discharges in brain cells. Seizures are brief episodes of involuntary and quick movement, like hyperemia or body angulation, that may mainly involve a part of the body or even the entire body. Seizures vary in happening frequency, from several times per day to one or two times per year [1] [2]. The major causes of epilepsy include brain malformations, tumors, infections in infants, head trauma and congenital conditions in adults, stroke and Alzheimer's disease in aged people. The health risks associated with epilepsy include sleeping problems, osteoporosis, reproductive problems, anxiety, and depression. Epileptiform transients (ET) or spikes are biomarkers of epilepsy. Their presence is predictive of recurrent seizure in patients after the first seizure, and epilepsy often occurs between seizures (interictal). One seizure does not signify epilepsy, it is defined as having 2 or more unprovoked seizures [4].

Currently, approximately 50 million people in the world suffer from epilepsy. The estimated proportion of the global population with active epilepsy at presence is between 4 and 10 per 1000 people in the world. Globally, over 2.4 million new patients are diagnosed with epilepsy every year. [3].

1.2 Background of EEG

Billions of neurons and cells comprise the brain. These highly specialized cells can make neurons prone to chemical and electrical stimulation easily by scientists. The data that passes through the Central Nervous System (CNS) is controlled and processed by various neurons only, and we can acquire these data by Electroencephalography (EEG) [5]. Many small electrodes are placed on the head scalp in order to record the current brain electrical signals due to voltage fluctuations of the neurons in the brain, as Figure 1.1 shows.

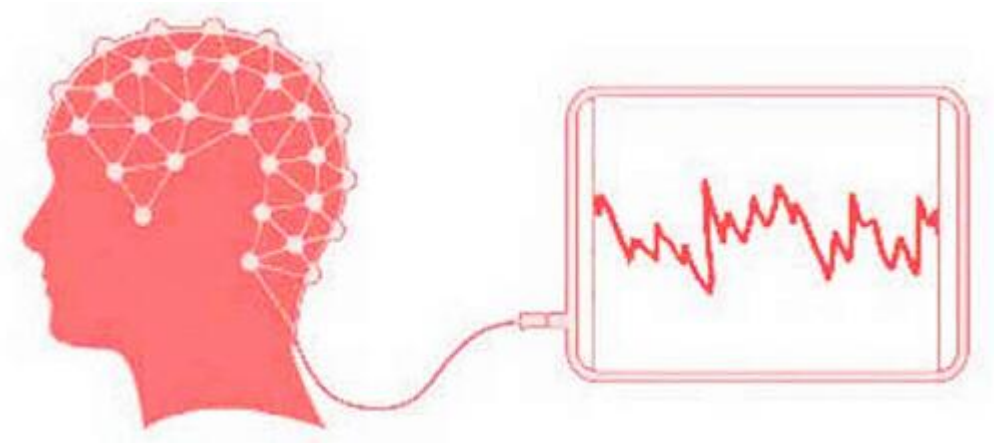

Figure 1.1: EEG data acquisition. [3]

EEG is the safest and most utilized technological way for detecting the functions of a brain. A waveform of EEG implicates an enormous summation of electric potentials that initiate from numerous neurons in the surroundings of the electrodes [6]. A typical adult human scalp EEG signal ranges from 10 μV to 100 μV in amplitude. A picture of EEG data is showed in the Figure 1.2.

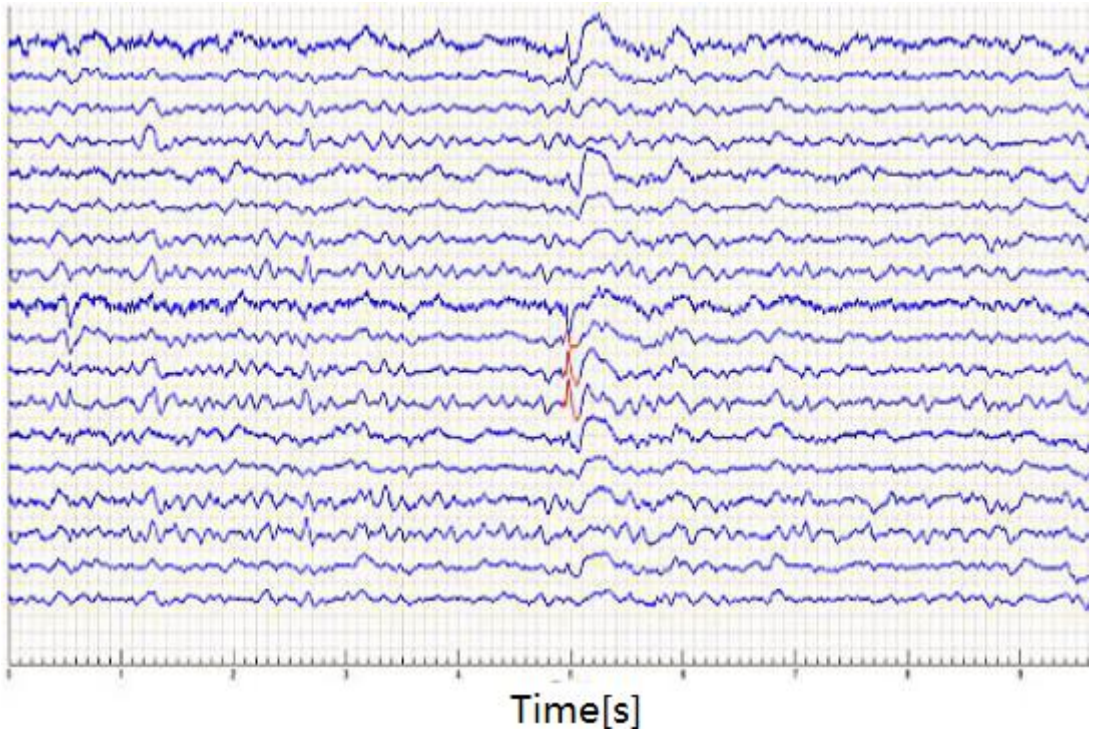

Figure 1.2: A EEG data example. [6]

Epileptiform transients (ETs) characterize EEG data of patients with epilepsy. In EEG of the epileptic patient, 20-70ms length or 70-200ms length of sharp waves (spikes) with pointed peaks can be found obviously if the EEG records the patients' brain electrical signals when they are in the condition of happening seizures. Figure 1.3 shows a picture of spikes in EEG data.

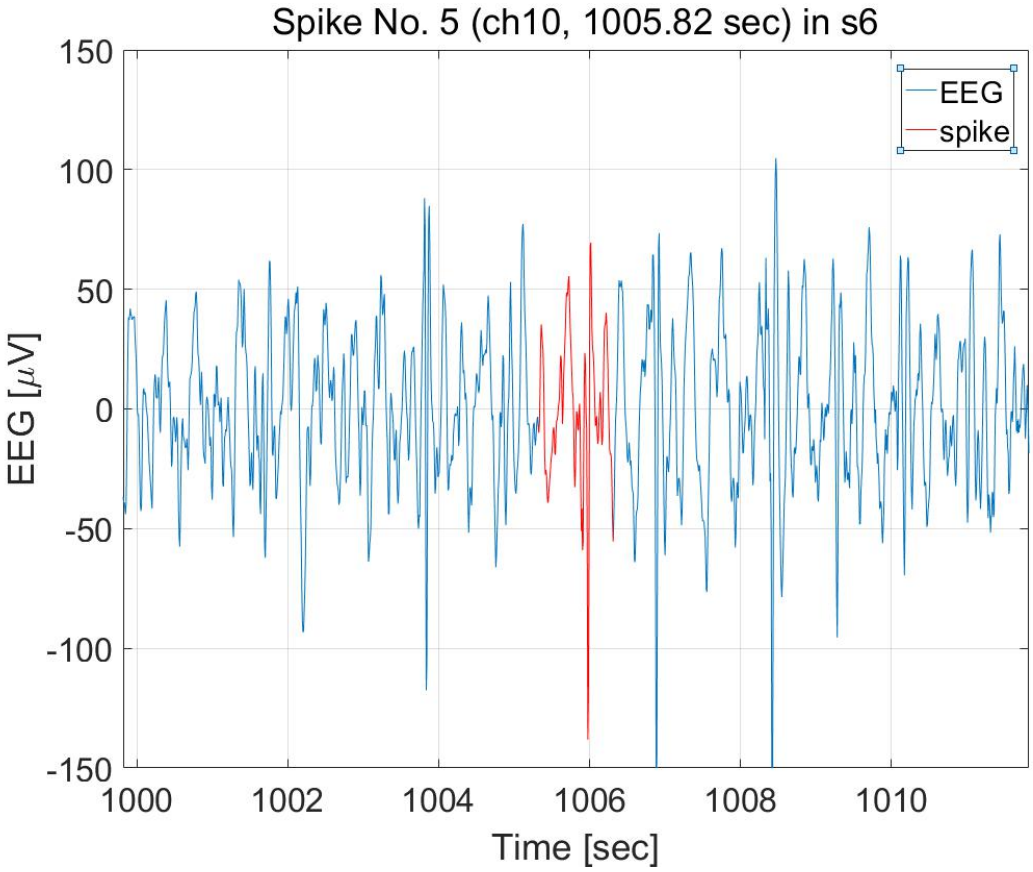

Figure 1.3: An epilepsy EEG example (a waveform in the red curve is a spike).

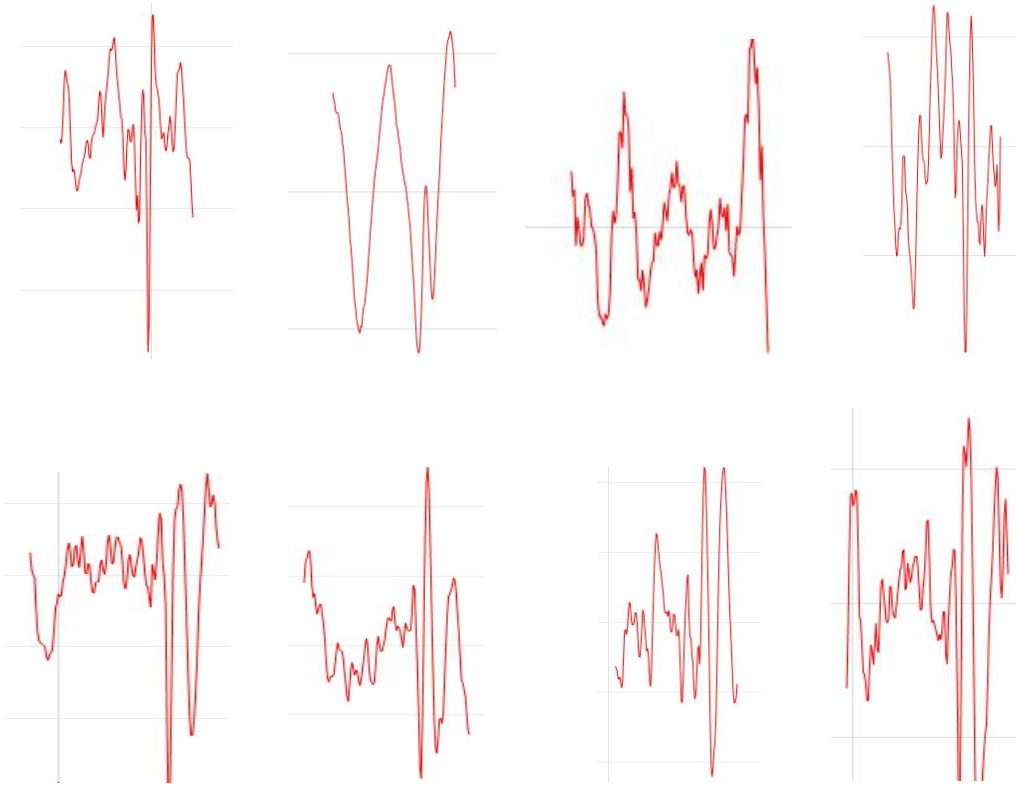

Figure 1.4: The morphological examples of spikes.

In Figure 1.4, we can find that spikes are of a huge variety in morphology. Thus, it is not easy to detect spikes in an automatic manner. Many experts have applied different ways, like deep dearning or neural network [7]. To find a spike detector with a good performance, a huge dataset is what we need to evaluate the detector. So far, the most significant problem is that a huge database of various different spike profiles is not available to validate the results, therefore the results are not reliable. The reason of lack of such database is huge cost for experts to annotate the EEG data. Since there is not any scientific conclusion for spikes occurring in most patients, even the same subject changes much from time to time. Thus, EEG interpretation still relies on EEG experts in short supply. Gold standard is still manual annotation which cost much time and is tedious.

In most of the cases, a continuous and long recording of the EEG is needed to be

analyzed. A general EEG recording is ambulatory recording that contains very long EEG data and meanwhile, it may cost much effort to trace the spikes of the epilepsy.

EEG diagnosis is an indispensable index for diagnosis of epilepsy. Some non-specific abnormalities of EEG are often considered as the basis for the diagnosis of epilepsy. For example, physical examination is also helpful, which includes general medical system examiner and neurological examination. Physical examination focuses on the nervous system, patient's mental state and intelligence, patient's speech, checking the eyes' motion. And CT examination greatly improves the diagnosis of intracranial lesions, which are important means of diagnosis of epilepsy. According to the large case data report, non-selective epileptic patients CT examination of the positive rate of brain lesions is generally more than 50%. .

1.3 Background of Machine Learning

Machine learning theory is mainly to design and analyze methods that computer can automatically "learn" an algorithm. The algorithm is a kind of automatic analysis from the data to obtain the basic law, and it uses the law of unknown data to predict the algorithm. Because the learning algorithm involves a large number of statistical theory, machine learning and inferred statistics are particularly similar, which is also known as a statistical learning theory. Algorithm design and machine learning theory can be concerned with the effective learning algorithm. Many of the problems show no basic law, so some of the machine learning studies can fix them with the approximation algorithm.

Learning strategies refer to the reasoning strategies used in the learning process. A learning system is always composed of two parts: learning and environment [14]. The environment provides the basic information, learning part converses information to easily understanding form, which is used to obtain useful information. The classification of learning strategies is based on the amount of reasoning and difficulty

of converting the information which can be classified. Following is the six basic types, more detailed information can be found in [15] [16] [17]:

（1） Rote Learning

Rote learning: Learners do not need reasoning or other knowledge to learn outside information, but directly absorb the information provided by the environment. Such as Samuel's Checkers program, Simon's LT system. This kind of learning system mainly considers how to index the stored knowledge and make use of it. The learning method of the system is to learn directly from pre-programmed and constructed programs. Learners do not need to do any work or learning and do not make any reasoning on the input information.

（2）Learning From Instruction

Learners acquire information from the environment, converting knowledge into internal representations, and combining new knowledge and original knowledge into a new one. So a learner has to do some reasoning, but the environment still needs to do a lot of work. 'Teachers' in some form will put forward and organize knowledge, so that learners can continue to increase the knowledge. The task of learning is to establish a system so that it can accept the knowledge and suggestions, and store and apply the learned knowledge effectively.

（3）Learning By Deduction

The reasoning form used by the learner is deductive reasoning. They do reasoning from the axiom, deriving conclusions through logical transformation. This reasoning is a specialization of the process so that learners in the process of reasoning can get useful knowledge.

（4）Explanation-Based Learning

Based on the concept of the goal provided by the teacher, like an example of the concept, the domain theory, and the operational criteria, Learners firstly constructs

an explanation to the reason why the example satisfies the goal concept and then extends the explanation to the goal of satisfying the operational criteria condition.

（5）Learning From Induction

Inductive learning is a case or an example of a concept that is provided by a 'teacher' or the environment, making the learners deduce a general description of the concept through inductive reasoning. To a certain extent, the work load of reasoning of inductive learning is also larger than that of analogy. It has been widely studied and applied in the field of artificial intelligence.

Supervised learning is one of basic classification learning model, it refers to the process of adjusting the parameters of a classifier using a set of known categories of samples to achieve the desired performance, also known as supervisory training.[18]

In machine learning, supervised learning model and unsupervised learning model are two common learning models. In a supervised learning model, the training data classes or attributes are already known. What we need to do is to analyze the inner property and obtain a classification rule in statistics or an inferred function. Then we apply this rule or function to map the unknown data.

The process of constructing a classification model is generally separated into two stages: training and testing. Before constructing the classification model, it is required to divide randomly dataset into training dataset and test dataset. \ Since the class designation of each training sample is provided, this stage is also referred to as a guided learning. Usually, this kind of model is provided in the form of a classification rule, which contains many classification algorithms and formulas. In this study, we apply decision tree, SVM and random forest to be the classification model, they are all supervised learning models. In the testing phase, the test dataset is used to evaluate the performance of the classification model. If the final performance of the model is considered acceptable, the model can be used to classify other dataset . In general, the cost of the test phase is far below the training phase.

There are many supervised learning classification models, like KNN, decision tree, Naive Bayes, SVM, random forest, etc. And these models are widely used in detecting the spikes of EEG. Nurettin Acir tried a two-stage procedure to automatically detect spikes of EEG with SVM [19]. J.Jose used fuzzy KNN and biorthogonal wavelet to detect the seizures of EEG [20]. K.Polat [21] did classification for epileptiform EEG with a hybrid system based on the decision tree.

Guleretal [22] applied recurrent neural networks to classify spikes and backgrounds of EEG data with 96.79% accuracy. Kannathal [23] took advantage of the ANFIS classifier to detect EEG seizures with an accuracy of 95%.

In order to improve the accuracy of the classification, the data is usually pre-processed before classification, including:

(1) Data Clean-up: The aim is to eliminate or reduce data noise and handle vacancy values.

(2) Correlation Analysis: Since some features in the dataset may not be relevant to the objective, especially noises or disturbance, the inclusion of these features will slow down and may lead the learning process to a wrong direction. Removing these irrelevant or redundant attributes is the aim of correlation analysis.

(3) Data Transformation： Data can be projected to higher-level dimension, which may be easier to be classified.

1.3 Motivation

There are many potential epileptic patients around the world, and EEG is still the primary way for analyzing and management of epilepsy. However, detecting spikes by visual inspection is tedious. Thus an automated detection system is needed. Actually, many algorithms have been tested in the literature to make a reliable automated spike detection system. And the lack of a sizable database for training and evaluating the performance of the detection models and the lack of a sufficient number of EEG interpreters are two main challenges in the current situations. A good automated spike detection system can help the early spike diagnosis of epilepsy. In this research, the motivation is to develop an automated and efficient diagnostic system for epilepsy. We study a spike database from the EEG data of 93 patients recorded and annotated at Massachusetts General Hospital (MGH). The database consists of 18164 spikes annotated by two neurologists independently. Because of the difficulties in spike detection, only one classifier might not solve the classification problem for EEG data. And in order to reserve spikes and reject backgrounds as much as possible, a cascade of classifiers can help in classification.

1.4 Objective

Considering the motivation, the objective of this project is to develop a cascade of classifiers to reserve spikes and reject backgrounds for test EEG data as many as possible. And decision tree, SVM and random forest are chosen classifiers.

1.5 Outline

The outline of the rest of the report is as follows: In Chapter 2, we talk about previous studies and basic theories. Then we cover a detailed description of data pre-processing in Chapter 3, which includes extracting spikes and backgrounds, building training set, test set andclassifiers. In Chapter 4, we detail the whole method

of the cascade of classifiers. And the results are displayed in the Chapter 5 with three classification models repectively..In Chapter 6, we conclude the report with a review of the entire project, followed by future commendations.

Chapter 2

Literature Review

2.1 Detection Methods

This section covers a review of the various methods explored in the previous studies, including single methods and combined methods.

2.1.1 Support Vector Machine (SVM)

Support vector machine (SVM) is one of most useful supervised learning models, and it is a linear machine used to do classification and regression analysis. An SVM model divides the classes by a gap or a plane that is as wide and clear as possible. Test data are mapped into the same space and classified to a class based on which area of the gap they fall or the plane they locate [24].

Actually, the whole algorithms of SVM are quite complicated, considering that we only classify the spikes and backgrounds, so using SVM to solve the binary classification problem is introduced as follows.

Firstly, we consider N training samples:$\{\{x(1), d(1)\}, \{x(2), d(2)\}, ...\{x(n\}, d(n)\}\}$ where:

$$x(i) = [x_1(i), x_2(i), ...x_n(i)]^T$$

$d(i)$ are class labels that samples x(i) belongs to. It is assumed that the value of +1 or -1 is assigned to $d(i)$.

The purpose of classification is to acquire a hyperplane as decision surface to separate the data of the two classes, as Figure 2.1 shows:

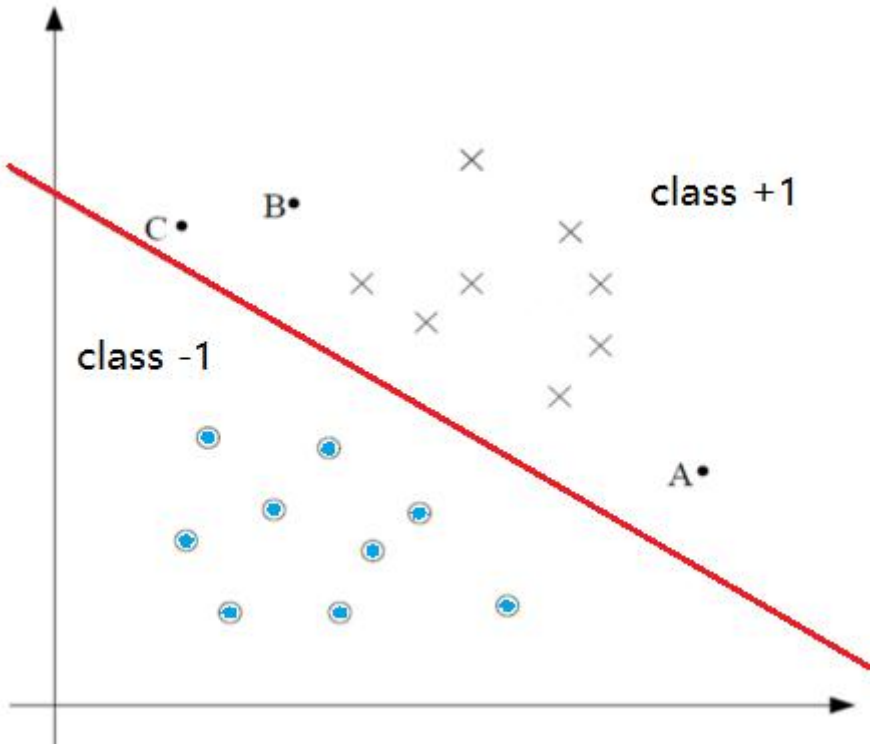

Figure 2.1: How a hyperplane separates two classes.

The equation of the hyperplane in the Figure 2.3 is given as (2.1) :

$$w^T x + b = 0 \tag{2.1}$$

Where x is the input data, w is the weight vector which can be changed, and b is a bias.

We can write :

$$w^T x + b \geq 0 \qquad \text{for } d = +1 \tag{2.2}$$

$$w^T x + b < 0 \qquad \text{for } d = -1 \tag{2.3}$$

For w , which is a weight vector decided by the designer ,and we use the term" margin of separation" or the bias "b" to implicate the separation between the nearest data point and hyperplane.

Finding the particular hyperplane which separates the most of two classes data is the final objective. We would like to find the optimal decision surface under above (2.2)

and (2.3) conditions.

The following diagram illustrates what an optimal hyperplane for a two-dimensional input space:

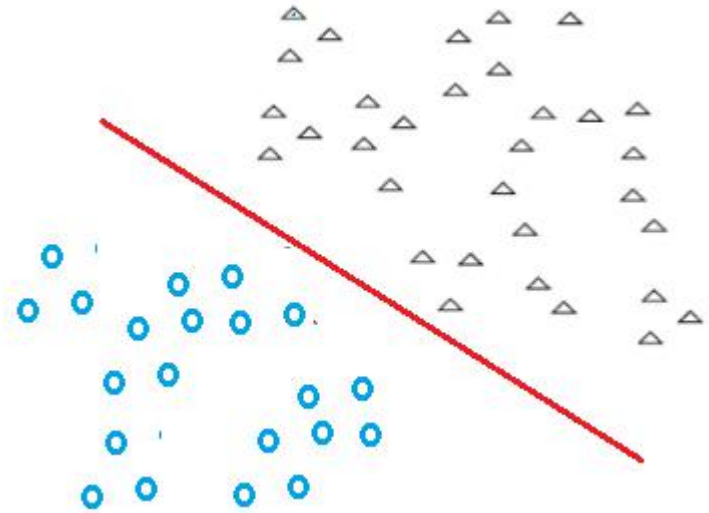

Figure 2.2: A normal hyperplane which separates the two classes.

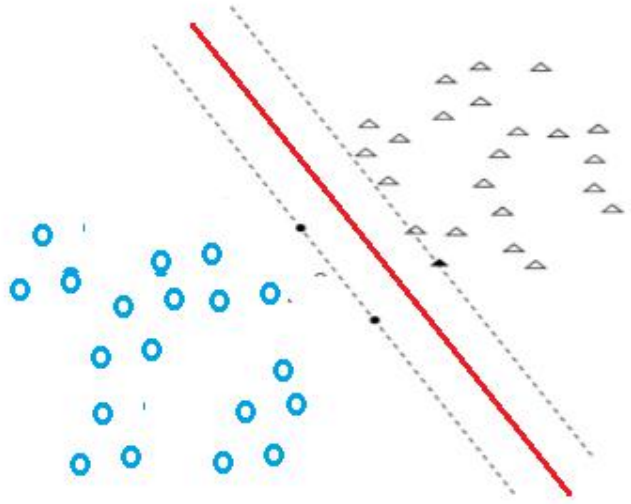

Figure 2.3: A optimal hyperplane which separates the two classes.

The discriminant function :

$$g(x) = w^T x + b \tag{2.4}$$

Equation (2.5) shows an algebraic calculation of the distance from x to the hyperplane. And we can express like this:

$$x = x_p + r\frac{w}{\| w \|} \tag{2.5}$$

Where x_p is the projection of x onto the hyperplane, and r is the distance.

r is positive if x is on the positive side of the optimal hyperplane, and negative if x is on the negative side. By definition, $g(x) = 0$. Then we have:

$$g(x) = r \| w \| \tag{2.6}$$

If we find the data linearly separable, w and b can be scaled like that：

$$w^T x + b \geq 1 \qquad \text{for } d = +1 \tag{2.7}$$

$$w^T x + b \leq -1 \qquad \text{for } d = -1 \tag{2.8}$$

The particular data points for which the above is satisfied with equality sign are called support vectors. In theory, the support vectors are those data points that locate closest to the decision surface.

Thus, the algebraic distance from support vector $x^{(s)}$ to the optimal hyperplane is:

$$r = \frac{g(x^{(s)})}{\| w \|} = \begin{cases} \dfrac{1}{\| w \|} & \text{for } d^{(s)} = +1 \quad (2.9) \\ \dfrac{-1}{\| w \|} & \text{for } d^{(s)} = -1 \quad (2.10) \end{cases}$$

In order to develop a computationally efficient procedure to find the optimal

hyperplane, a constraint is applied as (2.11):

$$d \times (w^T x + b) \geq 1 \tag{2.11}$$

Given the training samples $\{x(i), d(i)\}$, i = 1,2,3..,,n the optimal values of the weight vector w and bias b satisfy the constraints：

$$d(i) \times (w^T x(i) + b) \geq 1 \quad \text{for } i = 1,2,3.....n \tag{2.12}$$

and the weight vector minimizes the following cost function：

$$J(w) = \frac{1}{2} w^T w \tag{2.13}$$

We may solve the constrained optimization problem using the method of Lagrange multipliers. Firstly, we construct the Lagrange function:

$$J(w,b,\alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{N} \alpha(i)[d(i)(w^T x(i) + b) - 1] \tag{2.14}$$

We call $\alpha(i)$ Lagrange multipliers. The solution to the constrained optimization problem is determined by the saddle points of the Lagrange function $J(w,b,\alpha)$, which has to be minimized with respect to $\alpha(i)$ and b. but maximized with respect to w.

Differentiating $J(w,b,\alpha)$ with respect to $w$and $b$and setting the results to zero, we get the following two conditions of optimality：

Condition 1: $$\frac{\partial J（w,b,\alpha)}{\partial w} = 0 \tag{2.15}$$

Condition 2: $$\frac{\partial J（w,b,\alpha)}{\partial b} = 0 \tag{2.16}$$

Following Karush-Kuhn-Tucker (KKT) optimization theory, by the optimality condition 1, we have:

$$w^T w = \sum_{i=1}^{N} \alpha(i) d(i) w^T x(i) = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha(i)\alpha(i) d(i) x^T(i) x(j) \tag{2.17}$$

$$\alpha(i)[d(i)(w^T x(i) + b) - 1] = 0 \tag{2.18}$$

$$\alpha(i) \geq 0 \qquad \text{for } i = 1, 2, 3 \ldots n \tag{2.19}$$

Substituting the above into the expansion of $J(w, b, \alpha)$ yields

$$J(w,b,\alpha) = \sum_{i=1}^{N} \alpha(i) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha(i)\alpha(j) d(i) d(j) x^T(i) x(j) \tag{2.20}$$

Subject to the conditions:

$$(1) \sum_{i=1}^{N} \alpha(i) d(i) = 0 \tag{2.21}$$

$$(2) \alpha(i) \geq 0 \tag{2.22}$$

After acquiring the optimum Lagrange multipliers, we may compute the optimal weight vector w and b :

$$w^* = \sum_{i=1}^{N} \alpha(i) d(i)\ x(i) \tag{2.23}$$

$$b^* = 1 - w^* x^{(s)} \qquad \text{for } d^{(S)} = 1 \tag{2.24}$$

In the previous studies based on SVM, Yaozhang Pan [25] has researched the property of SVM and applied it to classify spikes of EEGs. He and his team extracted mean value, accumulated energy, six power, average nonlinear energy, curve length from the five groups of EEG dataset. Then they tried three kinds detection methods, the first method was to use SVM to do Novelty Detection, the second was ANN classification, the last one was Benchmark Detection. Comparing their results, using SVM resulted in an incredible 100% accuracy, which was much better than the other methods (accuracy rates ranged from 86.8%-97.4%). Nicoletta and his team [26] used 5 dataset [27].The accuracy of other previous studies used this dataset with

single-trial classification ranged from 92.2% to 100% [28-33]. Author introduced a Permutation Entropy as a feature for automated seizure detection with a Support Vector Machine (SVM). They used the PE values as an indicator to detect the seizures.And the maximum overall accuracy of all dataset was 86.10%.

2.1.2 Decision Tree

Decision Tree is based on the probability of occurrence of various circumstances, by making a decision tree to obtain the net present value of the expected value greater than or equal to the probability of zero. This method is a graphic method for probabilistic analysis of intuitive application. Figure 2.4 shows how a decision tree works.

In machine learning, the objective of decision tree is to map a relation between object attributes and object values. Entropy equals system messy degree, this measure is based on the concept of entropy in informatics theory [34] [35].

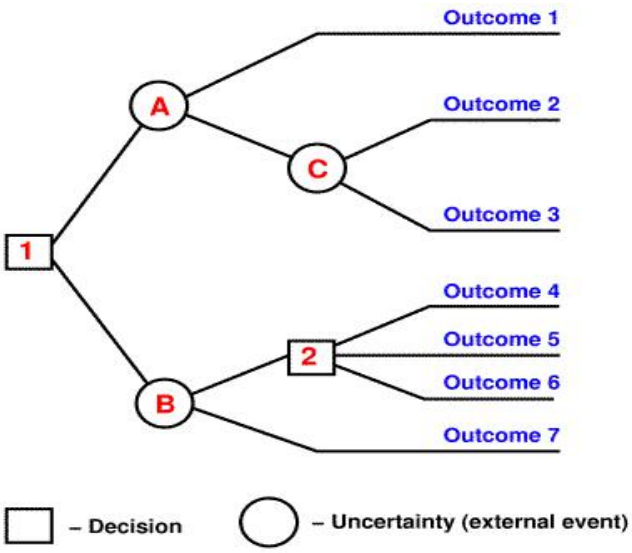

Figure 2.4: An example of a decision tree. [34]

In previous research based on Decision Tree, Kemal Polat and his team's research [29] proved that the decision tree can be a very valuable diagnostic decision support tool to detect seizures. They combined the hybrid medical decision-making system

with the FFT-based Welch method and decision tree classifier, which obtained the accuracy of 98.68% from 5-fold cross validation and 98.72 from 10-fold cross validation.

2.1.3 Random Forest

In fact, If we use decision tree to test a large dataset with many features, too many nodes in the tree will be generated, which lead to the problem of overfitting. What's more, the noise of the data would highly influence the results of training and testing, making training accuracy high but testing accuracy low. And random forest may solve the above problem.

Random forest classification model is formed by a number of decision trees and classification model-- $\{h(x,\theta_k\}, k = 1,..,..\}$ and parameters set { $\exists_k$ } are independent and identically distributed random vector. Under the given independent variable x, each decision tree classification model selects the optimal classification result by voting [36].

Constructing a Random Forest model is a very complicated process, I will introduce the whole process briefly.

Firstly, k samples are extracted from the training samples using bootstrap sampling, and the sample size of each sample is the same as the original training set.

Secondly, the k decision tree models are established respectively according to k samples, and we will obtain k kinds of classification results.

At last, the final classification was decided by voting according to the all classification results [37]. The whole procedure are shown in the Figure 2.5.

There are many advantages as follows if we apply random forest:

(1) When there is an imbalance in the situation, the random forest can provide an

effective way to balance the dataset error, it is suitable for multi-classification problems.

(2) It is fast in training and testing.

(3) The fault-tolerant ability of training data is an effective method to estimate the missing values. When the dataset is large, the accuracy can still be kept constant.

(4) Random forest can effectively handle large dataset.

(5) It can handle high dimensions (many features) of the data and do not have to do feature selection.

(6) It can detect the interaction between the features.

(7) There is much lower chance to cause the problem of overfitting.

Considering these advantages, we find the random forest is a suitably supervised learning model to do the EEG classification.

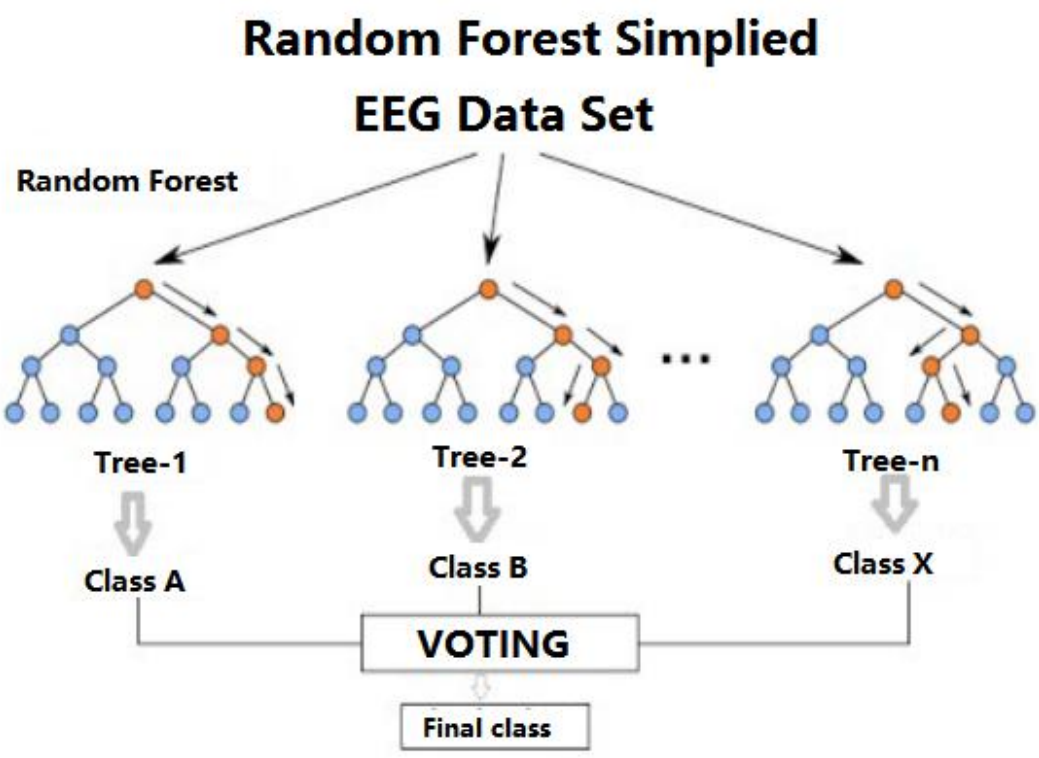

Figure 2.5: The Structure of a random forest.

In the previous study with random forest, Cristian Donos [38] and his team adapt random forest to test the 10 patients with different methods. They mainly extracted autocorrelation features and power-ratio features. The sensitivity of their study ranged from 20% to 100%. Although the interval was very large, almost half of the sensitivities were larger than 95%.

2.1.4 AdaBoost

AdaBoost is a machine learning method presented by Yoav Freund and Robert Schapire [41]. The AdaBoost method is an iterative algorithm in which a new weak classifier is added in each iteration until reaching a predetermined error rate. Each training sample is given a weight that indicates the probability that it is selected by next new classifier into the training set [42]. If a sample has been accurately classified, then the probability of being selected is reduced in constructing the next training set. Conversely, if a sample is not accurately classified, its weight will

be increased.

In fact, many algorithms with good performance have been put into effect based on the improvement of finding simple classifiers and boosting them. These include the enhancement of AdaBoost learning, taking advantage of new features, selecting optimal feature, etc. FloatBoost, a variant of AdaBoost, is proposed by Li and Zhang (2004) [43]. It achieves in finding a stronger classifier, which contains fewer weak classifiers and can acquire a higher accuracy. Ander (2004) [45] combines the AdaBoost framework with an evolutionary algorithm and found that a better classifier with new features was produced .

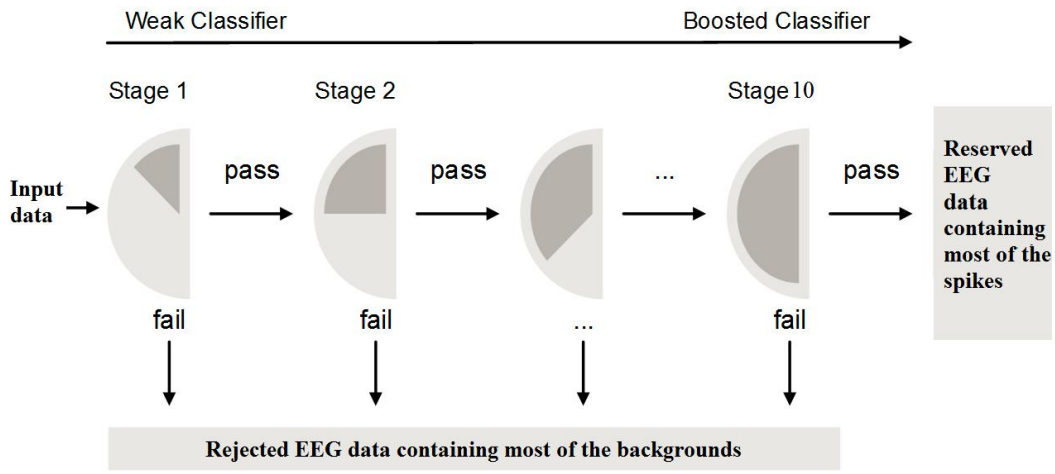

Figure 2.6: An AdaBoost classifier model.

2.1.5 Template Matching (TM)

Template Matching is one of the most common methods applied for detecting spikes. The objective of the method is to identify the waveforms that are similar to the waveforms in the template library. The template library plays the most important role in a template matching system. The reason is that the effectiveness of the system primarily relies on the quality of the templates in the library. The given waveforms and the templates are matched based on a prescribed distance measure. The matching between the given waveforms and the templates is based on prescribed distance measure. Correlation coefficient, Euclidean distance and Sum of Squared Error (SSE) are the common distance measures adopted for TM. The major problem associated with TM is the quality of the templates. The choice of the templates primarily determines the efficiency of the TM system.

Lodder et al. devised a self-adapting TM system based on a set of smart templates in [48]. The study was conducted on a set of 23 patients with an IED count of 723. The sensitivity was reported as 95% and Fp/minute as 6.6, after 15 iterations. Nonclercq et al. applied temporal clustering to a set of IEDs from a set of 3 patients (2500 spikes) and a template library was created. Subsequently, a system which adapts to

interpatient and intrapatient variation in spike morphology was devised based on the template library. The performance in terms of sensitivity and selectivity was reported as 93.7% and 93.7% respectively [49].

2.1.6 Artificial Neural Networks (ANN)

Artificial neural networks (ANN), or simply, neural networks are typical connectionist models. And it has following properties:

(1) It has certain characteristics in common with biological neuron networks.

(2) Information processing occurs at many same simple elements called neurons.

(3) Information are processed in parallel.

(4) It can do nonlinear information processing.

(5) It has learning capability .

In the Figure 2.1, u_j, the computing units, refer to be the neurons. The neuron receives one or more inputs (resembling dendrites), adds up the inputs according to the connection weights, and produces an output (resembling a neuron's axon). Generally, we will apply activation function for the output. The common activation functions are sigmoid function, piecewise linear functions, and step functions. By calculating the error or loss function between the outputs O_k and the targets Z_k , we can update the weight W_{ij} to adjust the output. The backpropagation algorithm is the most commonly applied learning rule for updating the weights. The weights are firstly initialized randomly. Next the weights are updated by minimizing the derivative of loss function with respective to the weight. Quite a lot of NN-based approaches have been explored for spike detection [50-54].

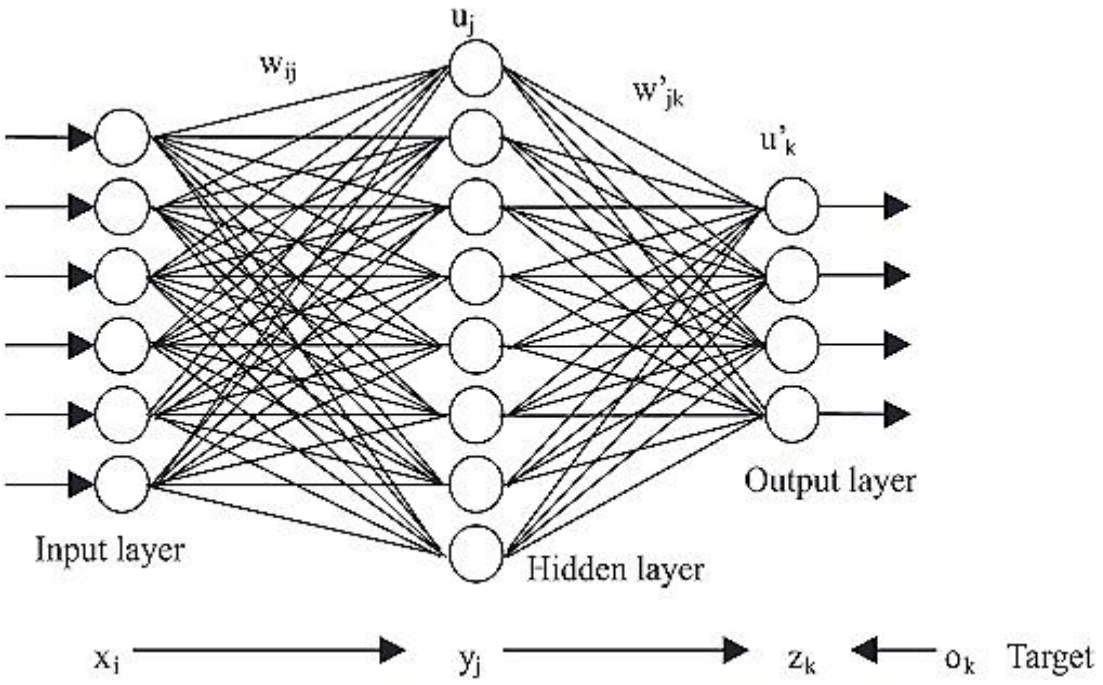

Figure 2.7 A basic model of a NN with single hidden layer. [52]

Wilson et al. developed a multiple monotonic NN-based spike detection method in [67]. The study was implemented on 50 epileptic patients (2400 spikes) and 10 control subjects. The comparison between the performance of the proposed NN-based method with traditional visual inspection and rule-based method is evaluated. The research achieved a sensitivity of 89.9% and specificity of 99.63%. Carey et al. employed a standard NN with back-propagation algorithm for spike detection. The NN was constructed with a 10-neuron single hidden layer. The input layer contained 120 neurons. The network was trained with features extracted from the spike and non-spike waveforms. The sensitivity and selectivity were reported as 82.68% and 72.67% (specificity is not reported) [52].

2.1.7 Mimetic Approach

Mimetic approach is a common analysis approach for spike detection [56-67]. This approach tries to analyze the spike waveform in terms of morphological characteristics. The spike waveform is decomposed into different transients and the features, namely, amplitude, duration, and slopes, are extracted. In James et al. [56] research, as shown in the Figure 2.2, the mimetic parameters extracted are amplitude ($\hat{A}$), the durations of half-waves (D1, D2, D3), rising and falling slopes (S1, S2) and sharpness (S_P). Three contextual parameters: average amplitude ($\hat{A}$), average duration ($\hat{D}$), and average slope ($\hat{S}$) are also extracted. Mimetic method has been applied independently [68-72] or in a combination with other methods [73-76].

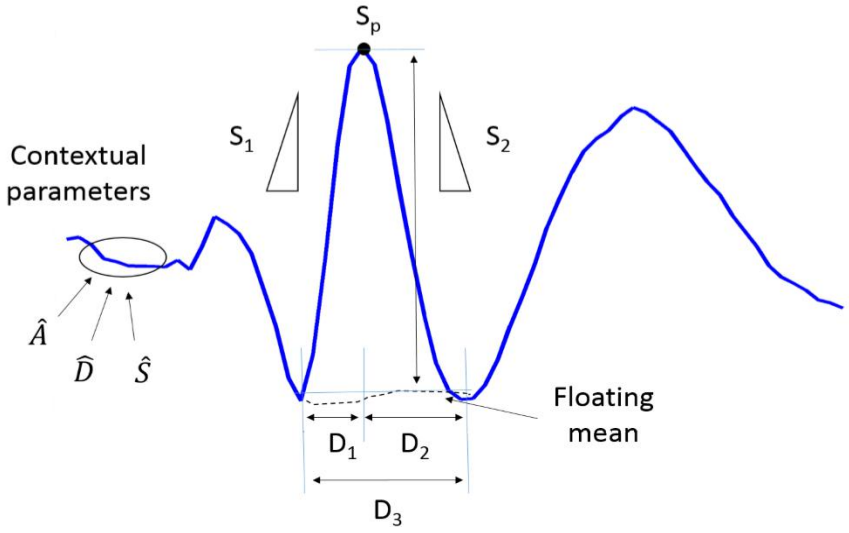

Figure 2.2: The mimetic parameters described by James et al. [56]

2.1.8 Results for Combined Methods

Some times, only one classifier can not acquire a good performance for detecting spikes in EEG. Table 1.1 shows some results with combined methods, we can find the results are acceptable. Consequently, it is difficult to compare the different methods in the table, due to the highly diverse dataset , methods and performance

evaluation characteristics.

Author	Method	Sensitivity	Specificity
Antoine Nonclercq (2012) [9]	TM+clustering	90	88.2
Shaun S. Lodder (2013)[10]	TM+SVM	90	-
Yung-Chun Liu (2013) [11]	NLEO+Mimetic+ AdaBoost	92.6	90.46
Jian Zhang (2013) [12]	Sequence merging +SVM	96.1	-
Bagheri,Elham (2016) [13]	MF + MT+ Cascade	90.6	98.65
James et al (1999) [55]	Mimetic+ ANN+Fuzzy logic	91.1	-

Table 2.1: Results of combined methods.

TM: Template matching. SVM: Support Vector Machine.

NLEO: Nonlinear Energy Operator.

MF: Morphological features

MT: Wavelet transform

2.2 Overfitting

Overfitting is a very normal problem during the classification procedure. It implicates that the performance is very good in training but very bad during testing. There are three main reasons leading to overfitting [47]:

(1) Training data may contains noise, the machine considers parts of the noise as a feature, then the classification rules are influenced.

(2) There are too many parameters, the model complexity is too high.

(3) The method of sample selection is incorrect, such as many errors in sample labels, which results in the selected sample data not good enough to represent the intended classification rules.

2.3 Cross Validation

In this study, we seek for a method for the combination of multiple classifiers (Wolpert 1992) [39] from the given dataset. For estimating the accuracy of a cascade of classifiers, the objective is that the accuracy will not be influenced by the noise or some data which are worthless, so we want the variance and bias in the prediction model are low. To select a classifier, the accuracy is less important and we may sacrifice bias for a low variance, assuming the bias affects all classifiers similarly (e.g., estimates are 5% pessimistic) [40]. Considering what I mentioned above and the problem of overfitting, Cross validation is the method we choose.

In the study of machine learning, the entire dataset is often divided into two parts, namely training set, and test set. Assuming that X is a set of all data, A is the nonempty true subset of the complete set X, then the nonempty set (X \ A) means removing set A from set X. So doing training and analysis on top of A should be put in the first place, and set (X \ A) is used for testing and verification. The set A is called the training set, and set (X \ A) is called the validation set or the test set [40]. There is an important point that only training sets can be used in the training of the model, and the test set must be used to assess the model's error after the model training is completed.

In this study, we use K-folds to do cross validation, we set k to be 5, means 5 groups of the different dataset. 93 patients' EEG data are separated randomly into 5 subsets.

A separate subset is retained as the data for the validation model, and the other 4 subsets are used for training, each group of the dataset is validated once, then we average 5 groups' results to a final accuracy. One thing should be mentioned is that the actual training data and test data are spikes and backgrounds extracted from the 5 subsets, and extraction ways are much different between the training data and testing data, I will talk it in the next chapter.

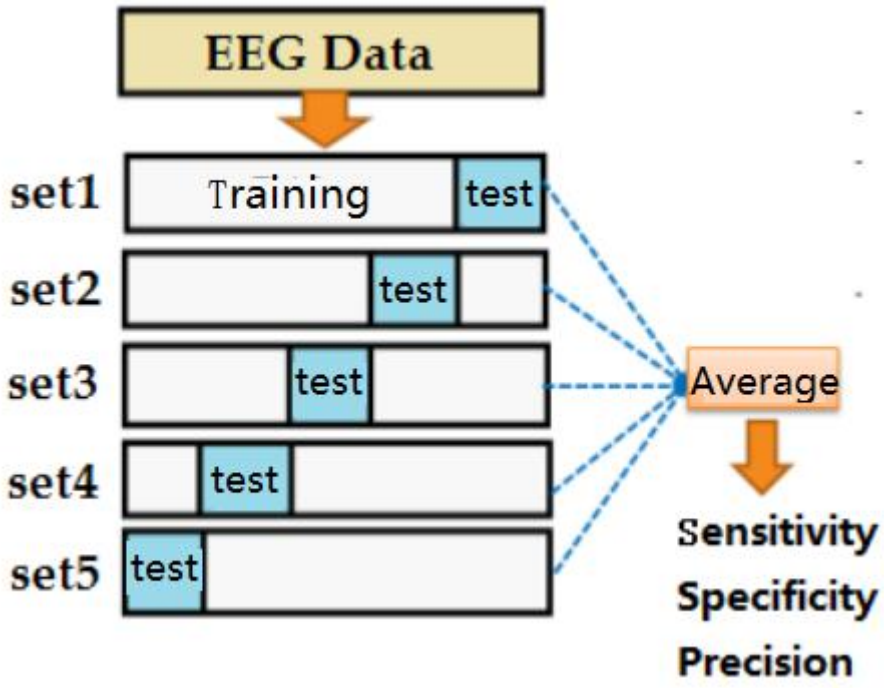

Figure 2.9: The cross validation model.

2.4 Performance Metrics

In this thesis, sensitivity, specificity, and precision are three terms we used to value the performance.

Following terminology and derivations are cited from [46].

True positive (TP): The quantity that test data are spikes and the corresponding prediction outcome are spikes.

False positive (FP): The quantity that test data are spikes and the corresponding prediction outcome are backgrounds.

True negative (TN): The quantity that test data are backgrounds and the corresponding prediction outcome are backgrounds.

False negative (FN): The quantity that test data are backgrounds and the corresponding prediction outcome are spikes.

Sensitivity	TPR = TP/(TP + FN)
Specificity	SPC = TN/(FP + TN)
Precision or Selective	PPV = TP/(TP + FP)

Table 2.2: Algorithms of sensitivity, specificity and precision.

Chapter 3

Data Processing

3.1 Introduction

This chapter mainly introduces how to acquire training data, test data and classifiers from the raw EEG data.

3.2 Extracting Spikes

We collect 93 epileptic patients' EEG dataset from Massachusetts General Hospital. The sampling frequency is 128Hz and a notch filter at 60HZ is applied to remove the power line interference. If one piece of brain electrical signals in any channel is detected as a spike by more than one expert, these experts will make an annotation for the waveform as a spike. we extracted 64-points from the onset of the annotation. Figure 3.1 shows a sample electrical signals from an epilepsy patient, the x-axis implicates the sampling time, the y-axis implicates the voltage value of sample points.

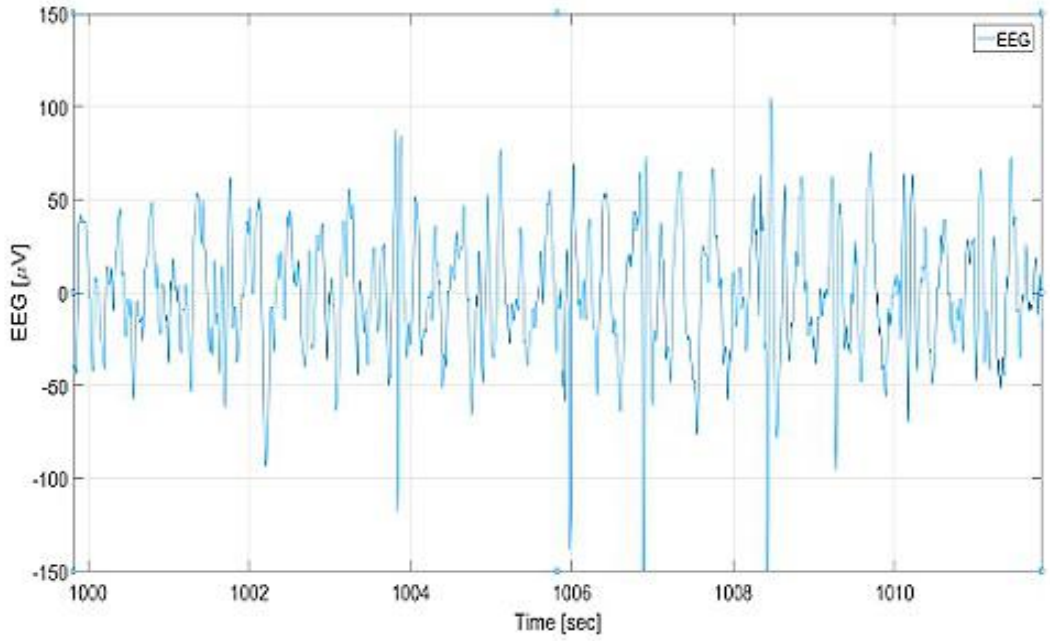

Figure 3.1: A clip of the one channel's raw EEG data

Figure 3.1 clearly shows the complication and disorder of the EEG data, so that normal people can not recognize what are spikes and what are backgrounds, and the task of annotating is accomplished only by the professional experts. In Figure 3.2, the red curve is a spike annotated by the experts.

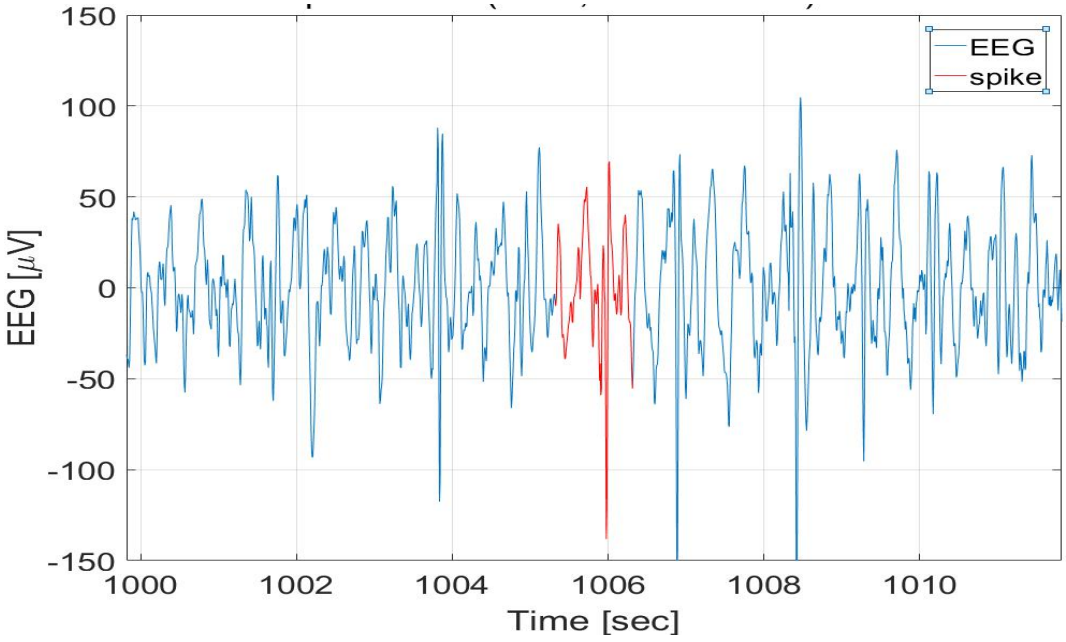

Figure 3.2: An example of an annotation for a spike（the red curve).

Then , we extract the 64-points spike and save the waveform as a statistical data.

The whole spike-extraction procedures are the above steps I mentioned, and we extract 18164 spikes in total from 93 patients.

3.3 Extracting Backgrounds

In the study, we want to extract backgrounds which do not overlap with spikes. Actually, the moment when a seizure happens, the all channels of EEG data may be highly influenced, that means if a waveform is confirmed as a spike in one channel, at the same time, the corresponding waveform recorded in all other channels can also be regarded as spikes. Thus, we remove the possible duplicates of the spike that appear in the neighboring channels. The duplicates are scaled shifted versions of the original spikes.

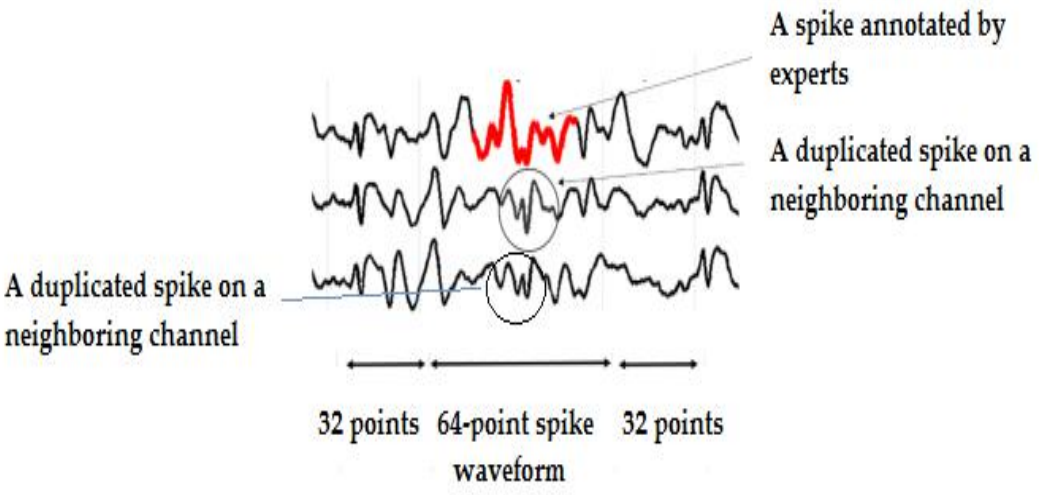

Figure 3.3: Removing possible duplicates of spikes on the neighboring channels.

In fact, a spike randomly happens and we can not predict how long it lasts and its shape, thus there may be some spikes with low peaks or curves close to the annotated spike, that means waveform of the spike may be more than 64 points. In order to

accommodate the shifted duplicates, we also remove 32 samples points before and after the 64 sample spike portions from all the 19 channels. It will do much help in the training stage where we hope to find the best threshold for discriminating spikes and backgrounds.

Then, after removing what we do not need in raw EEG data, we segmented the left EEG data at the rate of 64 points consequently. If some segments do not contain 64 points, for example, they may be overlapped with the removed spikes, then these segments will be discarded too.

3.4 Constructing the Data

Because we have 30 min EEG recordings of 93 patients with epilepsy and 18164 spikes, we separate patients with k-fold (k =5) method, then we have 5 folds of patients for training and testing. For example, now we have 93 patients EEG data, by k-fold (k = 5), we will randomly choose 74 patients' EEG data as training-patient set, the other 19 patients' EEG data will be the test-patient set.

After we divide the patients into the training parts and the testing parts, we will build training data, test data and classifiers as following steps.

(1) Building the training data: Training data stands in the main position of the classification, the steps of building training data are summarized as following:

- Extracting spikes: We extract all the spikes from training-patient set.
- Extracting backgrounds: If we have N spikes in total which are extracted from M training-patients' EEG data, then we will extract 3N backgrounds from each training-patient randomly. At last, we will have 3M groups of backgrounds ,each group contains N backgrounds, means 3*M*N backgrounds in total .
- The overall training data will be generated by combining all the N spikes and 3M groups of backgrounds.

(2) Building classifiers: Combining all the N spikes with N backgrounds (These N backgrounds are all from the same patient's EEG data) will be regarded as a classifier, then we will have 3M classifiers in total.

(3) Building the overall test data: In order to test the final overall performance, we extract all backgrounds from test-patient set with 50% overlapping, then combine these backgrounds with the spikes extracted from test-patient set to make the final overall test data.

(4) Labeling the data: For all classifiers, overall training data, and overall test data, we = use ''1'' to label the spikes and ''-1'' to label the backgrounds.

Chapter 4

Cascade of Classifiers

4.1 Introduction

This chapter describes the whole detailed procedures of how we use the method -- a cascade of classifiers, and how we use these classifiers to do the backgrounds rejection and spikes reservation. The basic idea of this method is showing in the Fig. 4.1. For the classifier, there are many methods, we choose decision tree, SVM, and random forest and compared the performance.

We develop a way to let all the classifiers compete with each other for 10 stages, through this way, 10 winning classifiers are found.

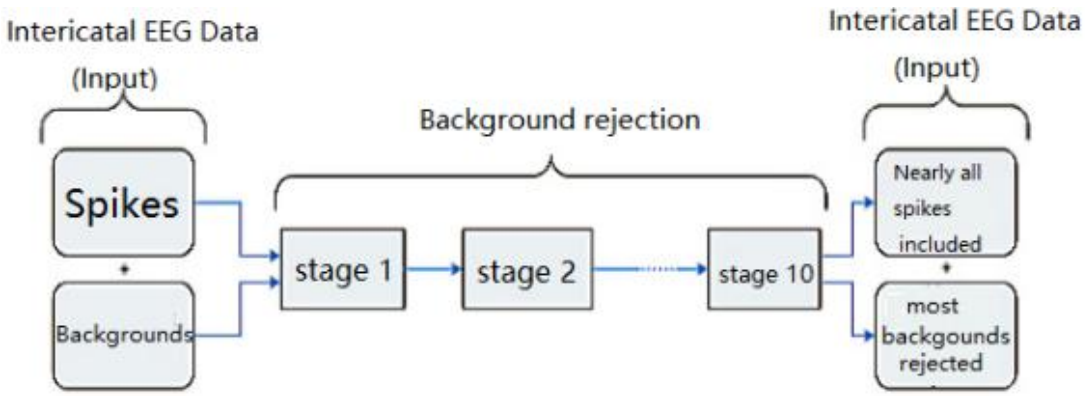

Figure 4.1: Schematic of background rejection method. [13]

4.2 Training

Firstly, we use a classification method to train one of the classifiers.

Secondly, we apply this trained model to predict the overall training data, except the prediction outcomes, all the data's possibility values are generated. We set a threshold to the possibility values. The possibility value implicates every waveform's possibility to be a spike. For example, if one waveform' s feature value is 0.9, then it means that this waveform is more possible to be a spike rather than a background. if a possibility value is larger than the threshold, the corresponding waveform is regarded as a spike, otherwise, a background. For example, if one waveform's score is 0.9, and the threshold is 0.6, then they regard the waveform as a spike.

4.3 Finding the Winning Classifier and Winning Threshold

We hope to find a threshold could classify the spikes and backgrounds for input EEG data as many as possible, but the threshold can not be calculated or designed according to the existing data. Hence, we use all the possibility values as thresholds to test input data, then calculate corresponding sensitivities and specificities. At last, we save specificity into an array whose corresponding sensitivity most proximate to 99.8% .

After testing for all the classifiers, the classifier with maximum specificity in the array is the winning classifier.

Then we use winning classifier to do classification for input data, the threshold for possibility values which make the sensitivity most proximate to 0.996， is the desired winning threshold. Since 0.4% of spikes is lost after each stage, in order not to lose too many spikes, we set the length of the cascade to be ten, that means ten stages of backgrounds rejection are applied, so that almost 4% of the spikes are

rejected at most after 10 stages of a cascade of classifiers. And losing 4% of the total spikes is accepted.

4.4 Rejection and Reservation

We apply the winning classifier to test the overall training data with the winning threshold, the waveform that is predicted as a spike (that means its possibility value is bigger than the winning threshold) will be reserved, and waveform that is predicted as a background will be rejected.

4.5 Iteration

Because the proportion of the backgrounds is much larger than that of the spikes in the overall training data, and the features of some backgrounds might be quite similar to the spikes, thus just only one simple classifier can not classify the spikes and backgrounds clearly, we have to seek for another classifier to classify the reserved data.

Next, we repeat the step 4.2 to step 4.4 nine more times, every iteration, we find another winning classifier in the all classifiers, and the reserved data will be next iteration's input data.

After ten iterations, ten winning classifiers are acquired with their corresponding winning thresholds. Combining them together results a cascade of classifiers.

4.6 Testing

We apply 10 wining classifiers with their corresponding winning thresholds to test overall test data consecutively. After 10 stages of background rejection, we calculate the final sensitivity, specificity and precision to value the performance of the method of the cascade of classifiers.

Chapter 5

Results

5.1 Final Results of the Cascade of Classifiers

In this section, a cascade of classifiers is applies to do training and testing. We choose decision tree, SVM, random forest as classification models. The entire MGH dataset is divided into 5 parts of 20 patient data each. 4 parts are employed for training and one for evaluating the cascade of classifiers. The 5-fold cross-validation results are also performed below to support f the proposed method. For more detail information, please see the appendix.

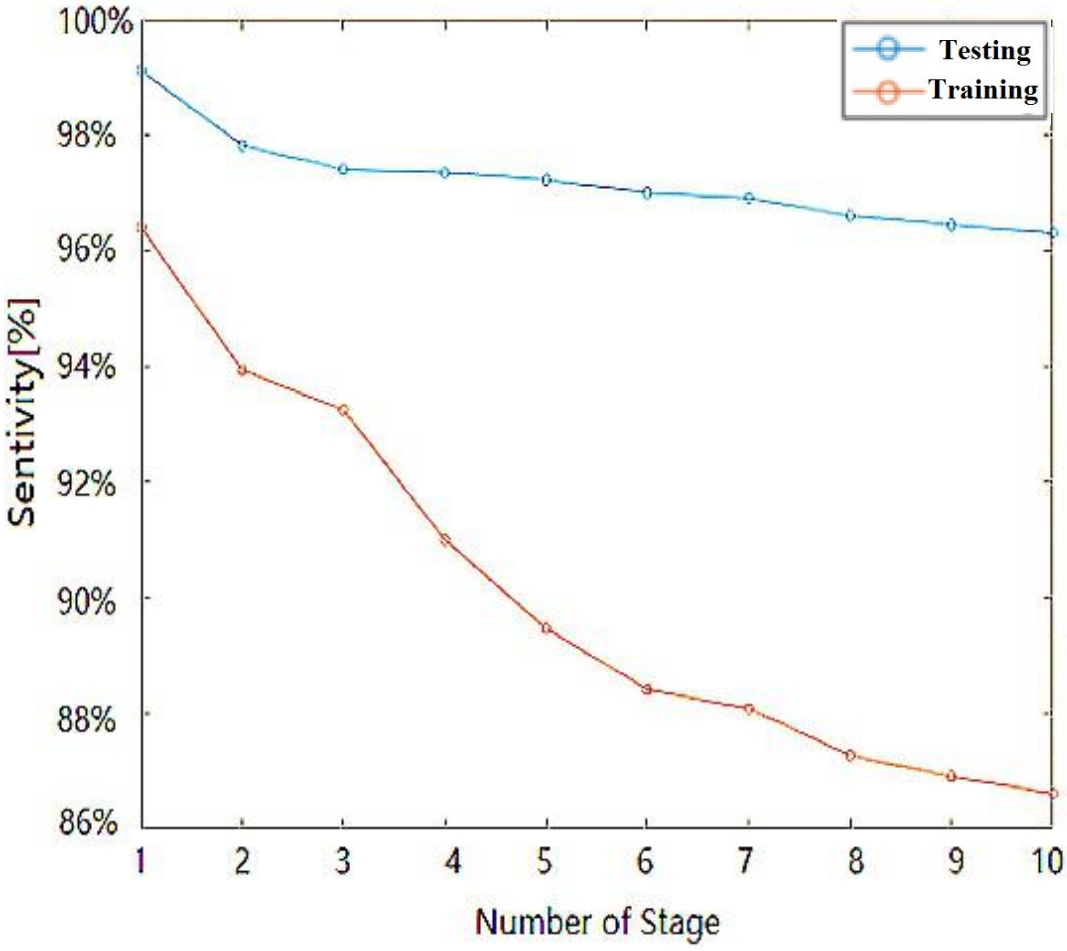

Figure 5.1: Average sensitivity of every stage for training and testing（decision tree）.

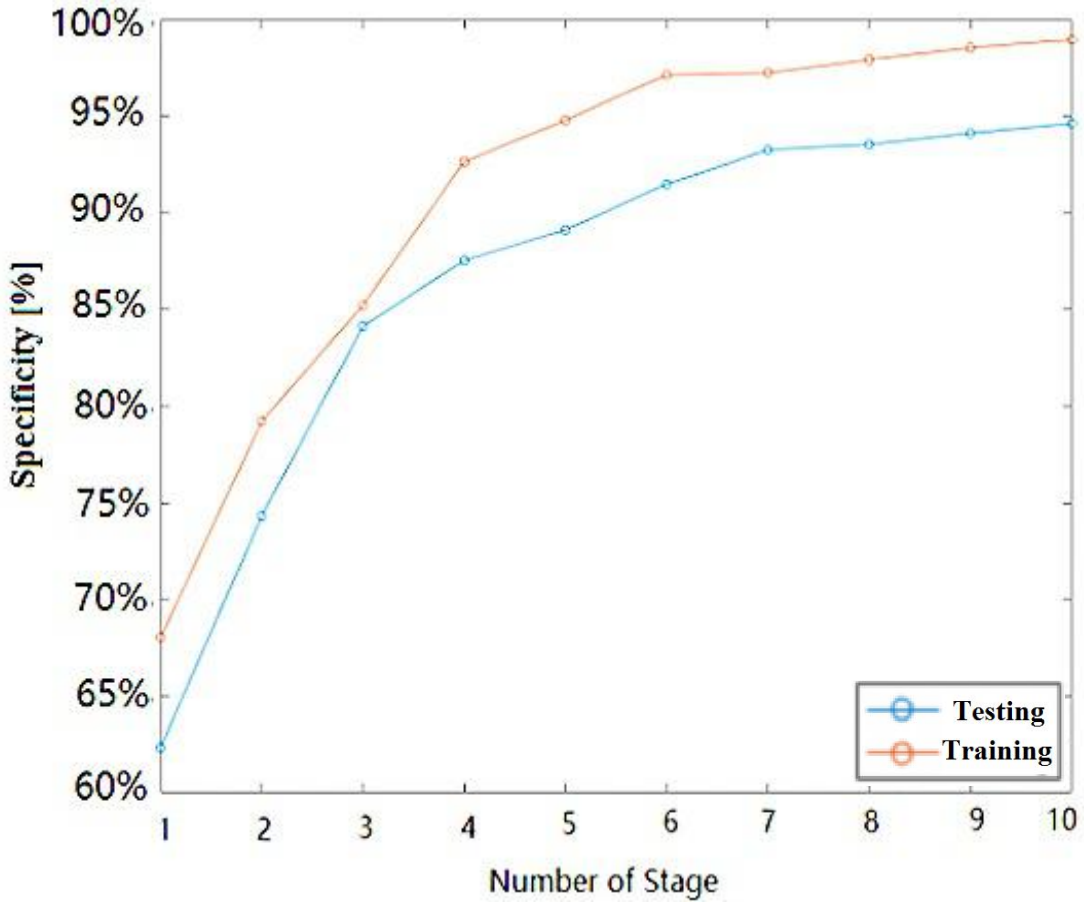

Figure 5.2: Average specificity of every stage for training and testing（decision tree）.

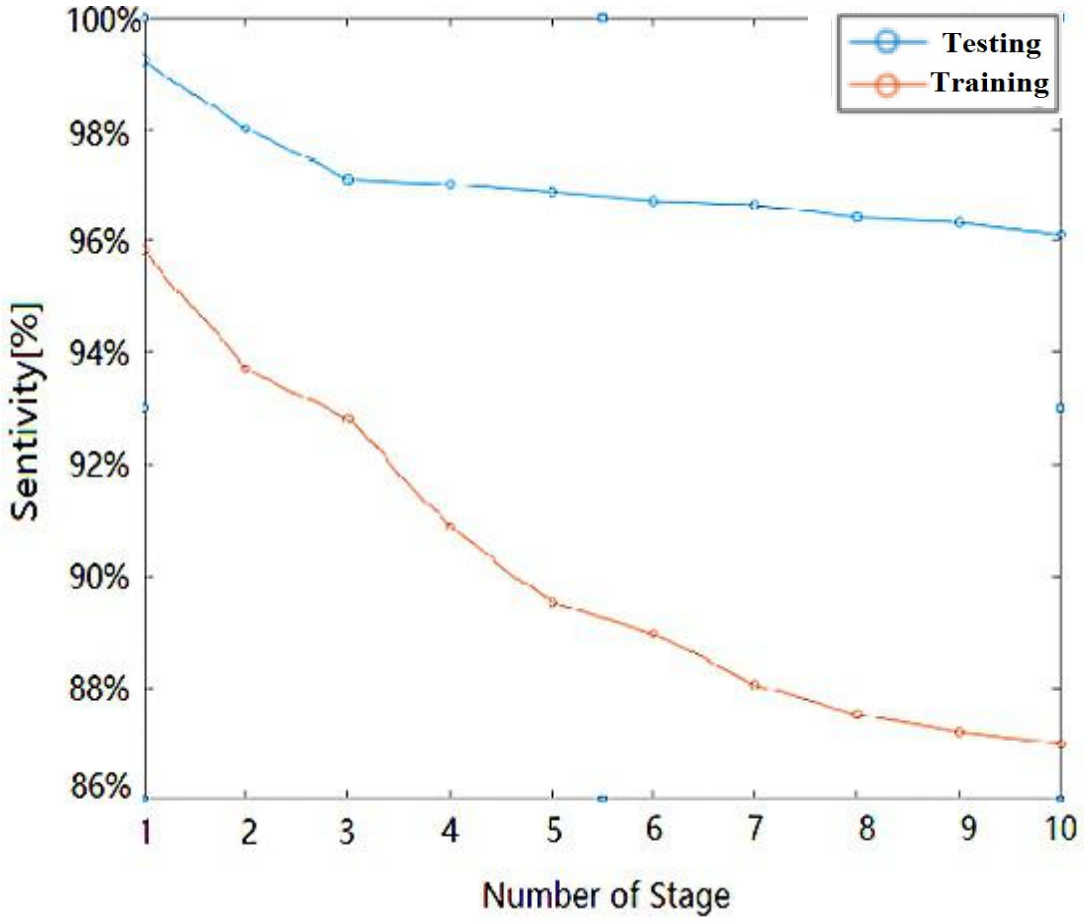

Figure 5.3：Average sensitivity of every stage for training and testing（SVM).

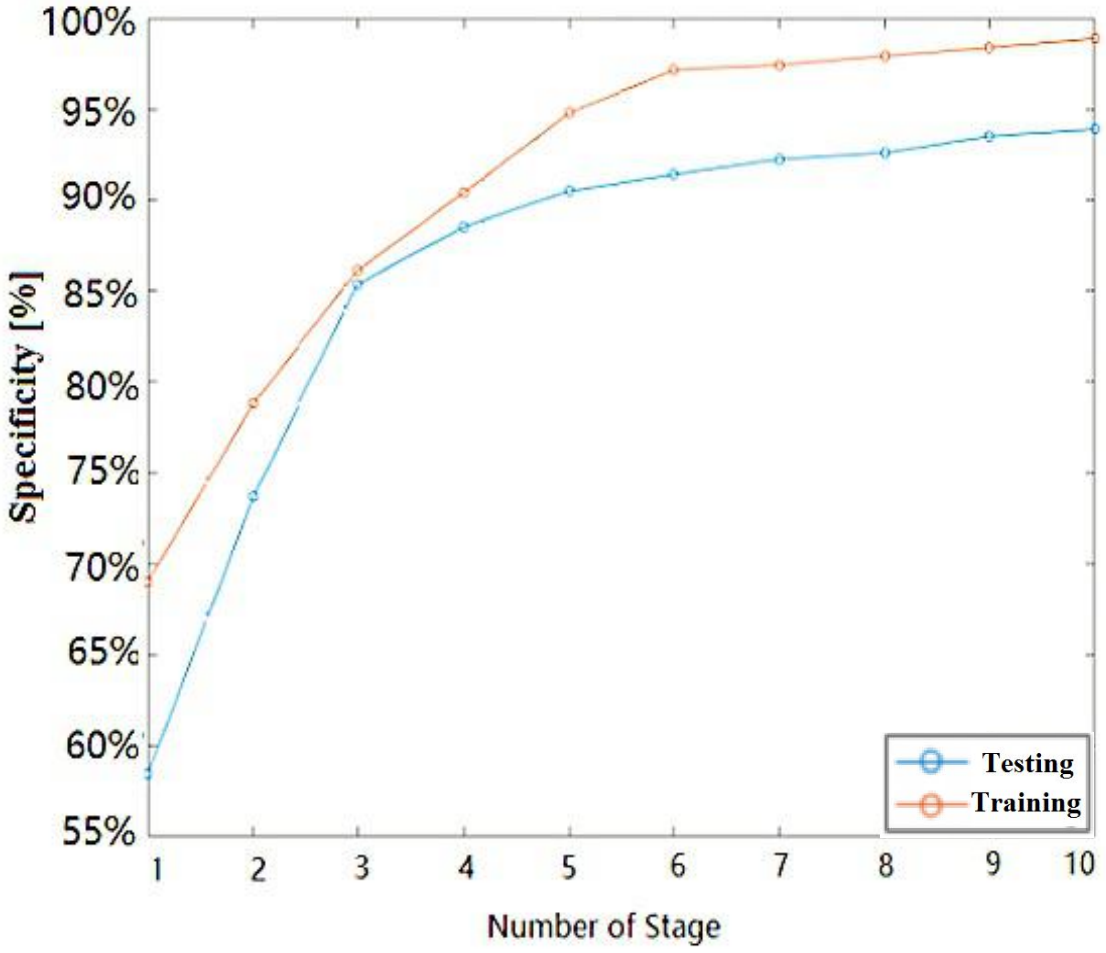

Figure 5.4: Average specificity of every stage for training and testing（SVM).

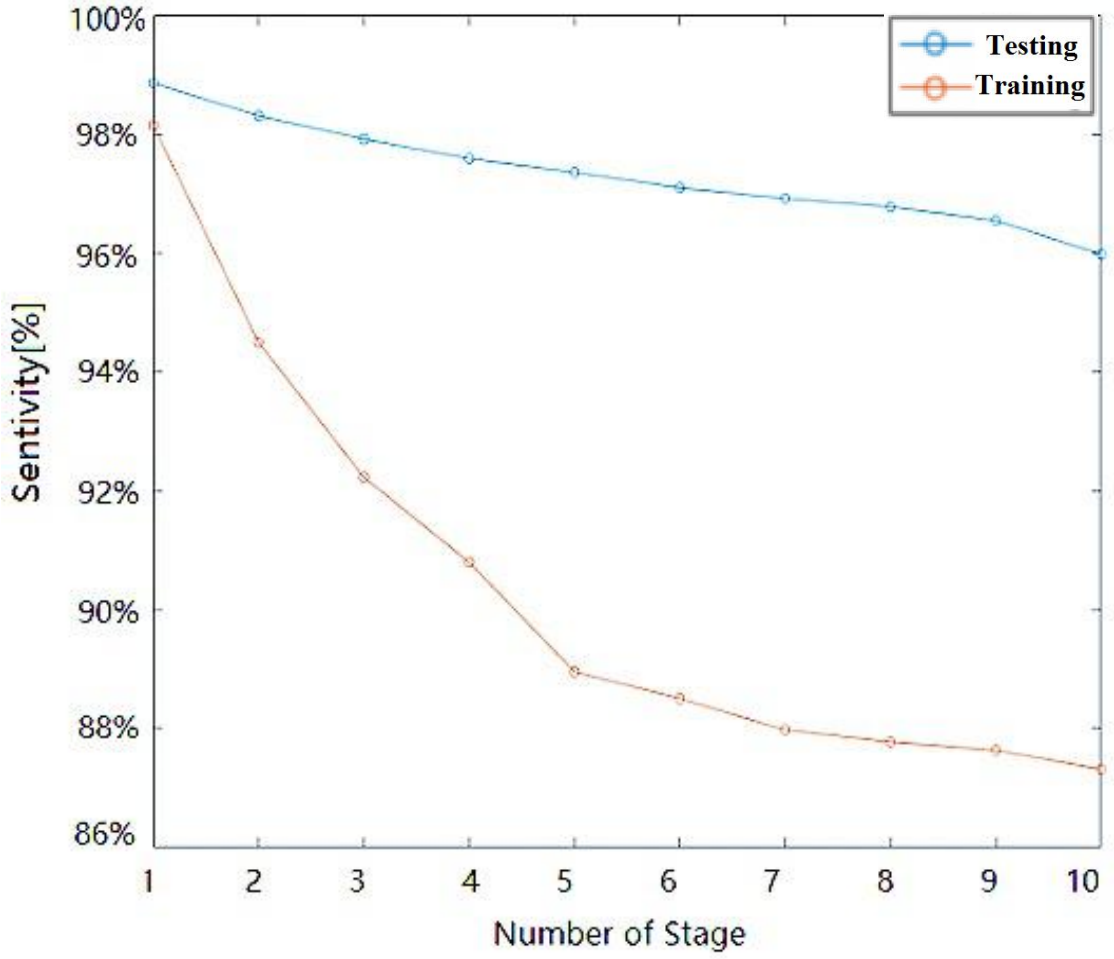

Figure 5.5: Average sensitivity of every stage for training and testing（random forest).

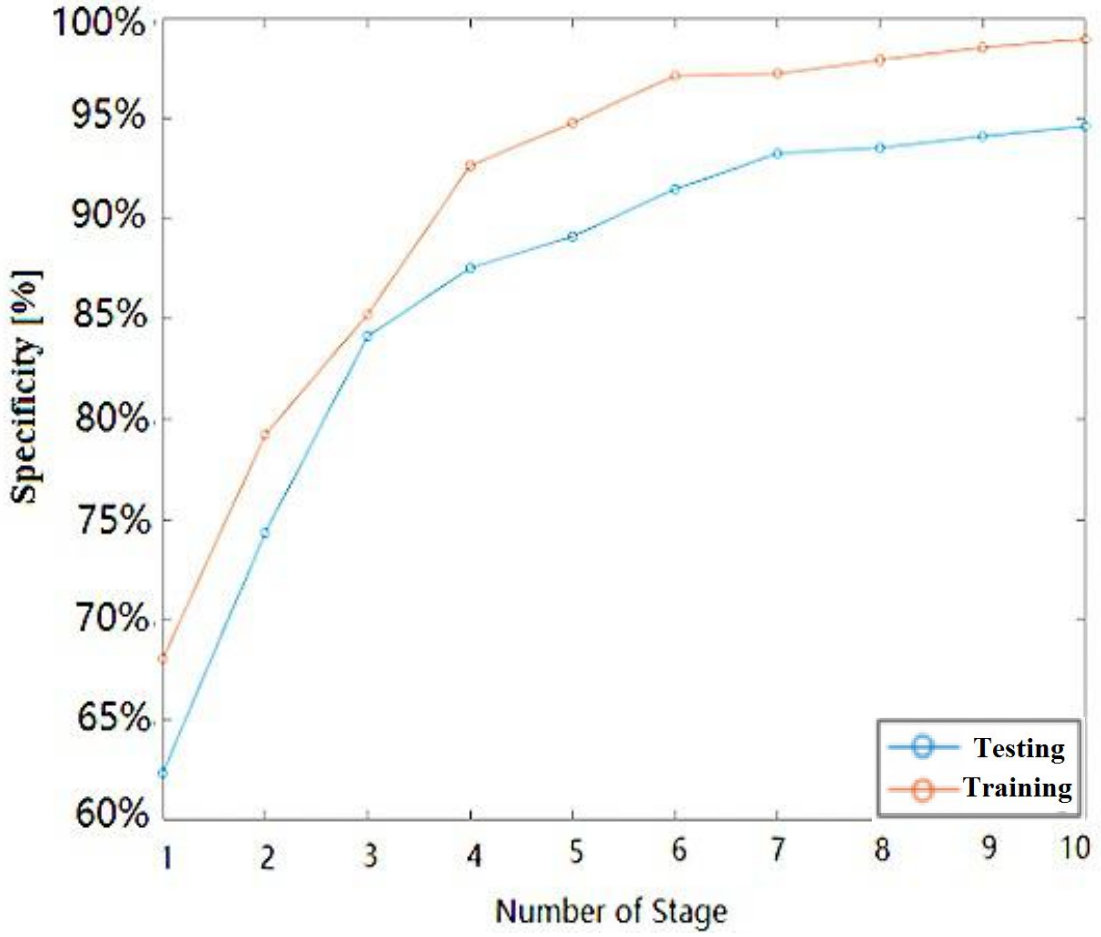

Figure 5.6: Average specificity of every stage for training and testing （random forest）.

In Figure 5.1--5.6, we compare the average specificities and sensitivities of every stage for training and testing with three classification models respectively. Here we explain some main results observed from the built cascades and their performances:

1. In testing phase, for the cascade of decision trees, the final average testing sensitivity is 86.22%, the average testing specificity was 98.94%. The average testing sensitivity and specificity are 86.97% and 98.89% respectively in the cascade of SVMs,. The cascade of random forests reserve 87.32% of the spikes, while rejecting 98.84% of the backgrounds. We can find all the sensitivities and specificities of the final stage are very high, therefore, we can conclude that the cascade of classifiers works on the spikes reservation and backgrounds rejection.

2. In training phase, in order to build the cascade, we need to find the winning classifiers as well as the corresponding thresholds. The thresholds are selected such that 99.6% of the spikes of the input EEG data are preserved. Then, after 10 stages, the estimated sensitivity should be 96%. And final sensitivities of the cascade of decision trees, SVMs, random forests are 96.32%, 96.11%, 95.98% respectively. Therefore, we can regard that we succeed in the training phase.

3. As expected, we can see that the specificity increases after each stage. However, the increase in specificity gets smaller and smaller after each consecutive stage, and saturates gradually. On the contrary, the sensitivity decrease after each stage. According to the results, in all cases, the training sensitivities are larger than the testing sensitivities at the same steps, and the training specificities are smaller than the testing specificities.

In the Table 5.1, it details average precision of every stage in cascade of decision trees (DT), support vector machines (SVM), random forests(RF) in testing phase. We can find that precision increases after each consecutive stage.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
DT	**0.019**	**0.031**	**0.034**	**0.051**	**0.086**	**0.097**	**0.106**	**0.111**	**0.119**	**0.128**
SVM	**0.018**	**0.030**	**0.033**	**0.052**	**0.072**	**0.087**	**0.091**	**0.105**	**0.109**	**0.112**
RF	**0.018**	**0.032**	**0.046**	**0.052**	**0.063**	**0.078**	**0.081**	**0.101**	**0.110**	**0.117**

Table 5.1: Average precision of every testing stage.

Chapter 6

Summary

6.1 Summary

In the thesis, the objective is to find a method which can be used an automated way to reserved spikes and backgrounds rejection as many as possible. Fortunately, we have a huge amount of epileptic patients' EEG data and labeled by experts. we proposed a method of the cascade of classifiers to do backgrounds rejection and spikes reservation on (interictal) EEG of epileptic patients. And, we applied decision tree, random forest, and SVM as the classification methods.

At last, we found 10 winning classifiers with corresponding thresholds, we applied the cascade of 10 winning classifiers to do testing. According to the results, the cascade of classifiers performed quite well, and the cascade of random forest performs better than the other two methods in this project, I think the reason is that random forest is suitable to classify the huge data with many features, but decision tree and SVM may cause the problem of overfitting.

6.2 Future Recommendation

As for the future work, there are some areas we can work on this project.

(1) Because in this study, we used the all raw data to do classification. All the spikes and backgrounds are extracted from the epileptic patients. Considering the overlap between the spikes and backgrounds or backgrounds of patients' EEG data may share some similar attributes with spikes, next step, we plan to extract the backgrounds from normal people EEG data or spike-free EEG data, this way may do help to build better classifiers.Backgrounds from pike-free EEG data may show some special features that epileptic patients' EEG data do not have.

(2) We plan to apply diverse features, like discrete wavelet transform, continuous wavelet transform, or nonlinear energy operator, etc. By taking advantage of variable features of the EEG will contribute to increasing the each stage's reservation rate. Another way to increase the feature types to adapt different types of mother wavelets. After that, we will do features selection. Selecting the most typical features will increase the efficiency of machine learning.

(3) What's more, we would like to reset the threshold to a bigger value, making more spikes reserved in every stage. By reset higher thresholds, we hope to make the length of the cascade longer, according to the estimation, longer length of cascade may achieve better performance.

(4) In addition, we plan to do post-processing for the reserved EEG data after background cascade of classifiers, like using deep learning.

(5) In this thesis, we applied random forest, the number of trees in the random forest function was 30, actually, this number can be optimized. More trees may be higher accuracy but need more time to process and sometimes overfitting may happen.

REFERENCE

[1] Engel J. A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: report of the ILAE Task Force on Classification and Terminology[J]. Epilepsia, 2001, 42(6): 796-803.

[2] Thomas J, Jin J, Dauwels J, et al. Clustering of interictal spikes by dynamic time warping and affinity propagation[C]//Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016: 749-753.

[3] Tibdewal M N, Tale S A. Multichannel detection of epilepsy using SVM classifier on EEG signal[C]//Computing Communication Control and automation (ICCUBEA), 2016 International Conference on. IEEE, 2016: 1-6.

[4] WHO，"Key facts”，2016, [Online],

Available :http://www.who.int/mediacentre/factsheets/fs999/en/

[5] Prabhakar S K, Rajaguru H. Conceptual analysis of epilepsy classification using probabilistic mixture models[C]//Brain-Computer Interface (BCI), 2017 5th International Winter Conference on. IEEE, 2017: 81-84.

[6] Prabhakar S K, Rajaguru H. Conceptual analysis of epilepsy classification using probabilistic mixture models[C]//Brain-Computer Interface (BCI), 2017 5th International Winter Conference on. IEEE, 2017: 81-84.

[7] Halford, J.J., Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation. Clinical Neurophysiology, 2009. 120(11): p. 1909-1915.

[8] Liu, Y.-C., et al., Model-Based Spike Detection of Epileptic EEG Data. Sensors, 2013. 13(9): p. 12536-12547.

[9] Nonclercq, A., et al., Cluster-based spike detection algorithm adapts to interpatient and intrapatient variation in spike morphology. Journal of Neuroscience Methods, 2012. 210(2): p. 259-265

[10] Lodder, S.S., J. Askamp, and M.J.A.M. van Putten, Inter-ictal spike detection using a database of smart templates. Clinical Neurophysiology, 2013. 124(12):

p. 2328-2335.

[11] Liu, Y.-C., et al., Model-Based Spike Detection of Epileptic EEG Data. Sensors, 2013. 13(9): p. 12536-12547.

[12] Zhang, J., et al., Automatic detection of interictal epileptiform discharges based on time-series sequence merging method. Neurocomputing, 2013. 110(0): p. 35-43.

[13] Chavakula, Vamsidhar, et al. "Automated quantification of spikes." Epilepsy & Behavior 26.2 (2013): 143-152.

[14] Urbanowicz R J, Moore J H. Learning classifier systems: a complete introduction, review, and roadmap[J]. Journal of Artificial Evolution and Applications, 2009, 2009: 1.

[15] Xue M, Zhu C. A study and application on machine learning of artificial intellligence[C]//Artificial Intelligence, 2009. JCAI'09. International Joint Conference on. IEEE, 2009: 272-274.

[16] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]//Ijcai. 1995, 14(2): 1137-1145.

[17] Wikipedia, “Machine learning,” 2017. [Online].

Available: https://en.wikipedia.org/wiki/Machine_learning.

[18] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN 9780262018258.

[19] Acr, Nurettin, and Cneytzeli. "Automatic spike detection in EEG by a two-stage procedure based on support vector machines."*Computers in Biology and Medicine*34.7 (2004): 561-575.

[20] Kumari, R., and J. Jose. "Seizure detection in EEG using biorthogonal wavelet and fuzzy KNN classifier." *Elixir Hum. Physiol* 41 (2011): 5766-5770.

[21] Polat, Kemal, and Salih Güneş. "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform." *Applied Mathematics and Computation* 187.2 (2007): 1017-1026.

[22] Güler, Nihal Fatma, Elif Derya Übeyli, and Inan Güler. "Recurrent neural networks employing Lyapunov exponents for EEG signals classification." *Expert systems with applications* 29.3 (2005): 506-514.

[23] Kannathal, N., et al. "Entropies for detection of epilepsy in EEG." *Computer methods and programs in biomedicine* 80.3 (2005): 187-194.

[24] Joachims T. Making large-scale SVM learning practical[R]. Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, 1998.

[25] an, Yaozhang, et al. "Detection of epileptic spike-wave discharges using SVM." *Control Applications, 2007. CCA 2007. IEEE International Conference on*. IEEE, 2007.

[26] Nicolaou, Nicoletta, and Julius Georgiou. "Detection of epileptic electroencephalogram based on permutation entropy and support vector machines." *Expert Systems with Applications* 39.1 (2012): 202-209.

[27] Andrzejak, R. G., Lehnertz, K., Rieke, C., Mormann, F., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Physical Review E, 64, 061907.

[28] Kumar, S. P., Sriraam, N., Benakop, P. G., & Jinaga, B. C. (2010). Entropies based detection of epileptic seizures with artificial neural network classifiers. Expert Systems with Applications, 37, 3284–3291.

[29] Polat, Kemal, and Salih Güneş. "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform." *Applied Mathematics and Computation* 187.2 (2007): 1017-1026.

[30] Srinivasan, V., Eswaran, C., & Sriraam, N. (2007). Approximate entropy-based epileptic EEG detection using artificial neural networks. IEEE Transactions on Information Technology in Biomedicine, 11, 288–295.

[31] Srinivasan, V., Eswaran, C., & Sriraam, N. (2005). Artificial neural network based epileptic detection using time-domain and frequency-domain features. Journal of Medical Systems, 29, 647–660.

[32] Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2007). Automatic seizure detection based on time-frequency analysis and artificial neural networks. Computational Intelligence and Neuroscience, 80510 (80513 p.).

[33] Übeyli, E. D. (2009). Combined neural network model employing wavelet coefficients for EEG signals classification. Digit Signal Process, 19, 297–308

[34] Safavian S R, Landgrebe D. A survey of decision tree classifier methodology[J]. IEEE transactions on systems, man, and cybernetics, 1991, 21(3): 660-674.

[35] Wikipedia, “Decision tree,” 2017. [Online].

Available: https://en.wikipedia.org/wiki/Decision_tree.

[36] Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.

[37] Liaw A, Wiener M. Classification and regression by randomForest[J]. R news, 2002, 2(3): 18-22.

[38] Donos, Cristian, Matthias Dümpelmann, and Andreas Schulze-Bonhage. "Early seizure detection algorithm based on intracranial EEG and random forest classification." *International journal of neural systems* 25.05 (2015): 1550023.

[39] Wolpert D H. Stacked generalization[J]. Neural networks, 1992, 5(2): 241-259.

[40] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]//Ijcai. 1995, 14(2): 1137-1145.

[41] Freund Y, Schapire R E. A desicion-theoretic generalization of on-line learning and an application to boosting[C]//European conference on computational learning theory. Springer Berlin Heidelberg, 1995: 23-37.

[42] Mozos O M, Stachniss C, Burgard W. Supervised learning of places from range data using adaboost[C]//Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on. IEEE, 2005: 1730-1735.

[43] Li S Z, Zhang Z Q. Floatboost learning and statistical face detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(9): 1112-1123.

[44] Lienhart R, Kuranov A, Pisarevsky V. Empirical analysis of detection cascades of boosted classifiers for rapid object detection[C]//Joint Pattern Recognition Symposium. Springer Berlin Heidelberg, 2003: 297-304.

[45] iola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001, 1: I-I.

[46] Wikipedia, “Confusion matrix,” 2017. [Online].

Available: https://en.wikipedia.org/wiki/Confusion_matrix

[47] Hawkins, Douglas M. "The problem of overfitting." *Journal of chemical information and computer sciences* 44.1 (2004): 1-12.

[48] S. S. Lodder and M. J. van Putten, “A self-adapting system for the automated detection of inter-ictal epileptiform discharges,” PloS one, vol. 9, no. 1, p. e85180, 2014.

[49] A. Nonclercq, M. Foulon, D. Verheulpen, C. De Cock, M. Buzatu, P. Mathys, and P. Van Bogaert, “Cluster-based spike detection algorithm adapts to interpatient and intrapatient variation in spike morphology,” Journal of neuroscience methods, vol. 210, no. 2, pp. 259–265, 2012.

[50]] J. J. Halford, R. J. Schalkoff, J. Zhou, S. R. Benbadis, W. O. Tatum, R. P. Turner, S. R. Sinha, N. B. Fountain, A. Arain, P. B. Pritchard et al.,

"Standardized database development for eeg epileptiform transient detection: Eegnet scoring system and machine learning analysis," Journal of neuroscience methods, vol. 212, no. 2, pp. 308–316, 2013

[51] F. I. Argoud, F. M. De Azevedo, J. M. Neto, and E. Grillo, "Sade3: an effective system for automated detection of epileptiform events in long-term eeg based on context information," Medical and Biological Engineering and Computing, vol. 44, no. 6, pp. 459–470, 2006.

[52] H. J. Carey, M. Manic, and P. Arsenovic, "Epileptic spike detection with eeg using artificial neural networks," in Human System Interactions (HSI), 2016 9th International Conference on. IEEE, 2016, pp. 89–95

[53] Y. Song and J. Zhang, "Automatic recognition of epileptic eeg patterns via extreme learning machine and multiresolution feature extraction," Expert systems with applications, vol. 40, no. 14, pp. 5477–5489, 2013.

[54] C. Castellaro, G. Favaro, A. Castellaro, A. Casagrande, S. Castellaro, D. Puthenparampil, and C. F. Salimbeni, "An artificial intelligence approach to classify and analyse eeg traces," Neurophysiologie Clinique/Clinical Neurophysiology, vol. 32, no. 3, pp. 193–214, 2002.

[55] S. B. Wilson, C. A. Turner, R. G. Emerson, and M. L. Scheuer, "Spike detection ii: automatic, perception-based detection and clustering," Clinical neurophysiology, vol. 110, no. 3, pp. 404–411, 1999.

[56] C. J. James, R. D. Jones, P. J. Bones, and G. J. Carroll, "Detection of epileptiform discharges in the eeg by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages," Clinical Neurophysiology, vol. 110, no. 12, pp. 2049–2063, 1999.

[57] C. F. Boos, F. M. de Azevedo Geovani R. Scolaro, and V. P. Maria do Carmo, Automatic Detection of Paroxysms in EEG Signals using Morphological Descriptors and Artificial Neural Networks. INTECH Open Access Publisher, 2011.

[58] Z. H. ˙Inan and M. Kuntalp, "A study on fuzzy c-means clustering-based systems in automatic spike detection," Computers in Biology and Medicine, vol. 37, no. 8, pp. 1160–1166, 2007.

[59] A. Tzallas, P. Karvelis, C. Katsis, D. Fotiadis, S. Giannopoulos, and S. Konitsiotis, "A method for classification of transient events in eeg recordings: application to epilepsy diagnosis," Methods of Information in Medicine, vol. 45, no. 6, pp. 610– 621, 2006.

[60] T. P. Exarchos, A. T. Tzallas, D. I. Fotiadis, S. Konitsiotis, and S. Giannopoulos, "Eeg transient event detection and classification using

association rules," IEEE Transactions on Information Technology in Biomedicine, vol. 10, no. 3, pp. 451–457, 2006.

[61] N. Acir, I. Oztura, M. Kuntalp, B. Baklan, and C. Guzelis, "Automatic detection of epileptiform events in eeg by a three-stage procedure based on artificial neural networks," IEEE Transactions on Biomedical Engineering, vol. 52, no. 1, pp. 30–40, 2005.

[62] M. Adjouadi, D. Sanchez, M. Cabrerizo, M. Ayala, P. Jayakar, I. Yaylali, and A. Barreto, "Interictal spike detection using the walsh transform," IEEE Transactions on Biomedical Engineering, vol. 51, no. 5, pp. 868–872, 2004.

[63] M. Adjouadi, M. Cabrerizo, M. Ayala, D. Sanchez, I. Yaylali, P. Jayakar, and A. Barreto, "A new mathematical approach based on orthogonal operators for the detection of interictal spikes in epileptogenic data." Biomedical sciences instrumentation, vol. 40, pp. 175–180, 2003.

[64] D. Flanagan, R. Agarwal, Y. Wang, and J. Gotman, "Improvement in the performance of automated spike detection using dipole source features for artefact rejection," Clinical Neurophysiology, vol. 114, no. 1, pp. 38–49, 2003.

[65] T. Sugi, M. Nakamura, A. Ikeda, and H. Shibasaki, "Adaptive eeg spike detection: determination of threshold values based on conditional probability," Frontiers of Medical and Biological Engineering, vol. 11, no. 4, pp. 261–277, 2001.

[66] M. A. Black, R. D. Jones, G. J. Carroll, A. A. Dingle, I. M. Donaldson, and P. J. Parkin, "Real-time detection of epileptiform activity in the eeg: a blinded clinical trial," Clinical EEG and Neuroscience, vol. 31, no. 3, pp. 122–130, 2000.

[67] S. B. Wilson, C. A. Turner, R. G. Emerson, and M. L. Scheuer, "Spike detection ii: automatic, perception-based detection and clustering," Clinical neurophysiology, vol. 110, no. 3, pp. 404–411, 1999.

[68] W. E. Hostetler, H. J. Doller, and R. W. Homan, "Assessment of a computer program to detect epileptiform spikes," Electroencephalography and clinical neurophysiology, vol. 83, no. 1, pp. 1–11, 1992.

[69] J. Gotman and L.-Y. Wang, "State dependent spike detection: validation," Electroencephalography and clinical neurophysiology, vol. 83, no. 1, pp. 12–18, 1992.

[70] P. G. De Oliveira, C. Queiroz, and F. L. Da Silva, "Spike detection based on a pattern recognition approach using a microcomputer," Electroencephalography and clinical neurophysiology, vol. 56, no. 1, pp. 97–103, 1983.

[71] J. Gotman, J. Ives, and P. Gloor, "Automatic recognition of inter-ictal epileptic

activity in prolonged eeg recordings," Electroencephalography and clinical neurophysiology, vol. 46, no. 5, pp. 510–520, 1979.

[72] J. Gotman and P. Gloor, "Automatic recognition and quantification of interictal epileptic activity in the human scalp eeg," Electroencephalography and clinical neurophysiology, vol. 41, no. 5, pp. 513–529, 1976

[73] M. Adjouadi, M. Cabrerizo, M. Ayala, D. Sanchez, I. Yaylali, P. Jayakar, and A. Barreto, "A new mathematical approach based on orthogonal operators for the detection of interictal spikes in epileptogenic data." Biomedical sciences instrumentation, vol. 40, pp. 175–180, 2003.

[74] T. Sugi, M. Nakamura, A. Ikeda, and H. Shibasaki, "Adaptive eeg spike detection: determination of threshold values based on conditional probability," Frontiers of Medical and Biological Engineering, vol. 11, no. 4, pp. 261–277, 2001.

[75] M. A. Black, R. D. Jones, G. J. Carroll, A. A. Dingle, I. M. Donaldson, and P. J. Parkin, "Real-time detection of epileptiform activity in the eeg: a blinded clinical trial," Clinical EEG and Neuroscience, vol. 31, no. 3, pp. 122–130, 2000.

[76] T. Pietil¨a, S. Vapaakoski, U. Nousiainen, A. V¨arri, H. Frey, V. H¨akkinen, and Y. Neuvo, "Evaluation of a computerized system for recognition of epileptic activity during long-term eeg recording," Electroencephalography and clinical neurophysiology, vol. 90, no. 6, pp. 438–443, 1994

APPENDIX A

Folds	Loss/reserved spike	Loss/reserved Backgrounds	Sensitivity	Specificity	Precision
Set1	452/4127	24310/ 1738521	0.9011	0.9862	0.1451
Set2	337/2721	21963/ 2568531	0.8878	0.9915	0.1102
Set3	514/3197	28618/ 2180211	0.8516	0.9870	0.1005
Set4	394/1809	11859/ 2297934	0.8201	0.9949	0.1324
Set5	631/3802	21193/ 1801315	0.8566	0.9884	0.1521

Table A.1: Testing results of the final stage of the cascade of classifiers (decision tree).

Folds	Loss/reserved spike	Loss/reserved Backgrounds	Sensitivity	Specificity	Precision
Set1	364/4394	34357/ 1728521	0.9234	0.9861	0.1134
Set2	357/2699	20750/ 2569732	0.8828	0.9920	0.1151
Set3	585/3124	22683/ 2186134	0.8402	0.9897	0.1211
Set4	435/1768	17245/ 2292514	0.8013	0.9925	0.093
Set5	436/3996	28950/ 1793541	0.9015	0.9846	0.1213

Table A.2: Testing results of the final stage of the cascade of classifiers (SVM).

Folds	Loss/reserved spikes	Loss/reserved Backgrounds	Sensitivity	Specificity	Precision
Set1	474/4284	29774/ 1733123	0.9003	0.9831	0.1258
Set2	356/2700	23142/ 2567542	0.8834	0.9911	0.1045
Set3	588/3122	20458/ 2188923	0.8413	0.9907	0.1324
Set4	296/1907	16757/ 2292812	0.8655	0.9927	0.1022
Set5	550/3882	28180/ 1794302	0.8759	0.9845	0.1211

Table A.3: Testing results of the final stage of the cascade of classifiers (random forest).

	1	2	3	4	5	6	7	8	9	10
Sensitivity	**0.991**	**0.978**	**0.974**	**0.973**	**0.972**	**0.970**	**0.969**	**0.966**	**0.964**	**0.963**
Specificity	**0.623**	**0.743**	**0.841**	**0.875**	**0.891**	**0.914**	**0.932**	**0.935**	**0.941**	**0.946**

Table A.4: Average sensitivity and specificity of every training stage.

	1	2	3	4	5	6	7	8	9	10
Sensitivity	**0.964**	**0.939**	**0.932**	**0.910**	**0.894**	**0.884**	**0.880**	**0.872**	**0.869**	**0.8622**
Specificity	**0.680**	**0.792**	**0.852**	**0.926**	**0.947**	**0.961**	**0.972**	**0.979**	**0.985**	**0.9894**

Table A.5: Average sensitivity and specificity of every testing stage.

	1	2	3	4	5	6	7	8	9	10
Sensitivity	**0.992**	**0.980**	**0.971**	**0.970**	**0.968**	**0.967**	**0.966**	**0.964**	**0.963**	**0.9611**
Specificity	**0.584**	**0.737**	**0.853**	**0.885**	**0.905**	**0.914**	**0.922**	**0.925**	**0.935**	**0.9392**

Table A.6: Average sensitivity and specificity Of every training stage.

	1	2	3	4	5	6	7	8	9	10
Sensitivity	**0.958**	**0.937**	**0.928**	**0.908**	**0.895**	**0.889**	**0.880**	**0.875**	**0.871**	**0.8697**
Specificity	**0.690**	**0.788**	**0.861**	**0.904**	**0.947**	**0.971**	**0.974**	**0.979**	**0.983**	**0.9889**

Table A.7: Average sensitivity and specificity of every testing stage.

	1	2	3	4	5	6	7	8	9	10
Sensitivity	**0.988**	**0.983**	**0.979**	**0.975**	**0.973**	**0.971**	**0.969**	**0.967**	**0.965**	**0.9598**
Specificity	**0.714**	**0.817**	**0.832**	**0.875**	**0.905**	**0.913**	**0.922**	**0.927**	**0.935**	**0.9401**

Table A.8: Average sensitivity and specificity of every training stage.

	1	2	3	4	5	6	7	8	9	10
Sensitivity	**0.981**	**0.945**	**0.922**	**0.908**	**0.889**	**0.885**	**0.879**	**0.877**	**0.876**	**0.8732**
Specificity	**0.730**	**0.831**	**0.863**	**0.885**	**0.935**	**0.961**	**0.969**	**0.978**	**0.983**	**0.9884**

Table A.9: Average sensitivity and specificity of every training stage.