Circuits and Systems
Mekelweg 4,
2628 CD Delft
The Netherlands
https://cas.tudelft.nl/

CAS-2022-

M.Sc. Thesis

Energy-efficient seizure detection for wearable EEG

Xiaoning Shi

Abstract

With the development of machine learning techniques, more and more classification models have been designed for seizure detection. The creation of these models has dramatically improved the convenience of epilepsy detection and made seizure labeling automation possible. However, many of the current researches in this field use EEG datasets with small data volumes and are mainly designed for scientific purposes, which do not have a good performance of actual medical data. Besides, most models require complex time-frequency domain transformation and feature extraction process, which result in low classification speed and makes it difficult to achieve real-time monitoring. Moreover, the excessive complexity also means higher power consumption, so most of these models cannot be implemented with wearable EEG devices.

This thesis proposed a new seizure detection algorithm based on the bidirectional long short-term memory(BiLSTM) technique. The seizure detection function is achieved using time-domain features and LSTM networks. The preprocessing steps of this model are simple, and the complexity is low. Thus its operation speed is significantly improved compared to other traditional models. Also, this model is developed and tested based on TUH EEG corpus, which is an open-access dataset. Therefore, the results are directly comparable to others in the literature.

Energy-efficient seizure detection for wearable EEG

Thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

in

Electrical Engineering

by

Xiaoning Shi
born in Henan, China

This work was performed in:

Circuits and Systems Group
Department of Microelectronics
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Delft University of Technology

Delft University of Technology
Department of
Microelectronics

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **"Energy-efficient seizure detection for wearable EEG"** by **Xiaoning Shi** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 21 Oct. 2022

Chairman: _____
dr. B. Hunyadi

Advisors: _____
dr. B. Hunyadi

_____
dr. O.C. Akgun

Committee Members: _____
dr. B. Hunyadi

_____
dr. O.C. Akgun

_____
dr. F. Fioranelli

Abstract

With the development of machine learning techniques, more and more classification models have been designed for seizure detection. The creation of these models has dramatically improved the convenience of epilepsy detection and made seizure labeling automation possible. However, many of the current researches in this field use EEG datasets with small data volumes and are mainly designed for scientific purposes, which do not have a good performance of actual medical data. Besides, most models require complex time-frequency domain transformation and feature extraction process, which result in low classification speed and makes it difficult to achieve real-time monitoring. Moreover, the excessive complexity also means higher power consumption, so most of these models cannot be implemented with wearable EEG devices.

This thesis proposed a new seizure detection algorithm based on the bidirectional long short-term memory(BiLSTM) technique. The seizure detection function is achieved using time-domain features and LSTM networks. The preprocessing steps of this model are simple, and the complexity is low. Thus its operation speed is significantly improved compared to other traditional models. Also, this model is developed and tested based on TUH EEG corpus, which is an open-access dataset. Therefore, the results are directly comparable to others in the literature.

Acknowledgments

When I first arrived in the Netherlands two years ago, everything around me was exactly new to me. It felt like I have opened the door to a new world. I had to adapt to the new environment and the mode of learning. When the courses were completed, a new challenge was presented to me, the thesis. In the past year, I have been struggling with a lack of background knowledge, the pain of learning new knowledge, and great self-doubts. But today I can finally say that I resisted all the pressure and finished it. Looking back on these two challenging and unforgettable years, I have gained much more than I have ever imagined. And this cannot be done without the countless people who have helped me.

First of all, I would like to thank my supervisor Professor Borbála Hunyadi for guiding me during the past year. She has led me step by step through meeting after meeting to complete this thesis. She has helped me not only in pointing out the direction of my research but also in gradually inspiring me and guiding me to learn how I should complete a research process independently. Also, I would like to thank my other supervisor Doctor Can Akgun for providing me with a lot of guidance on hardware. He taught me how to find a balance between desired goals and feasibility. Their selfless help has led me to complete this thesis.

Also, I would like to thank my friends. Chen and Kangjie, the regular weekly dinner/walking is some of my most memorable memories of delft. I will never forget those long nights of schmoozing. It is these warm memories that have sustained me through countless difficult times. Yongding, always dubbed by myself as the Roland daily supervisor, thanks to your help, I was able to find the bottleneck of the model. Thank you Yitong, we were working on the thesis at the same time, and your sharing made me feel that I was not alone in this journey. Also, I would like to thank my friends in China, Chenxuan, Kangqi, Yating, Xiangruo, Zheng, and so on. The company you have given me across the time difference has given me endless strength.

Last but most importantly, I would like to thank my parents. I come from a small city, but you have offered me the best life you can offer. It is by standing on your shoulders that I have been able to see the bigger and better world. You are the best parents in the world.

Writing here, my master student life is also coming to an end. I hope that in my future life I will take all the things I have learned in these two years to walk forward bravely.

Xiaoning Shi
Delft, The Netherlands
21 Oct. 2022

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context

Epilepsy is the fourth most common serious brain disorder worldwide, affecting about 50 million people in total. This number is greater than the total number of patients with Parkinson's disease, multiple sclerosis, and cerebral palsy. It's estimated that the proportion of the general population with active epilepsy is between 4 and 10 per 1000 people [1]. Americas had the highest percentage of people with epilepsy among WHO Regions in the world. Approximately 3.4 million persons, 1.2% of the whole population of the U.S., suffered from epilepsy, according to the 2015 National Health Interview Survey (NHIS) [2]. In Europe, this percentage is 0.82%, lower than the average value of the world, but still a significant number. A seizure is a burst of uncontrolled and abnormal electrical neuronal activity in the brain [3]. The seizure itself is not a kind of disease, but it is the most critical symptom of epilepsy, which may result in varying complications, from a few seconds of blank stare to a few minutes of losing consciousness or awareness. Epilepsy is characterized by repeated seizures. A patient can have more than one type of seizure, which differ from person to person [4]. These seizures are different in length and frequency. Therefore it is extremely important to keep a record of the occurrence of seizures either to analyze the cause of the disease or determine the best treatment.

Epileptic seizures result from abnormal electrical activities originating from the brain. When seizures occur, the patterns of electrical signals among neurons become abnormally synchronous, which are highly correlated with the type of epilepsy. Electroencephalography(EEG) is a method to record electrical activities of the brain by measuring the voltage fluctuations resulting from ionic current within the large neuronal population of the brain [5]. Routine outpatient EEG recording strategy for epilepsy diagnosis uses multi-channel electrodes attached to the scalp in clinical circumstances and usually lasts no more than half an hour. Because of the short testing time, these EEG data may sometimes not be sufficient to make a diagnosis. Furthermore, for some other types of EEG tests, it may take up to a few days. Consequently, a seizure detection device worn in daily life is useful for maintaining a seizure diary which can help the caretakers get more complete and accurate information about seizure activity. It also facilitates a more appropriate follow-up medication. Such a seizure detection device must be small and discrete so that patients can wear it in their daily life. Moreover, the device has to be powered by a small battery for an extended time period, which calls for energy-efficient algorithmic solutions.

The traditional methods for designing custom application-specific integrated circuits are standard digital cell methods and mix-signal methods employing analog processing techniques in CMOS technologies. Time-mode signal processing (TMSP) techniques

use time-mode(TM) circuits to take the place of standard digital design, which is more power-efficient[6]. Since instead of transferring the real values between nodes, the information transmission in TM circuit is a binary switching event. It compares two values between two times. The output will be either 1(rising) or 0 (falling). The temporal signal is discontinuous but carries the information in its phase and length, making it an energy-efficient way for long-distance transfer. In order to implement the machine learning model with the TM circuit, the features and computations of the whole pipeline should be simple enough. Multiple machine learning models like Convolutional Neural Network(CNN) and Long Short-term (LSTM) can be applied to develop an energy-efficient seizure detection algorithm. Convolutional Neural Network(CNN) is a kind of feed-forward neural network widely used in image detection. It was designed based on deep neural networks. CNN structure usually consists of convolutional, ReLU, max-pooling, and fully connected layers[7]. In order to use this algorithm in seizure detection, some preprocessing steps are also required. Input data matrix needs to be built up by extracted data features. It is the most popular model for seizure detection nowadays. Recurrent neural networks(RNNs) take outputs of the previous state while having hidden states. It is widely used to process sequence data, since its computation takes historical information into account. Besides, the model size of RNNs will not change when the size of input increases. However, the vanishing and exploding gradient phenomena often occur when the number of layers increases. Because the gradient is multiplicative in this model. LSTM is a type of modified RNNs architecture[8]. It uses different gates to forget the unimportant part of the last cell, remember the important part of the input, and output the important part. This special structure solves the vanishing and the exploding gradient problem of the traditional RNNs model. Thus this simple and common model is also a reasonable choice for seizure detection. Besides, other strategies like feature extraction and model compression can also be used in the whole pipeline to improve accuracy and decrease computational complexity.

There has been a lot of research [9, 10, 11, 12] about the machine learning algorithms application of seizure detection. However, few of these algorithms were developed with the use of the LSTM model. This research aims to create an energy-efficient seizure detection algorithm based on time-domain features and LSTM techniques.

1.2 Problem statement

The main research problem of this thesis is:

Create an energy-efficient seizure detection machine learning algorithm using the LSTM model.

To solve this main question step by step, some subquestions have been formulated as follows:

- **Reduce the input data size with some preprocessing strategies**
  Raw EEG data contains multiple kinds of noise and artifacts of eye movements, blinks cardiac signals, and so on. To get clear input data, some filtering step is

required at the beginning of the whole pipeline. Also, the size of input data has a considerable influence on computation time. For the CNN model, the choice of features should also be a great issue. So, some feature extraction methods, both traditional or ml-based, will be used to reduce the dimensions of features.

- **Propose proper postprocessing steps to get the final result**
  The raw predictions of the machine learning model are term-based. It means that the predicted labels correspond to the class of each second in the whole sequence data. However, the objective of this problem is to detect seizure events, which should be an event-based problem. So some postprocessing steps should be applied for the raw predictions. Also, proper evaluation metrics should be selected for this project.

- **Design an LSTM-based model structure for seizure detection**
  Sequence data is often dealt with by RNN models. But because of the gradient vanishing/exploding problem, it has a limit on the length of the input sequence. Thus, it's not widely used in EEG-related problems. However, as an advanced RNN structure, LSTM is able to solve this problem. In this thesis, a new machine learning algorithm using LSTM models will be designed.

- **Optimize the model to decrease its complexity and improve its performance**
  In order to increase the accuracy and decrease the complexity, the model should be optimized from several aspects. The type of each layer, the number of layers, and the number of hidden units in each layer are all required to be discussed.

1.3 Structure

This thesis is structured as follows:

- Chapter 2 introduces the background of this thesis. It will briefly review this research, including the background knowledge of seizure and epilepsy, the state-of-the-art seizure detection application, the datatype of EEG signals, and the most commonly used machine learning models in this area.

- Chapter 3 discusses the preprocessing strategies for this problem. Some filtering steps were used to get the clean signal from raw data. And feature extraction methods were compared to choose the most valuable features as the input for machine learning models.

- Chapter 4 provides postprocessing steps of the predictions. Besides, the evaluation strategies are also explained in this chapter.

- Chapter 5 proposes the selection of the feature set and model structure. Different feature sets and model structures are compared to get the optimal choice. Also, the methods to select the decision threshold are discussed.

- Chapter 6 talks about different strategies to optimize the model. The input size is reduced with the use of the channel selection method. Furthermore, the model's performance is also improved due to structure optimization.
- Chapter 7 discusses the model's performance from different aspects and gives a more detailed analysis of the scenarios for which the model is applicable.
- Chapter 8 summarizes the thesis and analyzes the possible future research based on this work.

2 Background

2.1 Epilepsy

2.1.1 Overview

Epilepsy is the fourth most common neurological disorder in the world. Around 50 million people worldwide of all ages are suffering from it. The main causes of epilepsy are infections, abnormal immune function, genetic problems, metabolic problems, and abnormal brain structure. But the causes of epilepsy are different from person to person. And each different cause has different symptoms, diagnoses, and treatments. For newborns, epilepsy is always caused by congenital conditions or abnormal behavior during pregnancy. For example, a newborn baby with congenital brain malformations or metabolism problems will have epilepsy. Maternal drug use and lack of oxygen during birth also result in newborn epilepsy. For infants and children, epilepsy may also result from fever, infections, or brain tumors. The common causes for adults are similar to those of children, but some adult patients are caused by head trauma and progressive brain disease. As for the older, some common diseases of old age, like stroke and Alzheimer's disease, are also the main causes of epilepsy. Epilepsy is also known as a seizure disorder. It results in recurring and unprovoked seizures. A diagnosis of epilepsy usually requires at least two seizures with no known trigger that happens at least 24 hours apart.
Seizures involve sudden bursts of unusual and excessive electrical activities in the brain. These abnormal activities can cause involuntary changes in body movement or function, sensation, behavior, or awareness. A person with epilepsy has a lower seizure threshold than normal people, meaning they have a higher risk of seizures. However, seizures can also be the symptom of other medical problems, such as syncope, sleep disorder, migraine, and trauma. Sometimes, psychological problems will also cause Psychogenic Nonepileptic Seizure(PNES).

2.1.2 Diagnosis and treatment

Brain waves are oscillating electrical voltages caused by electrical activities in the brain. EEG is the most commonly used non-invasive method of brain waves measurement. The amplitude changes, polyspikes activities, and spike-wave discharges may occur during the paroxysm of seizures. The identification and labeling of seizure EEG signals are mainly made manually by specialists. It is also the golden standard for EEG-based epileptic seizure detection. However, this process is inefficient and cumbersome. In order to overcome these problems, automated EEG-based seizure detection methods have been a perennial hot topic. This issue can be classified into two directions, seizure event detection(SED) and seizure onset detection(SOD)[13]. SED aims to identify

seizure signals as accurately as possible from multiple EEG recordings. This is also the main function of the algorithm proposed by this thesis.

Treatment for epilepsy includes anti-seizure medication, surgery, or neurostimulation devices[14]. The preferred epilepsy treatment is medication. Most epileptic seizure attacks can be avoided by taking anti-epileptic medicine regularly. Moreover, patients with more severe symptoms can also be treated with medication to reduce the frequency and intensity of seizures attack. The type and dosage of epilepsy medication need to be decided by the doctor based on the patient's age, pathogen, and condition. The prescription will also be adjusted from time to time according to the patient's state of illness. The medication can be stopped when the patient has no seizure recurrence for two consecutive years. And the selection of antiseizure medication can be adjusted promptly by monitoring the EEG response to treatment. For example, the spike burden of generalized spike-and-wave will reduce when the therapy is effective [15].

Surgery may also be an option when the condition is more severe, which cannot be effectively controlled with the medication and meets the surgery requirements. Epilepsy that can be treated surgically must be focal epilepsy, where the seizures originate in a small and well-defined area. The area cannot be associated with critical behavioral functions, such as speech, language, movement, vision, hearing, etc. In epilepsy surgery, the surgeon will remove the related area of the brain that triggers the seizure to prevent seizure. Interictal epileptiform discharges suggest a greater likelihood of seizure recurrence [15]. Thus, routine EEG measurement is able to provide valuable information after epilepsy surgery.

In addition to the above two options, neuromodulation is another option. Vagus nerve stimulation(VNS) is an approach to neuromodulation for epilepsy, which has been approved by the U.S. Food and Drug Administration (FDA) as an add-on therapy. With this therapy, a pacemaker-like device will be implanted under the skin in the left chest area. The device will deliver pulses and stimulation through the vagus nerve to the brain to suppress abnormal electrical activities[16]. This method is not able to cure the disease. However, it can effectively reduce the frequency of seizures attack. With routine EEG measurement, doctors can better determine whether a patient is eligible for surgical treatment. And by observing the EEG after implantation, the device can be programmed patient-specifically.

2.2 EEG

2.2.1 History and principle

Biomedical signals and images are essential tools for the diagnosis and treatment of many different diseases. The commonly used electrobiological signals are electroencephalogram(EEG), a recording of brain activity, electromyogram(EMG), picked up from muscles and nerves, electrocardiogram(ECG), a recording of heart rhythm and its electrical activity, electrogastrogram(EGG), a recording of stomach's movements. EMG, ECG, and EGG have similar principles in those sensors on the skin that pick up the electrical signals caused by muscular activities. And EEG is the measurement of electrical activities in the brain. Neuronal communication is in the form of electro-

chemical currents. These activities can be measured by the electrodes placed on the scalp or near-ear. It has a good temporal resolution [17]. Biomedical signals can also be acquired with ultrasound or radiography methods, such as computerized tomography(CT), magnetic resonance imaging(MRI), positron emission tomography(PET), and functional MRI(fMRI). Sometimes, these measurements will be made synchronously. For example, The EEG-fMRI technique synchronously uses EEG and fMRI techniques to record functional MRI image signals and detect abnormal EEG activity. It has been shown that the localizing ability has been dramatically increased by using EEG-fMRI in a pre-surgical context [18].

Vladimir Vladimirovich Pravdich-Neminsky published the first animal EEG signal in 1912. It showed the EEG and the evoked potential of the mammalian brain[19]. And the first human EEG was recorded in 1924[20]. The conventional scalp EEG acquires brain electrical activities by placing metal electrodes with a conductive gel or paste on the scalp. And each of these electrodes is connected with an individual wire to a differential amplifier and filters[21]. Then analog-to-digital converters will be used to convert the analog signals to digital signals, which can be visualized as multichannel waves in the following steps. Most seizure activities are in the frequency range between 0.5 Hz to 29 Hz[22]. So historically, there is a clinical focus on activities below 40 Hz. However, more and more research about high-frequency oscillations (HFO) in the frequency range between 80-500 Hz has come out [23]. Besides, in order to extract more frequency domain features, some high-resolution EEG measurements are performed with a sampling frequency of 2000Hz. Nowadays, conventional EEG signals' most commonly used sampling frequency is not less than 200Hz. In other words, at least 200 samplings will be recorded per second during the measurement.

Although EEG data has a smaller memory size than radiological image data, it is still very memory volume-consuming. Figure 2.1 is a recording of 32-channel EEG data. This recording has a time duration of 80s, with a 400Hz sampling frequency, one seizure event is included. Assuming each signal sample is stored with 16bits, a widely used value, the total memory size will be equal to $32 \times 80 \times 400 \times 16 \approx 16.61$Mbits $= 2.01$Mbytes, excluding other information like labels, electrodes info and so on. With the same measurement device, the memory size of the data per day will be up to 2.06Gbytes.

2.2.2 International 10-20 system

The international 10-20 system is a standard electrode placement method, widely used in clinical as well as academic areas. It was proposed by Jasper HH presented at the 2nd International Congress of IFSECN in Paris in 1949 and published in 1958. It is still considered the most standardized system[24].

The placement is based on standard skull landmarks, consisting of nasion, inion, the left, and the right. The first measurement is taken in the anterior-posterior plane. A line is taken from the nasal root to the occipital ridge, then 5 points are marked on the line from front to back. They are named frontopolar midline(Fpz) point, frontal midline(Fz), central midline (Cz) point, parietal midline (Pz) point, and occipital midline (Oz) point. The distance from the Fpz point to the nasal root and Oz point to

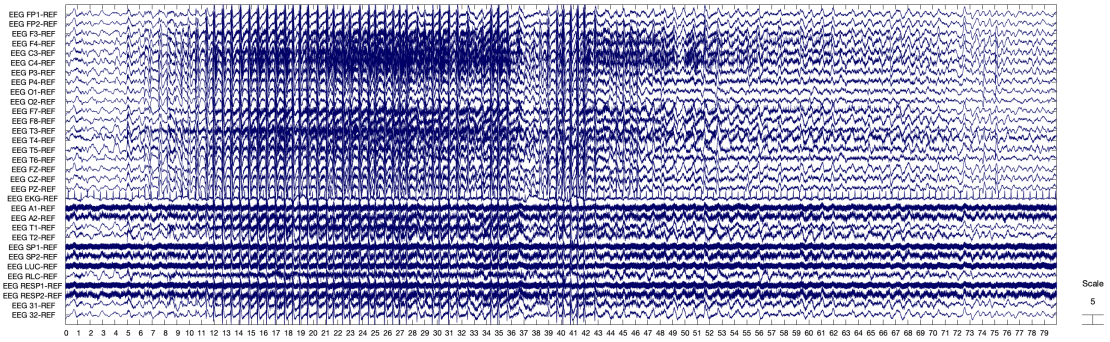


Figure 2.1: A recording of EEG data, 32 channels, sampling frequency = 400 Hz, duration = 80s

the occipital ridge account for 10% of the total length of this line, respectively. The remaining points Fz, Cz, and Pz are separated by 20% of the total length of this line.

The lateral measurement is taken from the Fpz point backward through the T3 and T4 points to the Oz point, the left and right side lines were taken, and then the left frontal pole (Fp1), right frontal pole (Fp2), left anterior temporal (F7), right anterior temporal (F8), left posterior temporal (T5), right posterior temporal (T6) and left occipital (O1) and right occipital (O2) points were marked symmetrically on this line from front to back. The distance from Fp1 and Fp2 to Fpz and from O1 and O2 to the Oz point each accounted for 10% of the total length of this line, and the remaining points were separated by 20% of the total length of this line.

The circumferential measurement is taken over the temporal lobes from the midline Fp position to the midline O position. The left temporal (T3) and right temporal (T4) points and the left central (C3) and right central (C4) points are marked symmetrically on the left and right sides of the line. 10% of the total length of the line is from the T3 and T4 points to the preauricular point, and the remaining points are separated by 20% of the total length of the line.

The remaining points, left frontal (F3) and right frontal (F4) points and left parietal (P3) and right parietal (P4) points are placed at the midpoint of the line between Fz point and F7 and F8, and the midpoint of the line between Pz point and T5 and T6, respectively. The left and right earlobe electrodes are indicated by A1 and A2, respectively.

The serial labels of the above electrodes are usually represented by an odd number for the left side and an even number for the right side. Eight electrodes are placed on the left and right sides of the head. Besides, there are three electrodes, Fz, Cz, and Pz, on the anterior and posterior plane and two earlobe electrodes on the left and right sides, so that a total of 21 electrodes are placed. Electrodes are placed in all major areas of the head[25].

The placement of the 10-20 system is illustrated as in Figure 2.2

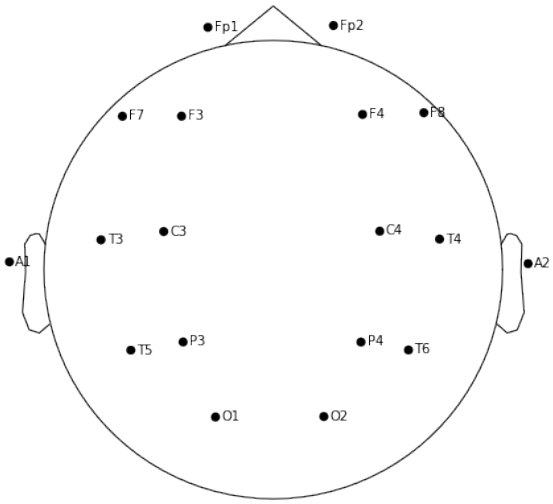


Figure 2.2: Electrode locations of International 10-20 system

2.2.3 Wearable EEG

Conventional on-scalp EEG uses wet electrodes to measure voltage fluctuations resulting from the brain's spontaneous activity. It is widely applied in medical diagnoses, such as epilepsy, anesthesia, sleep disorders, and so on[5]. However, the measurement device is expensive, massive, and needs to be operated by experts, restricting its non-medical application. The commonly used wearable EEG devices can be classified into three categories, headset-type, head-band type, and earphone-type. These devices are shown in Figure 2.3.

(a) iSyncWave, a wireless dry headset type EEG measurement and therapeutics device[26]

(b) BrainLink Lite v2.0, a head-band type EEG device[27]

(c) A ear EEG device from OpenBCI[28]

Figure 2.3: Three different types of wearable EEG devices

The headset-type EEG devices have the most comprehensive measurement range, measuring the brain signal from all related planes, including the prefrontal cortex,

parietal, temporal, and occipital areas. The placement of electrodes of these EEG devices is very similar to clinical EEG devices. Because of the variety of information extracted by these devices, they always have the highest accuracy compared with the other two types of wearable EEG devices. However, these devices are relatively bulky and heavy. Also, they require the wearers to keep their scalp tightly contacting with the device. So they are not suitable for long-term daily wear. Therefore it is mainly used for specific tasks within a short period. For instance, it performs as the brain-computer-interface(BCI) to control other devices or transmit information.

The head-band type devices measure EEG signals from the non-hair bearing areas, for example, the forehead and near-ear area. Compared with headset-type devices, these devices are lighter and more suitable for various daily scenarios. Also, it is more suitable for long-term wear. Besides, they are also very efficient. Because they acquire signals from the forehead to the prefrontal cortex, which is highly related to different mental states. However, on the other hand, the signals from the prefrontal cortex area also contain much more artifacts. Thus, this type of device cannot be used for applications that have requirements of high accuracy[29].

The headphone type EEG is also known as ear EEG. It is a new method that records EEG signals by the electrodes attached to the ear canal[30] or behind the ears [31]. It requires fewer electrodes and lower battery capacity. Some of them can be used without electrical gel[32, 33, 34, 35]. These characteristics give it great potential in non-medical applications, which always have higher requirements of wearable, portable, and user-friendly[30]. However, because of the limited number of electrodes, ear EEG includes fewer channel signals. So it has a lower accuracy compared with conventional multi-channel EEG. There is a trade-off between accuracy and wearability.

2.3 ML models

A general workflow of machine learning models for seizure detection is illustrated in Figure 2.4. The raw data are collected from the brain signals in the first step. Different types of data can be acquired with different types of monitoring tools, like EEG, ECoG, fMRI, and MRI. EEG is the most popular data to be used in this area. Most EEG datasets are collected with 10-20 International systems. Unwanted signals will contaminate the EEG data. So before extracting features from raw EEG data, the original signals must be filtered first. The unwanted signals can be divided into two classes: physiological artifacts and instrumental noise. Physiological artifacts are the result of physiological factors. The most common physiological noise types are cardiac, muscle, and ocular signals. The unwanted signals can also be caused by the measurement equipment itself, which is known as instrumental noise. All electrical equipment used in the experiment will emit electromagnetic fields more or less, resulting in unwanted signals from AC power lines, sensors, computers, etc. And during the measuring process, some motion noise can be detected when the tools of measurement are moved. In this step, the signal-to-noise(SNR) ratio should increase as much as possible. The feature extraction step will extract the features from different domains. Time domain features are the most simple and intuitive features. Frequency domain features are more diverse since they can be extracted from different frequency bands as essential characteristics of EEG

signals. But they can only be extracted after Fourier Transformation. So considering the convenience, some time-frequency features are also widely used in seizure detection models. Although frequency-domain features always contain more robust information [36, 37, 38]. The most important part is classification. In this step, a classifier will be built to detect seizures from signals. Both simple Machine Learning(ML) models and advanced Deep Learning(DL) models are applied for different objectives. These models will be introduced in this section. And finally, the predictions will be evaluated with different evaluation metrics.

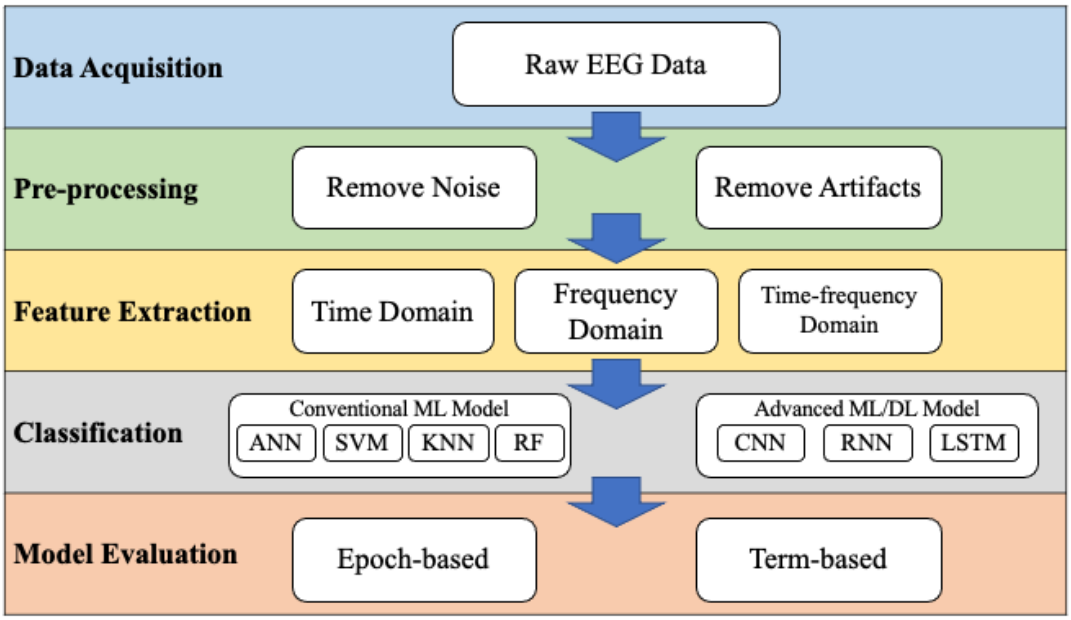


Figure 2.4: Basic machine learning-based epileptic seizure detection model

2.3.1 Conventional ML models

2.3.1.1 Artificial Neural Networks(ANN)

ANN model is one of the most simple and commonly used classifiers. It is a network structure formed by a large number of interconnected processing neurons, which is a kind of abstraction, simplification, and simulation of the organization and operation mechanism of the human brain. And the ANN model is an information processing system based on imitating the structure and function of neural networks in the brain, which simulates the neuronal activity with mathematical models. There are also many different types of ANNs. Hundreds of different ANN models have already been proposed since McCulloch and Pitts firstly built the first artificial neuron network in 1943[39]. The ANN models discussed in this part are simple ANN models only.

The first ANN model used for EEG seizure detection was developed by Webber et al. in 1996[40]. It consisted of three layers, an input layer, a hidden layer with 30 neurons, and an output layer with eight neurons. This research proves the feasibility

of utilizing the ANN model to detect seizures automatically. With a more complex feature extraction strategy, Guo et al. proposed an even simpler ANN model for the same task[41]. This work extracted line length features based on wavelet transform multi-resolution decomposition. The hidden layer of this ANN model only has ten neurons, which is much simpler than ever before. However, it has good performance because of its effective feature extraction method.

2.3.1.2 Support Vector Machine(SVM)

SVM is a supervised binary classifier. In the classification problem, each sample consists of multiple features, which create a feature space. The objective of SVM is to classify the samples into negative and positive classes. The SVM model represents instances as points in space so that the mapping allows two groups of points to be separated by as wide an apparent interval as possible. The new samples are then mapped into the same space, and the categories to which they belong are predicted based on which side of the interval they fall into[42]. This classifier can be used to discriminate between seizure and non-seizure epochs. As the principle of SVM, transforming data into high-dimensional space may make the model complex. So representative features should be extracted first.

Temko et al. used an SVM classifier to build a multi-channel patient-independent neonatal seizure detection system[43]. The feature vectors contained both time domain and frequency features. And based on its term-based evaluation metric, subsequent decisions made predictions for classification. Liu et al. conducted wavelet decomposition of multi-channel intracranial(iEEG) with five scales and selected three of them to extract features. These time-frequency features were sent to train an SVM classifier[44]. Zhang and Chen applied local mean decomposition(LMD) to decompose raw EEG signal data into s series of product functions(PFs), and features were calculated with different PFs. A genetic algorithm SVM was used to make classifications for the data[45].

2.3.1.3 K-Nearest Neighbor(KNN)

KNN is one of the simplest supervised machine learning algorithms that can be used for classification and regression. When the model starts running, the dataset will be loaded, and the classification starts directly. The principle is that if the majority of the K most similar samples in the feature space belong to one category, then the other samples will also belong to this category. In other words, the category of each sample is decided by the nearest one or more neighboring samples. Consequently, the KNN model is more suitable for sets of samples to be classified with more crossover or overlap of class domains compared with other ML models. On the other hand, it needs to count the distance between each test point and the training set, which makes it incredibly computationally intensive and time complex if the training set is large. This method is still an effective model when the dataset is not large.

Based on the univariate linear features in eight frequency sub-bands, Ghaderyan proposed a method by making classification with KNN model[46]. A good performance was achieved with this method. Birjandtalab et al. used a random forest algorithm

to select channels first and then applied a non-linear dimension reduction technique to reduce the feature dimension. In this research, KNN was used as an effective classifier to discriminate between seizure and non-seizure events[47]. Morphological features have also been used for the KNN model. In Shanir et al.'s research, local binary patterns(LBPs) assisted in capturing the rising and falling edges of EEG signals, and morphological features were extracted with them. The dimension of features is very low. As a result, a KNN classifier was used[48].

2.3.2 Random Forest (RF)

The decision tree is a tree-structure model used for classification or regression. In the classification tree, each internal node represents a judgment on an attribute to split data, each leaf represents the decision of a judgment, and finally, each leaf node represents a classification result. A random forest is a classifier that contains multiple decision trees whose output class is determined by the plurality of the output classes of the individual trees. Random forests were proposed by Tin Kam Ho in 1995 [49]. It is a classification model based on decision trees. The complexity of a single tree cannot grow arbitrarily because of the loss of generalization accuracy on unseen data. However, the classifier's capacity can be expanded arbitrarily by using multiple trees in multiple randomly selected subspaces, which is known as a random forest.

Donos et al. extracted 11 time domain and band power features for a random forest classifier [50]. It provided a relatively simple model for microcontroller implementation. Mursalin et al. proposed a method using the Improved Correlation-based Feature Selection method (ICFS) and Random Forest classifier (RF) together to detect epileptic seizures from EEG signals [51]. In this model, both time and frequency domain features are selected. The random forest can also be applied with autoregression (AR) based quadratic features. Zhang et al. have proposed a random forest classifier that uses the AR-based quadratic features extracted in the Variational mode decomposition (VMD) domain [52].

2.3.3 Advanced ML/DL models

2.3.3.1 Convolutional Neural Network (CNN)

CNN is one of the most influential models in image classification. It uses a convolution layer to extract features from the input image, which preserves the relationship between pixels. One-dimensional CNNs is a special type of CNNS, which performs 1D convolutions (scalar multiplications and additions) only[53]. It has already been widely used in the application of biomedical data classification. Instead of extracting different types of features first, CNNs perform not only the classifier but also the feature extractor. The model can directly be fed with time or frequency data.

Acharya et al. proposed a CNN model with five convolution layers, five max-pooling layers, and three fully-connected layers to make classification of normal, preictal, and seizure EEG signals[12]. This is the first research to apply CNNs to this field and made a great success. Different from Acharya's work which is based on raw time domain EEG signals, Zhou et al. proposed a CNN model for seizure detection based on EEG

data in frequency domain[9]. Since the EEG data contains more effective information in the frequency domain, this model performs considerably well.

2.3.3.2 Recurrent Neural Network (RNN)

Conventional DNN or CNN model consists of an input layer, hidden layers, and an output layer. In these models, layers are connected in one direction. Each layer takes the output of the last layer as the input and delivers the output to the next layer. Figure 2.5a shows the structure of a simple ANN model with three layers. The value of the output layer only depends on the value of the input layer. But different from conventional structure, the RNN model will only do outside loop but also loops within the hidden layers. The output of the RNN layer depends both on the input layer and the previous hidden states. It is widely used in sequence data processing problems. EEG is also a kind of multichannel sequence data. So some seizure detection RNN models also have been proposed.

(a) Structure of simple ANN model

(b) Structure of RNN model

Figure 2.5: Structures of simple ANN and RNN

A patient-specific seizure detection model was developed by Minasyan et al. with a variety of input features and RNNs. A 3-layer RNN was implemented and connected with the Levenberg-Marquardt learning algorithm in this model[54]. Vidyarante also proposed a patient-specific seizure detection model using the CNN model. However, it is different from the model used in this research is bidirectional RNNs, which contain the information from each side[55]. Both of these models have shorter run-time compared with the CNN model, which has a similar performance.

2.3.3.3 Long Short-term Memory (LSTM) networks

The backpropagation algorithm is a kind of gradient descent method that finds the steepest path to reach the minimum of the function. Considering the error value as a function of the network weights, the gradient descent method is used to minimize the error of different weights by taking steps with the size of the error gradient. Thus, the error gradient decides the direction and magnitude to calculate during the training process, which is used to update the weights of the network. This value is accumulative. So if the network is very deep and the value of the error gradient is larger than 1, the

accumulated value will be extremely large and result in an overflow of memory. This problem is called exploding gradients. Moreover, vanishing gradients has a similar cause to them. If the error gradient is close to 0, the accumulated value will be smaller and smaller and finally approach zero. In this case, the initial or lower layers will not be changeable anymore. In order to solve these problems, LSTM networks can be used instead of RNNs. Long Short-term Memory (LSTM) model is an advanced version of the RNN model. Its recurrent unit is much more complex than RNNs. The structure is displayed in Figure 2.6. As RNNs, LSTMs also have hidden-states that represent the working memory capability that transmits information from the most recent time step and is updated in each step. But LSTMs have a new type of state, cell-state. It represents the long memory capability that aggregates all data in the previous time steps. Besides, the major difference between LSTM recurrent units is the use of gates. The model has three gates: forget gate, input gate, and output gate. The cell-state c_t in each step conveys the information through the entire chain and is regulated by these gates.

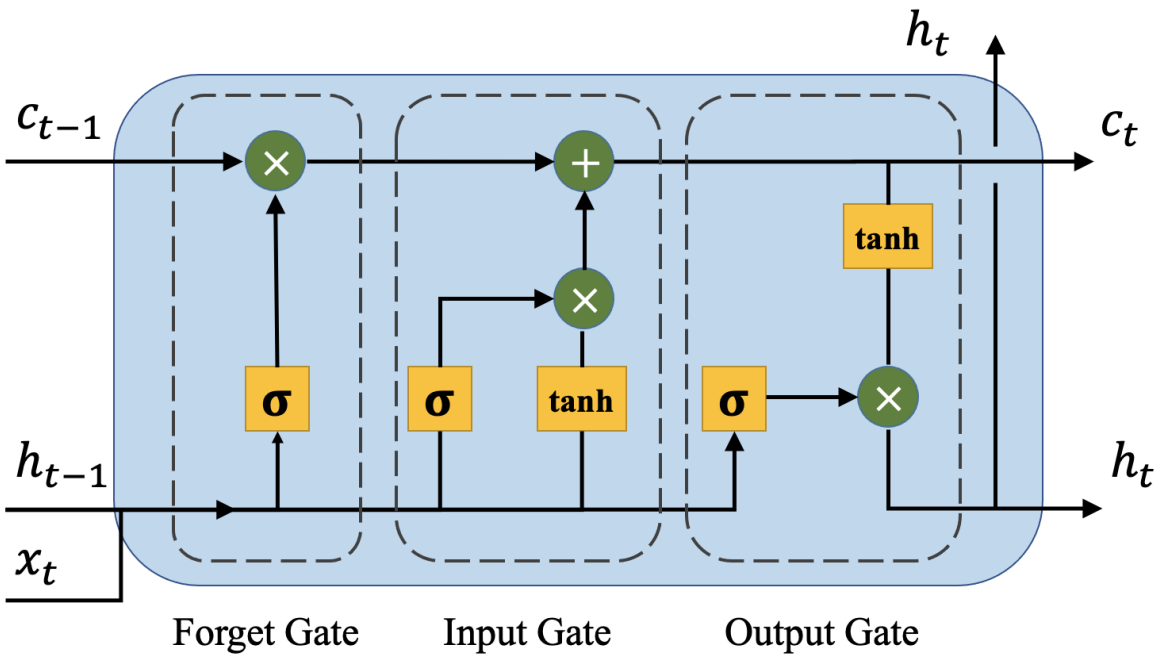


Figure 2.6: Recurrent unit of LSTM network

The forget gate decides which part of the long-term memory will be remembered and how much of this memory will be remembered. This gate has two inputs, h_{t-1}, the output of the previous hidden state, and x_t, the new input data. The sigmoid block is a sigmoid-activated neural network consisting of a weight matrix and bias matrix. The sigmoid-activated neural network is trained to make decisions on the input data. The value of matrix elements is within the range from 0 to 1. The relevant input will get an activation function whose output is close to 1. If the output of the activation function is close to 0, it means this information is irrelevant, and most of it will not be passed

to the next cell. The activation function of the forget gate is as follows, where W_t is the weight matrix at t, b_f is the bias vector of the function, and x_t is the input vector to the LSTM unit.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2.1}$$

The input gate decides which part of the new input data will be added to the cell-state, given the previous hidden state and new input data. The sigmoid block works as the input gate, and the tanh block works as the new memory network. The tanh block is a tanh-activated neural network that will be trained to combine the previous hidden state h_{t-1} and new input data x_t together. The output range of the tanh function is from -1 to 1, which makes it possible to reduce the impact of specific input data components by using negative activation function output. And in order to check the importance of new input data, a sigmoid block (input gate) is used meanwhile. The structure of the input gate is similar to the structure of forget gate. Then the outputs of these two blocks are pointwise multiplied. The result decides which part of the information will be updated to the cell-state. Equation 2.2 shows the details of these computations. i_t is the activation function of input gate. C_t and $\tilde{C}_t$ are the cell-state and cell-input vectors, respectively. b_i and b_C are bias vectors for these functions. W_o and w_C are weight matrices of them.

$$\begin{aligned} i_t &= \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \\ \tilde{C}_t &= \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \end{aligned} \tag{2.2}$$

The output gate makes the decision about the new hidden state. The input of this gate is newly updated cell-state, previous hidden state h_{t-1} and new input data x_t. And the activation block (output) gate is also similar to forget gate, which performs as a filter to decide which part to pass through. The newly updated cell-state will multiply with the output of the tanh activation function whose interval is [-1,1]. The process can be translated as Equation 2.3. o_t is the activation function of the output gate. W_o is the weight matrix of it, and b_o is the bias vector of it.

$$\begin{aligned} o_t &= \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \\ h_t &= o_t * \tanh\left(C_t\right) \end{aligned} \tag{2.3}$$

Since the raw EEG data is too complex, the applications of the LSTM model in seizure detection are always combined with the use of other feature extraction methods. The most commonly used combination is STFT and CNN, Shahbazi et al., Abdelhameed et al. [56] and Li et al. [57]'s researches all use this strategy. The difference is the type of LSTM model they used. The first study used a simple LSTM model. The second one used the Bidirectional LSTM (BiLSTM) model. And the last one used a nested LSTM model. There are also researches use statistical features. For example, the BiLSTM model proposed by Hu used the features obtained from local mean decomposition EEG data [58].

3 Preprocessing

3.1 Dataset overview

3.1.1 data structure

Temple University Hospital EEG(TUH-EEG) Corpus is one of the most extensive EEG archives in the world, which contains more than 30,000 clinical EEG recordings collected at Temple University Hospital(TUH) from 2022 - the present. It includes kinds of different subsets, such as The TUH EEG Artifact Corpus(TUAR), The TUH EEG Epilepsy Corpus(TUEP), The TUH EEG Events Corpus(TUEV), The TUH EEG Seizure Corpus(TUSZ), The TUH EEG Slowing Corpus(TUSL) and The TUH EEG Abnormal EEG Corpus(TUAB). TUSZ is the subset used in this thesis. It is a subset of TUH-EEG that contains manually annotated EEG data for seizures[59]. The dataset comes with a reference document, where metadata such as data type, recording location(ICU or clinical), recording type(long-term or routine), etc., are mentioned for each recording. This information can help us to do some related error analysis. For example, if the model fits too much with one or few specific types of seizures and if it really has the ability to detect all-length seizures or long seizures only. The signal data are stored as European Data Format(EDF) files. The metadata information about each patient session is also displayed in the header of each EDF file.

The corpus has 5610 different recordings, which come from 1423 different sessions. These sessions originate from 642 different patients. This works out to 2.2 sessions per patient on average. The basic statistics of the data are shown in Table 3.1.

Table 3.1: Statistics about TUSZ dataset

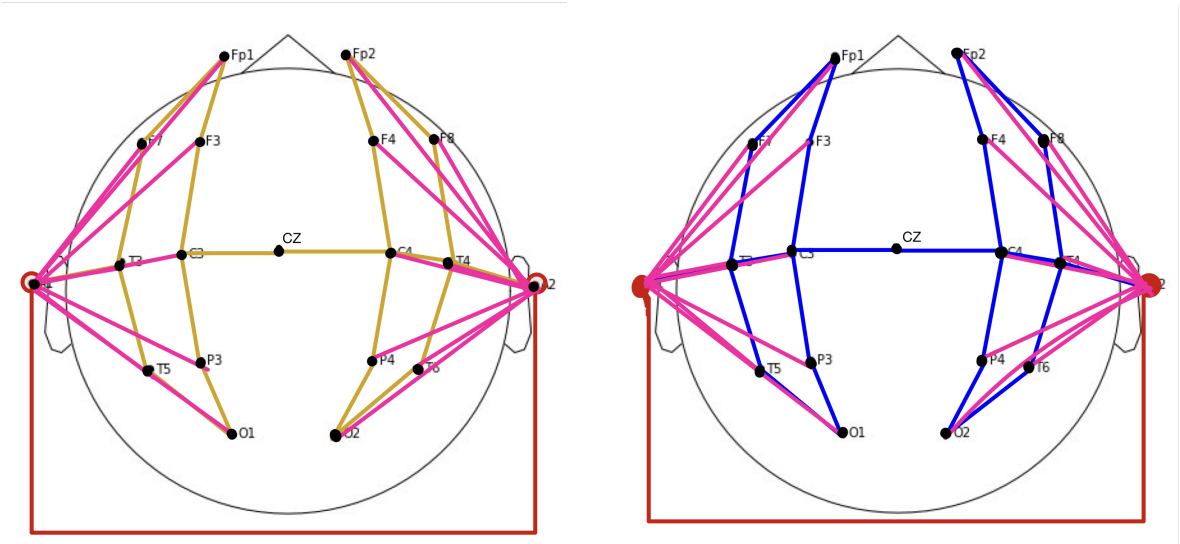
	TRAINING SET	DEVELOPMENT SET
Total files	4597	1013
Total sessions	1185	238
Total patients	592	50
Files with seizures	867	280
Patients with seizures	343	104
Total number of seizures	2370	673
Total duration	2,7008,284.00 secs	613,232.00 secs
Total background duration	2,540,144,77.00 secs (93.79%)	554,786.89 secs (90.47%)
Total duration of files with seizures	635,490.00secs (23.46%)	230,031.00 sec (30.51%)
Total seizure duration	168,139.23 secs (6.21%)	58,445.11 secs (9.53%)

3.1.2 montage and channel

The International 10-20 system is the most commonly used placement of electrodes in EEG recording. There are also some newly developed systems. For example, the 10-10 system increases the number of electrodes from 21 to 74. This change also enriches the richness of the dataset. However, whether 10-10 or 10-5 system is not as commonly used as the 10-20 system in clinical applications. In order to collect more valid EEG data with the same standard, the bulk of this dataset is collected with the 10-20 system. The EEG signals are micro-voltage signals, which should be recorded as differential voltage since this process can effectively reduce the noise while recording. In order to record the differential voltage with 16 bits of digital signals, a proper arrangement or array of channels on the EEG machine display, which is also named as montage, should be selected.

The EEG channel voltage shows the difference in potential voltage between two input electrodes. Theoretically, the electrode with zero potential, which means it has no EEG or any other bioelectrical activity, should be taken as the reference electrode. In this case, the voltage data of this channel will reflect the EEG signal of the measured electrode directly. However, every part of the human body suffers from bioelectrical activities continuously. It is impossible to find a surface with zero potential. So the parts which are less affected by various bioelectric fields and have fewer movements can also be considered as the location to place the reference electrode. Average Reference(AR) and Linked Ears Reference(LE) is the most widely used references. The AR montage connects a resistor in series with each electrode and then connects them in parallel. With this montage, the reference is the average of a certain number of electrodes. The LE montage takes earlobes as a reference. For the left cerebral hemisphere, the left earlobe is used as a reference. And for the right cerebral hemisphere, the right earlobe is used as a reference. Both of these two montages are unipolar montages. Figure 3.1 displays these two montages.

The above unipolar montages are always used for recording. However, for EEG data analysis, bipolar montages are more commonly used. The bipolar montage takes two adjacent electrodes and subtracts one from the other. Since the noise and artifacts are canceled due to a common reference point, they are more stable and have fewer artifacts than unipolar montages. One of the most popular bipolar Temporal Central Parasagittal (TCP) montages is used for EEG interpretation in this dataset. The TCP montage is also known as the double-banana montage. Figure 3.2 shows the configuration of it.



(a) AR montage

(b) LE montage

Figure 3.1: Location of reference for two montages

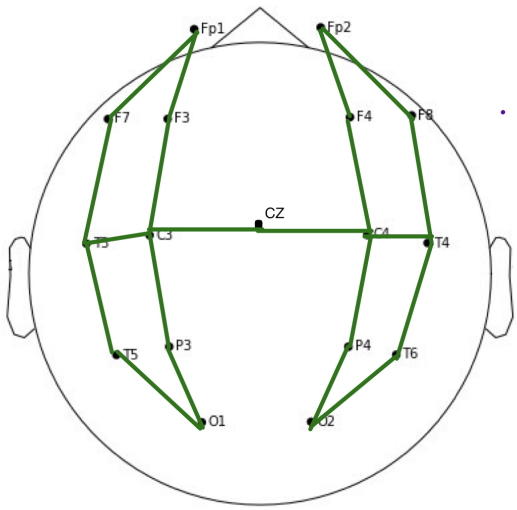


Figure 3.2: Electrodes placement and channels of 10-20 system with TCP montage

In order to reduce the size of input data, some channels are removed in this experiment. As the American Clinical Neurophysiology Society (ACNS) proposal, an 18-channel EEG dataset is selected. The information of this montage is shown in Table 3.2.

Table 3.2: Unipolar montage used in this project

Channel No.	Label	Montage Component
0	FP1-F7	EEG FP1-REF – EEG F7-REF
1	F7-T3	EEG F7-REF – EEG T3-REF
2	T3-T5	EEG T3-REF – EEG T5-REF
3	T5-O1	EEG T5-REF – EEG O1-REF
4	FP2-F8	EEG FP2-REF – EEG F8-REF
5	F8-T4	EEG F8-REF – EEG T4-REF
6	T4-T6	EEG T4-REF – EEG T6-REF
7	T6-O2	EEG T6-REF – EEG O2-REF
8	T3-C3	EEG T3-REF – EEG C3-REF
9	C3-CZ	EEG C3-REF – EEG CZ-REF
10	CZ-C4	EEG CZ-REF – EEG C4-REF
11	C4-T4	EEG C4-REF – EEG T4-REF
12	FP1-F3	EEG FP1-REF – EEG F3-REF
13	F3-C3	EEG F3-REF – EEG C3-REF
14	C3-P3	EEG C3-REF – EEG P3-REF
15	P3-O1	EEG P3-REF – EEG O1-REF
16	FP2-F4	EEG FP2-REF – EEG F4-REF
17	F4-C4	EEG F4-REF – EEG C4-REF

3.2 Data cleaning

The EEG data in the TUSZ corpus has been collected since 2002. Because of its long history, its origins are complicated and vary from intensive care unit(ICU) to outpatient service. As a consequence, these signals contain a wide variety of noise. Some cleaning processes should be applied to the raw data first. In this section, two data cleaning steps are introduced.

3.2.1 Filter tiny fluctuations

The multichannel EEG data is encoded and stored as 16-bit integers in the dataset. And after being loaded with python, this data will be stored as floating-point numbers series. The floating-point numbers are actually stored in binary in a computer, which uses a fixed number of binary digits to represent the decimal numbers. Inevitably, some decimal numbers cannot be represented correctly, some small round-off errors will occur. The nearest floating point number can result in maximum rounding error when it is halfway between two representable numbers. And in this case, the rounding error is 0.5 ULP. ULP is the unit in the last place. So the range of floating point error is fixed. Since the EEG data is recorded in scientific notation with the same number of significant figures. The floating point error is neglectable when the number is respectively large enough. However, when the real value is tiny, the error should be taken into consideration. And this error is accumulative, which will finally lead to

great error or even model crush in the following processes. So some signals with too many tiny values cannot be used in this project.

The number of tiny values in each recording is illustrated as Figure 3.3. In this figure, the data less than 1e-7 is considered tiny data, and the number of tiny data in each recording is counted firstly. Then the number of recordings with different amounts of tiny data is shown as a histogram in the figure. It can be concluded that there is an apparent gap between doable and undoable recordings, in which most of the recordings have no tiny data or only a few tiny data. As a result, the threshold in this step is set as 1e3. It means the recordings with less than 1e3 tiny data can be selected as normal data, and the recordings with more than 1e3 tiny data will be filtered in this step. There are 3175 sessions in the training set among which 47 sessions are undoable recordings. So after this step, 3128 sessions are reserved.

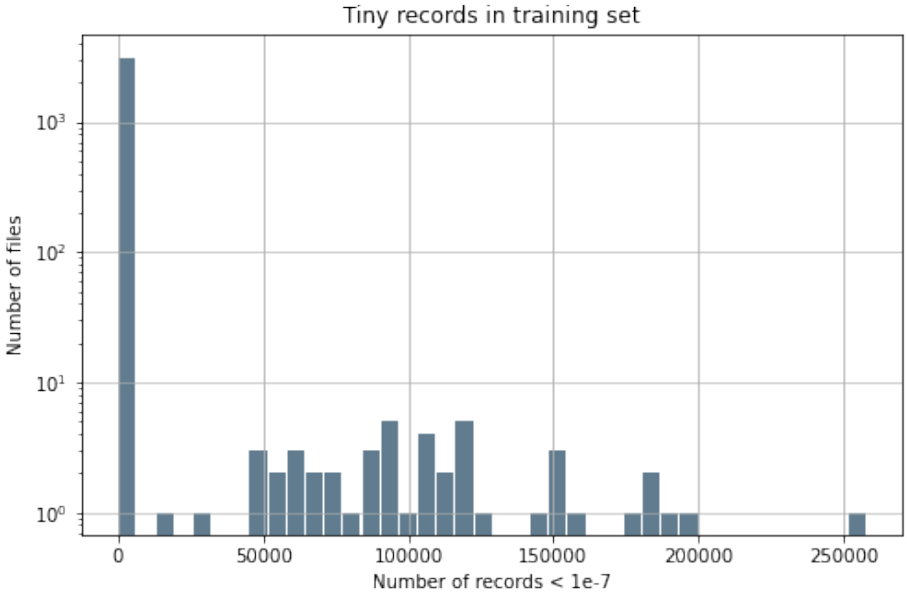


Figure 3.3: Number of recordings with different amounts of tiny data (Training set)

The same process is also applied to the test set. There are 897 sessions in the test set among which 5 sessions are undoable recordings. 892 sessions are selected in this step. Figure 3.4 displays the distribution of different number of tiny data in recordings.

3.2.2 Filter seizures less than 10s

It's conventional that seizures in automatic detection method should be at least 10 seconds long and contain no gaps longer than 3 seconds. However, since the original dataset is labeled manually, some EEG signals less than 10 seconds are also labeled as seizures in this dataset. It results in a conflict with the post-processing steps in Section 4.1, which will filter the seizure signals in less than 10 seconds. Considering both the background knowledge and the following steps, the seizures of less than 10 seconds are

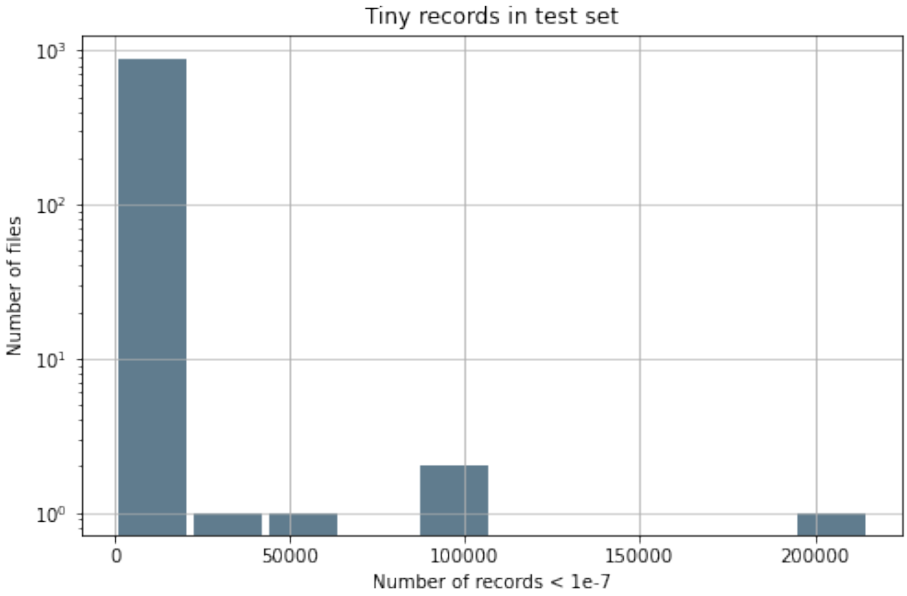


Figure 3.4: Number of recordings with different amounts of tiny data (Test set)

filtered in this step. Since this data is also different from normal data, it will not be used as a normal signal either.

Figure 3.5 shows the length of seizures signals in training set. There are 458 sessions containing seizures in total. It can be seen that the majority of seizures are more than 10 seconds and less than 150 seconds. The longest seizure is 2031.64 seconds. In this step, 3 sessions with short seizures are filtered, and 455 sessions with seizures are reserved. Finally, the training set contains 3125 sessions consisting of 455 seizure sessions and 2670 normal sessions.

The length of seizure signals in the test set is shown in Figure 3.6. There are 241 sessions containing seizures in total. The distribution is similar to the training set but contains more short seizures. And short seizures are always more challenging to be detected. After this step, 240 sessions with seizures are reserved, which means only 1 session is filtered. And finally used test data have 891 sessions consisting of 240 sessions with seizures and 651 normal sessions.

3.3 Feature extraction

There are mainly two strategies to extract features. The first one is using machine learning models such as CNNs to directly extract features from raw data in the time domain or frequency domain. This strategy is powerful and has high accuracy. However, the computational complexity is also very high. The second strategy should be used to make the algorithm more energy efficient, which extracts specific types of features to feed the classifier. Depending on the different domains of EEG signals in which features are extracted, the features can be classified into three categories, time

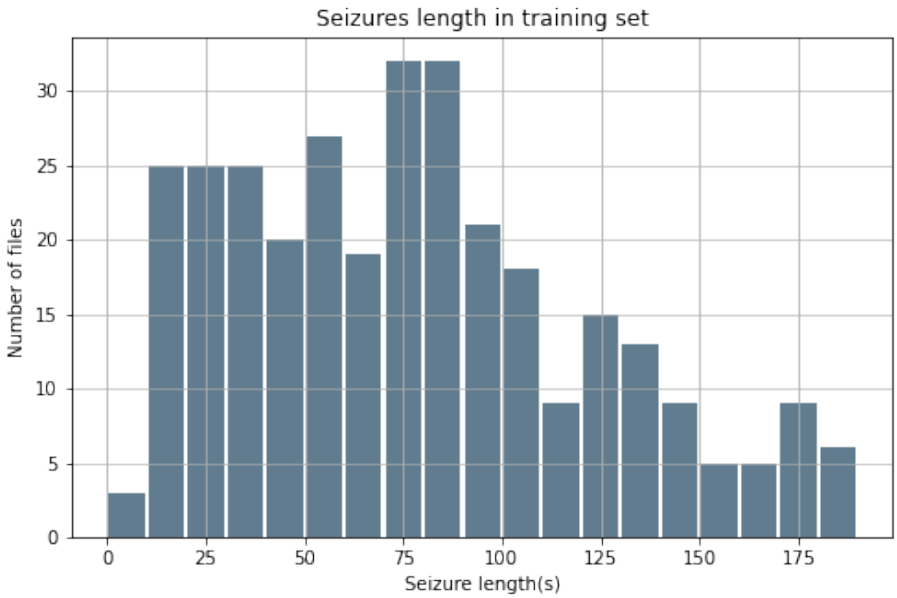


Figure 3.5: Length of recordings in the training set)

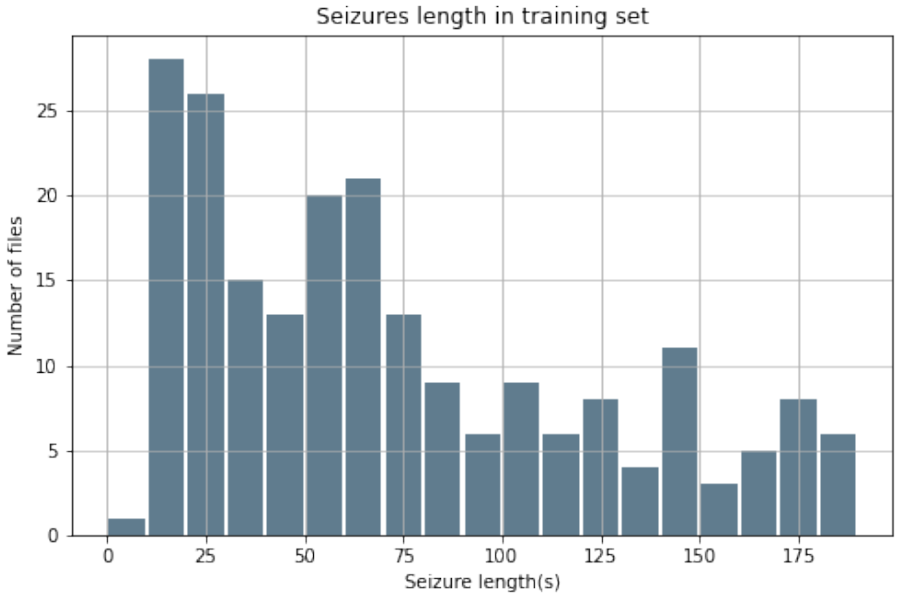


Figure 3.6: Length of recordings in the test set)

domain features, frequency domain features, and time-frequency domain features. In this section, different types of EEG features will be discussed.

3.3.1 Time domain features

Time domain features are extracted from raw EEG data in the time domain. These features are the simplest and most intuitive features. Most of them are statistical features, such as Mean Absolute Value(MAV), maximum, minimum, skewness, kurtosis, and so on. Some other features, like Wave Length (WL) and Zero Crossing (ZC) are also time domain features. The model proposed by this thesis uses only the time domain features to reduce the model's complexity and make it realizable in the time-mode circuit. In this model, 11 types of time domain frequency are acquired in the feature extraction step. There are 18 channels of the EEG signals are used. So 198 features are extracted in total. These features are extracted from the segment that consists of N samples. x_i is the ith sample in each segment. The definition and mathematical equation of these features are as follows.

(i) **Waveform Length (WL)**
Waveform length is the cumulative length of the waveform over the EEG segment[60]. It intuitively indicates the information of amplitude, frequency, and duration. The representation of this feature is shown as Equation 3.1, N is the number of sample points in each segment.

$$wl = \sum_{i=1}^{N} |x_i - x_{i-1}| \tag{3.1}$$

(ii) **Number of zero-crossing (ZC)**
A zero-crossing is the point where the value of a mathematical function changes from positive to negative or from negative to positive. It is an intercept of the function and axis. The number of zero-crossing represents the appearance frequency of zero-crossing in a time segment. If $x_i \cdot x_{i-1} < 0$, where $1 < i < N$, N is the number of sample points in each segment, a zero-crossing within the range (x_{i-1}, x_i) exists.

(iii) **Number of Slope Sign Changes (SSC)**
The number of slope sign changes is defined as the number of times where the sign of slope changes. The slope sign is determined by the difference of adjacent samples for discrete signals. For three consecutive samples x_{i-1}, x_i, x_{i+1}, if $(x_i - x_{i-1}) \cdot (x_{i+1} - x_i) < 0$, a slope sign change happens. The SSC indicates the frequency of slope sign changes in each time segment.

(iv) **Mean Absolute Value (MAV)**
The mean absolute value is the average of the absolute value of the signal. The formula is Equation 3.2.

$$MAV = \frac{1}{N} \sum_{i=1}^{N} |x_i| \tag{3.2}$$

(v) **Root Mean square (RMS)**
The RMS of a signal is defined as the square root of the mean square of the

samples in this signal. Equation 3.3 represents the computation of this value.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} {x_i}^2} \tag{3.3}$$

(vi) **Maximum (MAX)**
MAX is the value of the maximum sample in the signal.

(vii) **Minimum (MIN)**
MIN is the value of the minimum sample in the signal.

(viii) **Standard Deviation (SD)**
The standard deviation represents the amount of variation or dispersion of the signal. It indicates how close the samples in the signal are to the mean. The mathematical expression is as below, where mu is the mean of this segment.

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2}, \mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.4}$$

(ix) **Kurtosis (kurt)**
Kurtosis is the fourth central moment of the standard score. If Fisher's definition is used, then 3.0 is subtracted from the result to give 0.0 for a normal distribution. It indicates the "tailedness" of the probability distribution. The mathematical expression is shown as Equation 3.5, where μ and σ are the mean and standard deviation of this segment.

$$kurt(X) = \mathbb{E}\left[(\frac{\sigma}{X - \mu})^4\right] \tag{3.5}$$

(x) **Skewness (skew)**
The skewness is determined by the asymmetry of the probability distribution of a real-valued random variable about its mean. The value of the right-skewed distribution is positive. The value of the left-skewed distribution is negative. And if the distribution is symmetrical, the skewness will be zero. The skewness of signal X is the third central moment of the standard score of X. The formula is as follows: μ and σ are this segment's mean and standard deviation.

$$skew(X) = \mathbb{E}\left[(\frac{\sigma}{X - \mu})^3\right] \tag{3.6}$$

(xi) **Shannon Entropy (SE)**
The Shannon entropy is a measure of uncertainty of a random process. The higher the Shannon entropy is, the bigger the information is given by a new value in the process[61]. The feature is defined as follows.

$$SE = -\sum_{i=1}^{N} p(x_i) log(p(x_i)) \tag{3.7}$$

3.3.2 Frequency domain features

In order to extract frequency domain features, the raw EEG signal should be transformed into the frequency domain first. There are several methods to do this transformation. Fourier transform is the most common method, which is usually calculated with Fast Fourier Transform (FFT). And it decomposes the signal into sine waves. There are five major brain waves distinguished by their frequency ranges, delta (δ, 0.5-4 Hz), theta (θ, 4-7.5 Hz), alpha (α, 8-13 Hz), beta (β, 14-26 Hz) and gamma (γ, 30-45 Hz)[62]. Each band contains different information. With the analysis of the power spectral density (PSD) in each frequency band, the statistical features, like MAV, SD, skewness, and kurtosis, used in the time domain can also be applied to extract features from the PSD of each frequency band. Unlike FFT, Wavelet Transform (WT) decomposes the signal into wavelets, which are a kind of rapidly decaying, wave-like oscillations. Different wavelets can be used for different applications. In this case, wavelet coefficients are the most important features. Besides these non-parametric methods, some parametric models like the aggressive (AR) models, multivariate autoregressive models, and autoregressive-moving average (ARMA) models can also be used to do the transformation. With these models, frequency domain features like reflection coefficients and partial correlation coefficients can be extracted[63].

3.3.3 Time-frequency domain features

By analyzing the signal in the time and frequency domain simultaneously, the relationship between frequency and time can be observed intuitively. Short-time Fourier transform (STFT) is the simplest method to observe this relationship. The frequency information within each window is extracted along the time axis. A spectrogram will be obtained as the output of this method.

3.3.4 Time-mode circuit

Time-mode circuit is the circuit that uses the difference between the time instants at which two digital events take place to represent an analog signal, which is shown as Figure 3.7. Both the input and output signals are in the temporal domain and the input difference is linearly proportional to the amplitude of the analog signal. The time-mode circuit is able to deal with addition, multiplication, amplification, integration, and quantization of time variables. So it's possible to use the time-mode circuit to make fundamental computations[6]. The time-mode circuit is a digital system, which performs mixed analog-digital signal processing. Compared with other computation modes like voltage-mode or current-mode, this mode is simpler and more energy efficient. Besides, digital transistors have a better ability to process timing signals. So it's also easier for hardware implementation.

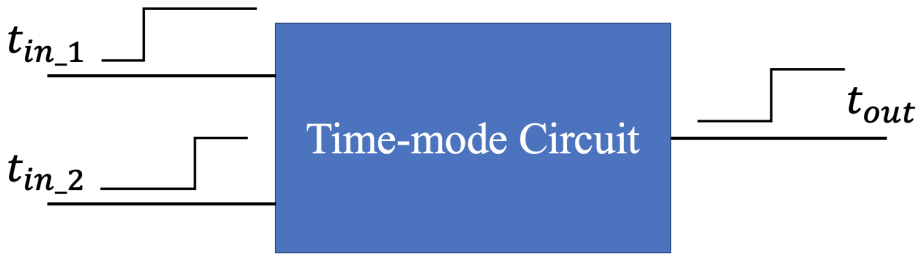


Figure 3.7: Time-mode computation block

Ravinuthula has proposed several time-mode circuits for analog computation [64]. Linear computations are easier to be realized with the time-mode circuit. He designed circuits for linear computations, including weighted averaging, weighted subtraction, weighted sum, scalar multiplication, and maximum/minimum. He also talked about two ways to implement non-linear arithmetic. The multiplication and division were implemented in time-mode circuits by introducing non-linearity in the existing linear computational blocks. And the second method, using time-mode multi-layer perceptrons (MLPs), can realize more complex non-linear arithmetic. In [65], the translinear principle was used to implement root and power computations with the time-mode circuit. And based on approximate computing, some non-linear computations like logarithm and exponential can be calculated by linear computations[66]. Consequently, they are able to be implemented with time-mode linear computation locks. Also, the logarithm computation is possible to be implemented in the time-domain directly[67].

The definitions and formulas of the selected features have been illustrated in Section 3.3.1. It can be seen that Waveform Length (WL) is the sum of the differences between the sampling points, which can be implemented by a sum circuit and subtraction circuit. The number of zero-crossing (ZC) is the number of points where the value of functions changes from positive to negative or from negative to positive. So it can be implemented by a simple subtractor, comparator and counter. By introducing non-linearity in a scalar multiplication circuit, the time-mode multiplication and division circuit can be realized[64]. The number of Slope Sign Changes (SSC) is a minor modification of ZC circuit. Based on ZC circuit, it calculates the multiplication of two consecutive sample differences first, and a time-mode multiplication circuit will be used. The time-mode multiplication and division circuits are also used in the models of root mean square(RMS), Standard Deviation (SD), Kurtosis (kurt), Skewness(skew) and Shannon Entropy (SE). The mean absolute value can be realized with the use of a weighted subtraction and scalar multiplication circuit. Maximum and minimum can be easily implemented by Maximum(MAX)/Minimum(MIN) Circuit. For the logarithm computation of Shannon entropy, it can be realized by linear approximation or implemented in the time-domain directly as it was discussed in the last paragraph. Consequently, most of the time-domain features in subsection 3.3.1 can be calculated with these fundamental computations.

4 Post-processing

4.1 Process

4.1.1 Filter and merge seizures

Different types of seizures have different symptoms. Most of the seizures have a length in the range of 30 seconds to 2 minutes. The seizures longer than 200 seconds are long seizures. A seizure longer than 5 minutes will be considered as a medical emergency. As a consensus, the abnormal signals of less than 10 seconds will not be labeled as seizures. In other words, the discrete short seizure signals will be ignored [68]. So the predicted seizures less than 10 seconds will be filtered. Besides, the adjacent seizures with an interval of fewer than 60 seconds will be merged together and considered as one seizure signal.

4.2 Evaluation strategy

4.2.1 Evaluation metrics

Evaluation metrics are powerful tools to assess the performance of machine learning models. There are tens of metrics that can be used for binary classification problems. Ferri et al. have classified these metrics into three families, 1) threshold metrics, which are used to quantify and understand the error, 2) probability metrics, which are used to quantify the uncertainty of predictions, 3) ranking metrics, which are used to evaluate the effectiveness of separating classes [69].

The main objective of using threshold metrics is to minimize the model's error, which is also the main objective of the model itself. So threshold metrics are the most generally used evaluation metrics. And they are usually used in direct applications of classifiers. This experiment uses these metrics after the post-processing step to evaluate the model's performance based on its final result. The metrics that have been used are as follows.

(i) **Confusion matrix**
The confusion matrix is a common method for presenting true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions. True positive is the outcome where both the prediction and the label are positive. False positive is the outcome where the prediction is positive, but the label is negative. True negative is the outcome where both the prediction and the label are negative. And false positive is the outcome where the prediction is negative, but the label is positive. These four values are presented in the form of a matrix. The x-axis shows predicted values, and the y-axis shows actual values.

(ii) **Accuracy**
Accuracy shows the frequency of correct predictions among all predictions. However, it's not a good measure of performance for imbalanced datasets. Since even for inseparable data, any model which makes dummy predictions that all samples belong to the major class will still have high accuracy equal to the percentage of the majority. The accuracy will be high in our project. So it is not enough to use this method only.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

(iii) **Sensitivity (Recall)**
Sensitivity, also known as recall, measures the ratio of positive predictions in all positive samples.

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.2}$$

(iv) **Precision**
Precision is the ratio of true positive predictions to all positives.

$$Precision = \frac{TP}{TP + FP} \tag{4.3}$$

(v) **F1 score**
Both sensitivity and precision are class-level metrics. There is a trade-off between these metrics. Increasing the sensitivity will pull down the precision and vice versa. The decision should be made based on which metric is prior to the problem. And to make a balance between them, F1 score is proposed. F1 score is the harmonic mean between precision and sensitivity. By optimizing the F1 score, a model with good performance of both sensitivity and precision is acquired.

$$F1 = \frac{2 * Sensitivity * Precision}{Sensitivity + Precision} \tag{4.4}$$

(vi) **FAs/24hrs**
Besides the commonly used metrics above, FAs/24 hrs is also an important metric in the healthcare field. It is similar to the false alarm (FA) rate, which indicates the number of false alarms per 24 hours.

The objective of probability metrics is to evaluate the uncertainty rather than the correctness of the predictions. They can be used to penalize the false predictions with high confidence. The probability metric used in this model is binary cross-entropy. It has been used during the training and validation process to perform as a loss function. The mathematical expression of this metric is shown as Equation 4.5, where y is the label, $p(y)$ is the probability of predictions and N is the total number of samples to be evaluated.

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\left(p\left(y_i\right)\right) + \left(1 - y_i\right) \cdot \log\left(1 - p\left(y_i\right)\right) \tag{4.5}$$

The ranking metric will score the classifier on its effectiveness in separating classes. Based on this score, different thresholds for the machine learning model's output can be selected to do the classification. Receiver Operating Characteristic (ROC) is the most popular ranking metric. ROC curve shows the relationship between true positive rate (TPR) and false positive rate (FPR) in Figure 4.1. Each point on the curve represents a threshold. The x label is FPR, and the y label is TPR. A diagonal line from the bottom left to the top right represents a classifier without skill. The area under the curve (AUC) is calculated to show the classifier's ability to make predictions. The closer the ROC curve is to the top left, the higher AUC is, and the better the classifier is. An ideal classifier has an AUC equal to 1.

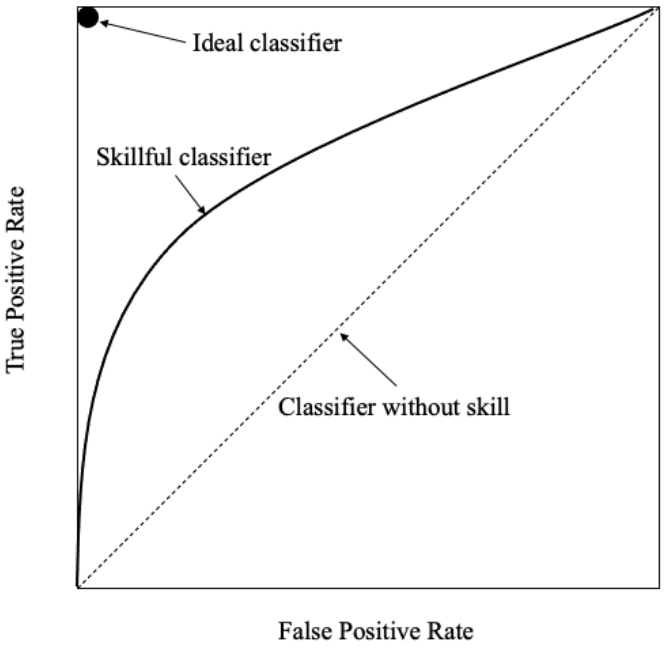


Figure 4.1: ROC curve

4.2.2 Scoring metrics

The evaluation metrics discussed above have provided various methods to evaluate the performance of the classifier. However, a proper scoring metric should also be chosen to get the final result. The scoring metrics can be grouped into two categories: epoch-based and term-based. Two examples of predictions are shown in Figure 4.2. The reference is the sequence of the signal's labels. The hypothesis is the sequence of the signal's predictions.

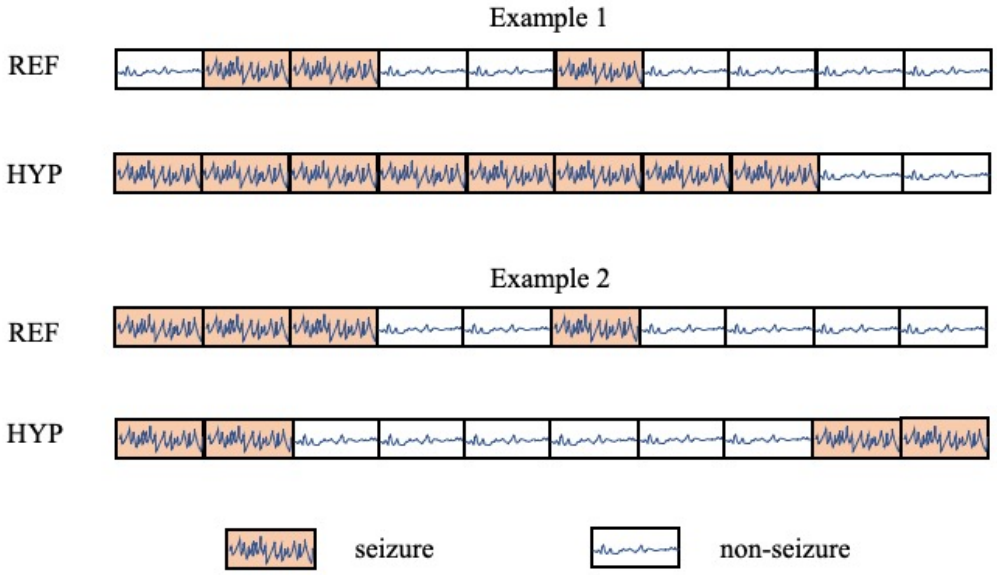


Figure 4.2: Two examples of reference and hypothesis

With the epoch-based scoring method, the signal is regarded as continuous units of epoch over time. It directly calculates the evaluation metrics based on the correctness of each unit. In this case, there are 3 TPs, 2 TNs, 5 FPs and 0 FN in example 1. The accuracy is 0.5. The sensitivity is 1. And the precision is 0.375. For example two, there are 2 TPs, 4 TNs, 2 FPs and 2 FNs. So the accuracy is 0.6. The sensitivity is 0.5. The precision is 0.33. Since this scoring method is rigorous, the value of each metric is small.

Term-based scoring is a more permissive method. It does not assess every epoch in the signal. Each successive epoch with the same label is aggregated as a term first. And then, the assessment is based on each term rather than each epoch. A successfully detected seizure will be counted as one TP only in this case. This method focuses more on the number of seizures detected rather than the length of TP duration. Any-Overlap Method (OVLP) is a typical term-based scoring approach [70]. TPs are counted if any overlaps between reference and hypothesis exist. FPs are positive hypotheses that have no overlap with any references. And a FN is counted when a positive reference is fully covered by a negative hypothesis. As a result, there are 1 TP, 1 TN, 0 FP and 0 FN in example 1. Its accuracy, sensitivity, and precision are 1, 1, and 1, respectively. And for example 2, there are 1 TP, 0 TN, 1 FP, and 1 FN. So the accuracy, sensitivity, and precision are 0.33, 0.5 and 0.5, respectively. This method tends to have a higher sensitivity. OVLP is widely applied in neuroengineering community, which is also the scoring metric used in this project.

4.2.3 Criteria in seizure detection field

The EEG records are labeled by experts manually. So it is inevitable that some individual biases exist in every dataset, which means labels of the data are not ideally correct. In order to minimize the incorrect evaluation caused by it, it is more proper to use relative performance rather than absolute performance. So the evaluation seizure detection methods should follow the manners below [71].

(i) The seizures within 60 seconds of each other are grouped together so that the continuous positive predictions will not lead to over-represent.

(ii) Due to the individual bias, the onset and offset of seizure is ambiguous, so the positive detection occurs within 60 seconds of the onset, and the offset of a marked seizure is also reported as a true positive detection.

(iii) An undetected seizure is considered a false negative detection.

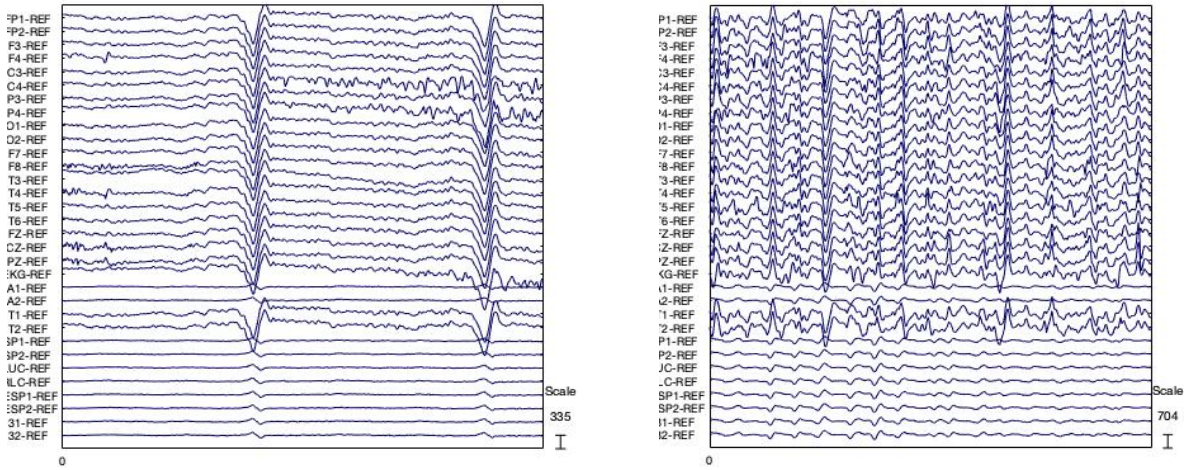
(iv) The positive detection before or after 3 minutes of the marked seizure is reported as false positive detection.

(v) Only one false positive detection will be counted for continuous 60 seconds.

5 LSTM model

5.1 Feature set selection

EEG features can be selected from the time domain, frequency domain, and time-frequency domain. To extract features from the time domain, the sequence signal need to be segmented into epochs with the same length first. And the epoch length has a great influence on the effectiveness of the extraction. This length should be short enough to make sure the signal within the epoch is stationary. The epoch of 1s is the most popular choice for time segmentation, which has been widely used in many related researches [72, 36, 73]. And in order to form a time series as the input data for the ML model, the length of segments should also be discussed. The short length means short series, which may result in insufficient information. The long length means long series, in which case some short seizures are fully contained within the time series and the data will be labeled as seizure even if most part of this signal is normal, so there is a trade-off.



(a) 1s epoch of normal EEG

(b) 1s epoch of seizure EEG

Figure 5.1: 1s epoch of EEG signal

Many ML seizure detection algorithm apply Short-time Fourier transform(STFT) for raw EEG data first and then extract features from the time-frequency data[74]. STFT has the ability to show the sinusoidal frequency and phase content of a short

frame over time. The objective of this LSTM-based algorithm is to imitate its principle, which extracts features from short time frames first and then takes the feature series over time as the input of ML models. A simple hidden layer model only uses the input of each time point to get the output. However, the recurrent neural networks have an inside loop that can obtain information about all past history to compute the output. So, it can be concluded that in the LSTM model, both the time domain feature extraction step and the LSTM layer work as a feature extractor. The time domain feature extraction step is performed in pre-processing step to extract features from raw EEG data. In this step, the EEG signal should be segmented into epochs first and then the epochs will form a new series with specific length. This process in shown in Figure 5.2. The sampling frequency is f_s. The length of epoch, overlap and segment are l_{ep}, l_{ol} and l_{sm}. So, there are $l_{ep}f_s$, $l_{ol}f_s$ and $l_{sm}f_s$ samples in each epoch, overlap and segment respectively. And the LSTM layer performs as a machine learning-based feature extractor to obtain information from feature series. In order to find a proper combination of the time length in two steps, four feature sets were acquired and worked as the input of several LSTM models. By comparing their performance, the optimal feature set was selected in Section 5.2.4. The structure of the feature set is shown in Table 5.1.

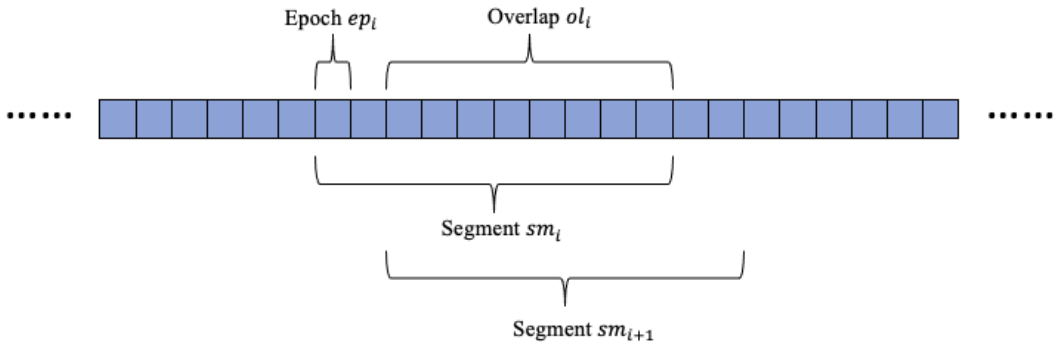


Figure 5.2: Segmentation of EEG signal to extract features

Table 5.1: Structure of feature set

Feature set No.	Epoch (s)	Segment length (s)	Overlap (s)	Series length (epochs)
1	0.5	5	4	10
2	1	5	4	5
3	0.5	10	9	20
4	1	10	9	10

5.2 Model structure selection

In this section, three different LSTM-based seizure detection models are proposed and discussed. These different models are combined with four different feature sets to make

a comparison. The result shown in Section 5.2.4 indicates the optimal choice of the feature set and model structure.

LSTM layer is able to remember the important information from newly input data and forget the unimportant information from a previous memory. So it's a sequence processing model that processes the data in a forward direction, which means the previous samples are used to predict the label of the next sample's value. In this case, it's possible to get a seizure prediction even though most samples in the sequence are normal and vice versa. For example, there are two 10s signals, one has 8 seconds seizure signal at the beginning, and the other has 8 seconds seizure single in the end. They have a different probability of being predicted as a seizure. If a simple LSTM model is used to predict the label of the last second in the sequence in the first signal, it is probably to be predicted as a seizure because of the high percentage of seizures in the signal. But actually, the label of the last second is normal. And for the same reason, the last second of the second signal is more likely to be predicted as normal. This result cannot represent the sequence data properly. So, in this thesis, the label of the input series should be extracted from the middle time. In other words, the purpose of the model is to use time series with a length of t seconds to predict the label of $\frac{t}{2}s$, which is illustrated in Figure 5.3. In this case, the simple one-way propagation LSTM structure cannot meet the requirement.

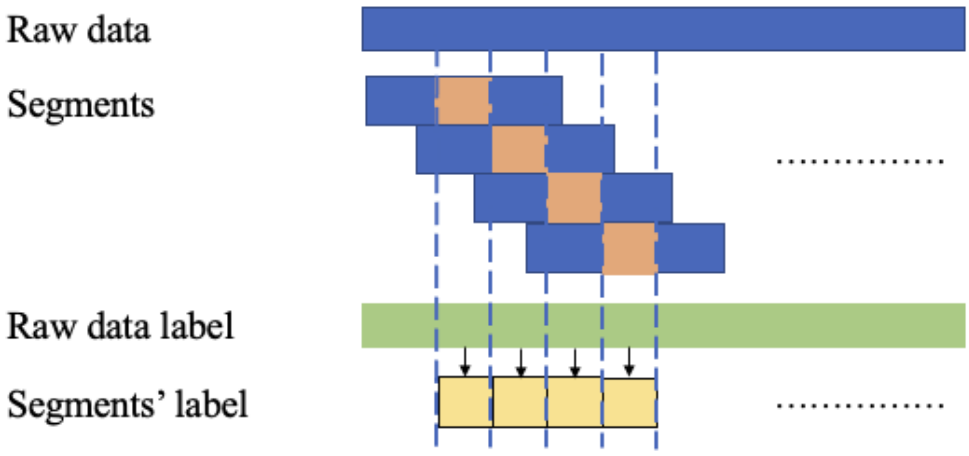


Figure 5.3: Label extraction process

Bidirectional Long Short-term memory (BiLSTM)is an advanced LSTM structure. It consists of two LSTMs: one is in the forwarding direction, and the other is in the backward direction. The structure of BiLSTM is as Figure 5.4. The LSTM layers used in this thesis are all BiLSTM layers. The features sequence is a $N \times 198$ 2D input vector of the model, where N is the number of time samples in each sequence. And in order to compare the model structures more intuitively, the input vector size in Section 5.2.3, 5.2.2 and 5.2.1 are fixed as (20, 198). The number of time samples N in each sequence is equal to 20.

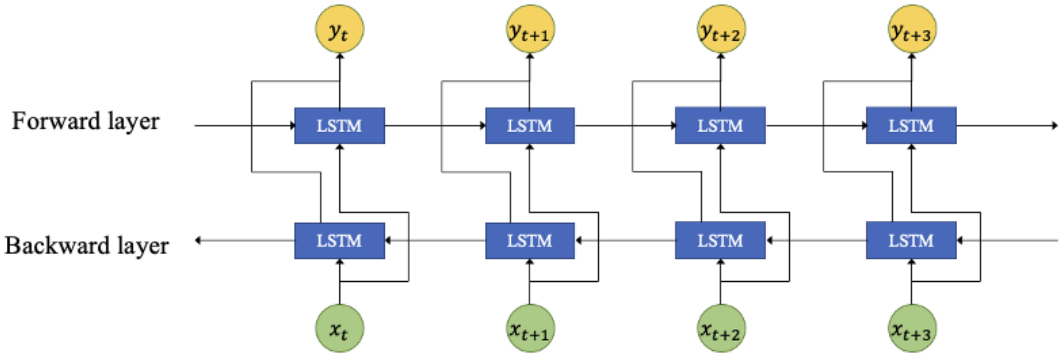


Figure 5.4: BiLSTM structure

5.2.1 DNN-BiLSTM model

The first proposed LSTM model is the DNN-BiLSTM model. It uses two dense layers to compress the size of the features vector first. The 198 features of each timepoint are reduced according to the number of neurons in DNN layers. Then a bidirectional LSTM layer is added to extract a 1D feature vector from the 2D input feature series. And last, a DNN layer works as an output layer with the sigmoid activation function to give a binary classification of the signal. The architecture of the DNN-BiLSTM model is illustrated in Figure 5.5. In order to make a comparison among different models, the number of variable parameters in each model is close. For this DNN-BiLSTM model, there are 58625 trainable parameters and no non-trainable parameters, 58625 parameters in total when the input size is (20, 198).

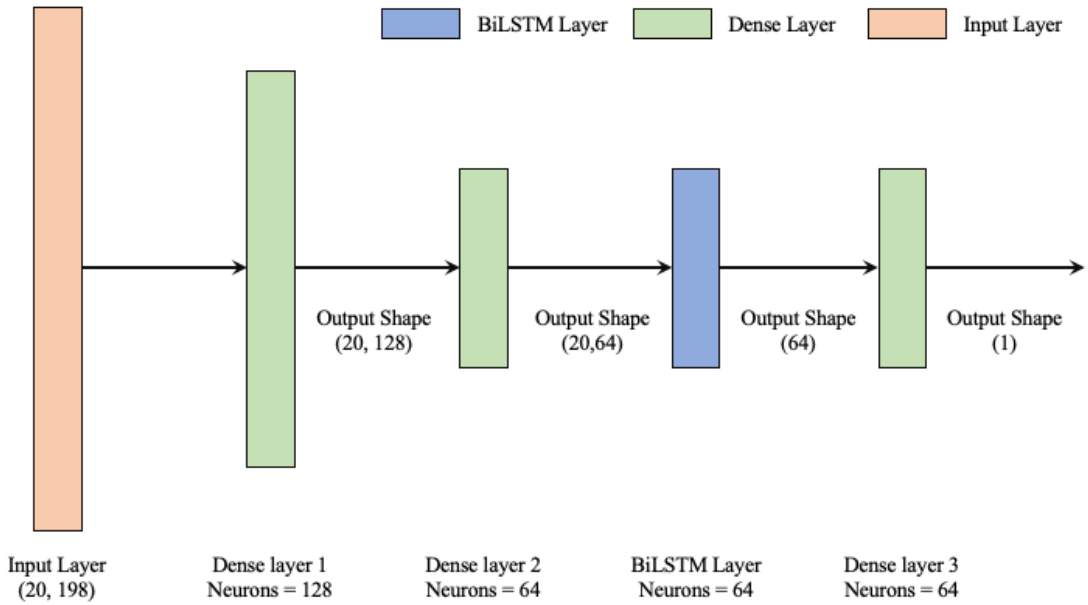


Figure 5.5: DNN-BiLSTM model architecture

5.2.2 Multi-layer BiLSTM model

The second proposed LSTM model is the Multi-layer BiLSTM model. In this model, BiLSTM layers are the only type of layer that performs as the ML feature extractors. Two BiLSTM layers are used successively, which the first layer compresses the vector in the feature direction and the second layer compresses the vector in both feature and time direction. The output of the second layer is several independent features without relationships over time. And same as the first LSTM model, the output layer in this model is also a DNN layer with a sigmoid activation function. This architecture is an intuitive application of the BiLSTM layer. The classification ability is related to the complexity of the model, which means the depth and the number of neurons in each BiLSTM layer. Figure 5.6 illustrates the architecture of this Multi-layer BiLSTM model. There are 69537 trainable parameters and no non-trainable parameters, 69537 parameters in total when the input size is (20, 198).

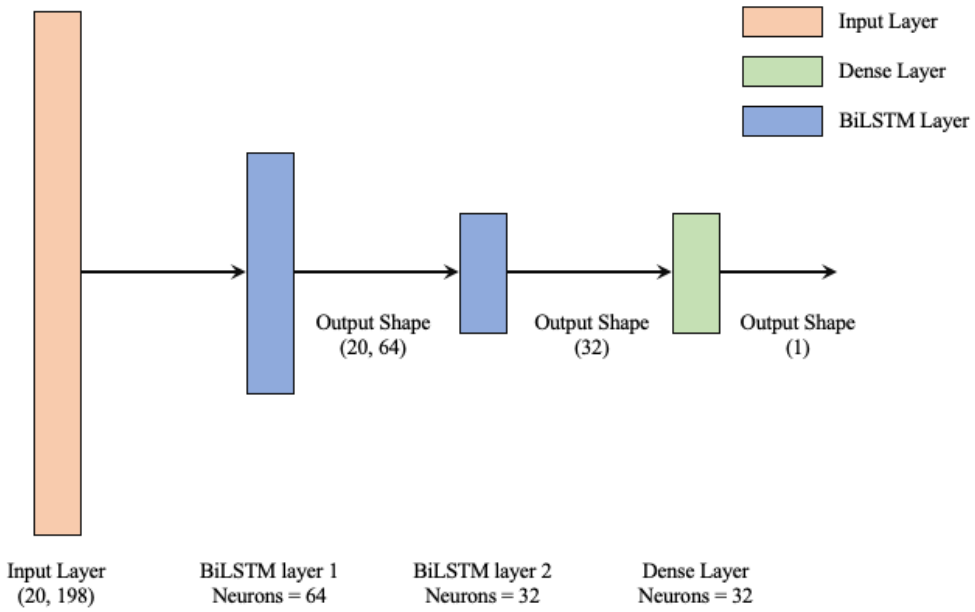


Figure 5.6: Multi-layer BiLSTM model architecture

5.2.3 BiLSTM-DNN model

The third proposed LSTM model is the BiLSTM-DNN model. It's different from the second model which also has BiLSTM layers and DNN layers successively. The DNN layers in this model work not only as the output layer but also as the feature extractor. The first part of this model is similar to the multi-layer BiLSTM model, both of them have two BiLSTM layers to extract features from input data. But the difference is that the BiLSTM layers of this model only compress the features in the feature direction. The output of this part is still a 2D vector. Next, the 2D vector is flattened by a flatten layer. And then, the input of the following dense layers is a 1D vector. This structure has more layers, and the size of the input vector is compressed more slowly. However,

the disadvantage of this model is that the time information among different time points is lost in the flattening step. The architecture is displayed in Figure 5.7. This model has 79777 trainable parameters and no non-trainable parameters, 79777 parameters in total when the input size is (20, 198).

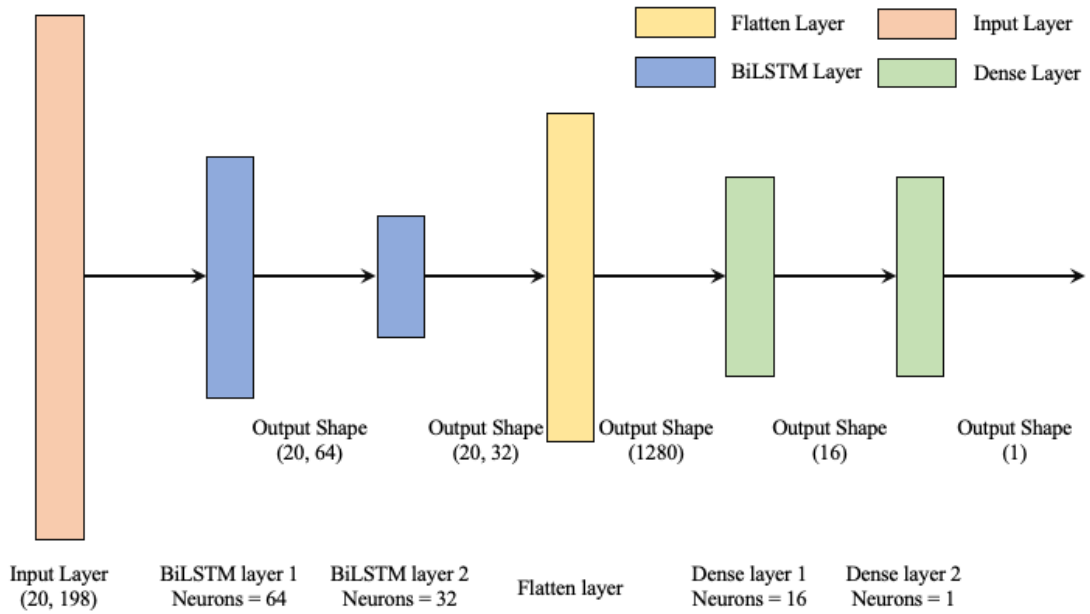


Figure 5.7: BiLSTM-DNN model architecture

5.2.4 Result

An epoch in this section means one complete pass of the training dataset through the algorithm. Every parameter in each epoch can update the internal model. There are 12 combinations of 4 feature sets and 3 models. The models were trained for 50 epochs. The training data were randomly split into training and validation sets randomly. 20% is the validation set, and the rest is preserved as the training set. the performance of each model is compared in this section. The DNN-BiLSTM model doesn't converge with all 4 feature sets, which can be proven not to be a doable architecture for this problem. And the result of the rest of 8 combinations is shown in Table 5.2.

As seen from the table, both models with the input from the feature set 3 have good performance compared with the other three sets. It means feature set 3 has a great ability to extract more meaningful features. Besides, the Multi-layer BiLSTM model performs better with the feature set 1 and set 3. And with the data from the feature set 2 and 4, the BiLSTM-DNN model has a better performance. The combination of feature set 3 and the Multi-layer BiLSTM model has the best performance, so the Multi-layer BiLSTM model should be regarded as the optimal model. Taken together, these results suggest that the combination of feature set 3 and the Multi-layer BiLSTM model is the optimal choice for this problem.

Table 5.2: Comparison of performance about different feature sets and models

	Set No.	Params	Training loss	Training acc	Test loss	Test acc
Multi-layer BiLSTM	set1	69537	0.1861	0.9299	0.4803	0.8252
	set2	69537	0.3493	0.8446	0.5046	0.7793
	set3	69537	0.1529	0.9403	0.4913	0.8420
	set4	69537	0.2744	0.8935	0.4937	0.8004
BiLSTM-DNN	set1	74657	0.3138	0.8717	0.6045	0.7525
	set2	72097	0.2782	0.8892	0.4736	0.8041
	set3	79777	0.2204	0.9122	0.4570	0.8306
	set4	74657	0.1978	0.9245	0.4515	0.8265

5.3 Decision threshold

The sigmoid function is used as the activation function in these models. It guarantees that the output of the output layer is always within the range from 0 to 1. The smaller the input is, the closer the output will be to 0, and vice versa. There are two mutually exclusive classes in this problem, which is known as a binary classification problem. The decision threshold of the output should be selected. The output value greater than the threshold will be converted to class label 1, and the output value less than the threshold will be converted to class label 0. The default value of the threshold is equal to 0.5. However, the class distribution of this problem is severely skewed, and the evaluation metrics are different from the metrics used in the training process. Thus, the default threshold hold cannot interpret the probabilities effectively. This section will discuss three different methods to obtain the optimal threshold of the predicted probabilities. Section 5.3.1 introduces the strategy of selecting a threshold based on ROC. Section 5.3.2 talks about the threshold for the F1 curve with and without the post-processing step.

5.3.1 ROC of raw predictions

As explained in Section 4.2.1, the ROC curve is actually a combination of the results of multiple confusion matrices. Every point on the curve represents the decision threshold. Its x coordinate is the true positive rate with this threshold, and its y coordinate is the false positive rate with this threshold. Consequently, the closer the point to the top left is, the bigger the difference between TPR and FPR is, and the more accurate the predictions are. So the optimal threshold appears to be closest to the top left of the figure and has the largest value of $TPR - FPR$. In this experiment, this method was applied to raw predictions. It is the most intuitive and simplest method for threshold selection. And the result of it will be explained in Section 5.3.3. The ROC curve of the model is displayed as Figure 5.8. The black point on the curve represents the optimal decision threshold.

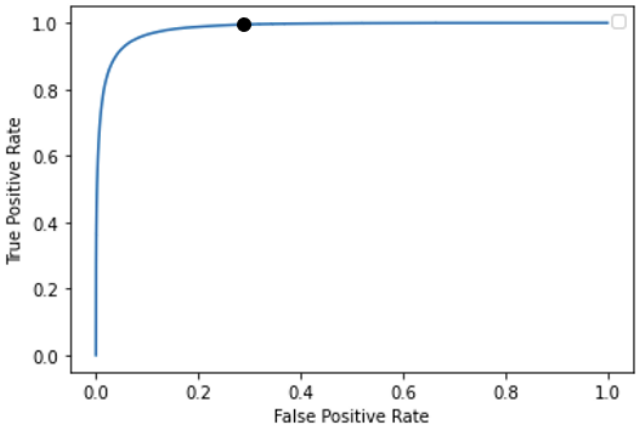


Figure 5.8: ROC to acquire optimal threshold

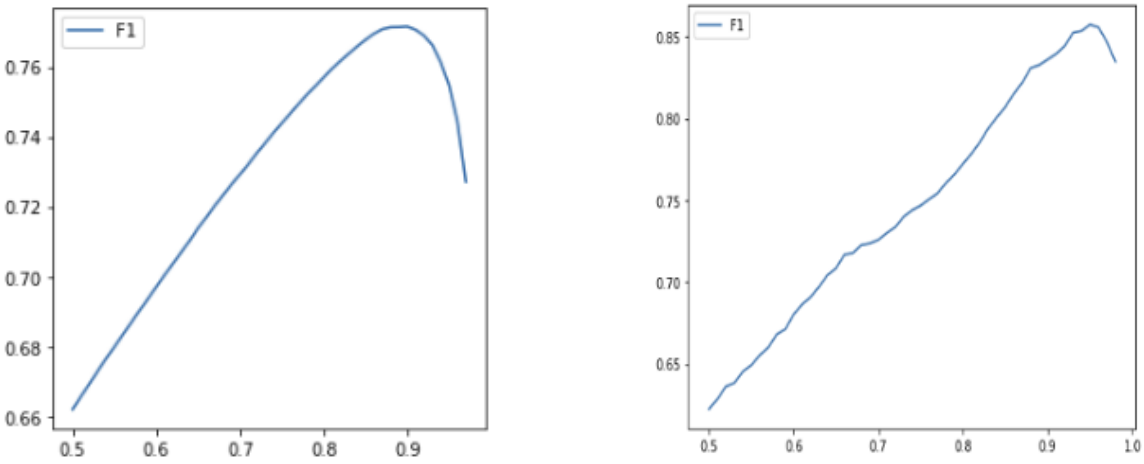
5.3.2 F1 curve of raw predictions and predictions with post-processing

As was mentioned in Section 4.2.1, the F1 score is an improved evaluation metric that combines both precision and sensitivity evaluation metrics. It works well on imbalanced data since it focuses on the performance of the minority class only. The function is similar to a precision-recall curve, created by using crisp class labels for predicted probabilities across a set of thresholds. The x and y coordinates of the precision-recall curve are recall (sensitivity) and precision, respectively. The closer the point is to the top-right, the better the performance is. The naive approach to finding the optimal threshold is to calculate the F1 score for each threshold. So, the F1 curve is used as a straightforward interpretation of the precision-recall curve to find the optimal threshold.

The F1 curves are plotted with raw predictions and predictions with post-processing. Figure 5.9 presents the F1 curves. And the result is discussed in the next section.

5.3.3 Result

As was pointed out in Section 5.2.4, the feature set used is set 3 and the model is a multi-layer BiLSTM model. The performance discussed in this section is the performance of the result after post-processing. In other words, it is an event-based evaluation. It is also the final result we want to get. Table 5.3 compares the performance obtained with different decision threshold selection methods. It is apparent from this table that the best performance is obtained by using the threshold selected by the F1 curve of predictions with the post-processing step. It has the best precision, F1 score, and FAs/24hrs. So the decision threshold of the models will be acquired with this method.



(a) raw predictions

(b) predictions with post-processing

Figure 5.9: F1 curves for raw predictions and predictions with post-processing

Table 5.3: Comparison of different decision threshold selection methods

Method	Acuuracy	Precision	Sensitivity	F1 score	FAs/24hrs
ROC	0.7872	0.4272	0.8617	0.5712	79.03
F1 raw	0.8434	0.6219	0.6868	0.6527	25.24
F1 post-processing	0.8504	0.6875	0.6286	0.6567	16.60

6 Model optimization

6.1 Channel selection

6.1.1 Channel selection techniques

EEG data used in machine learning and medical areas are always multi-channel signals. The EEG dataset of this experiment is the 18-channel signal. Signals of different channels are extracted from different sites of the cerebral cortex. And each channel has different contributions for different applications, which means the importance of each channel is different. So the major purpose of channel selection is to reduce the model's complexity and computational cost. Besides, the signals of specific channels have consistency and synchrony over time, which means that these channels contain the same information. Using multiple signals that contain the same information may cause the model to overfit the signals from these channels while not fully utilizing other types of information. Thus, channel selection can also improve model performance by reducing the use of redundant channels [75].

Channel selection is a special kind of feature selection technique. The objective of feature selection is also to select a subset from the dataset. The supervised feature selection can be classified into three main categories wrapper method, filter method, and embedded method. The wrapper method uses a specific machine learning algorithm to find the optimal features, for example, Recursive Feature Elimination (RFE) and forward selection are two wrapper methods widely used in feature selection. The filter method uses independent evaluation criteria such as information measure, dependency measure, and distance measure to evaluate the contribution of each channel. Its accuracy is relatively low since it does not take the combinations of different channels into consideration. However, the computational complexity of this technique is low and it's independent of different classifiers. Because of its convenience and high speed, in this section, it is used to rank the channels. The embedded method embeds features during the model-building process. It evaluates the channels' importance on their usefulness by cross-validation. It is similar to the wrapper method, but the computational complexity is lower.

The filter method applied for seizure detection can be achieved using signal statistics. In this experiment, three statistical values, variance, the difference in variance and entropy, are used[75]. Var_1 is the variance of seizure signals in the channel, which is referred to as the variance in this experiment. Var_0 is the variance of non-seizure (background) signals in the channel. And the difference in variances is the difference between Var_1 and Var_0. Last, the entropy refers to the entropy of an EEG channel that measures the uncertainty of the signal. All these three statistics are normalized first to combine these values. The average of these normalized values is used to evaluate the importance of each channel. The N input channels are selected based on the maximum

average of these statistics. The table below presents the statistics of each channel, and the result is also plotted in Figure 6.1.

Table 6.1: The statistics of each channel

Channel No.	Channel name	Var_1	Var_{diff}	$Entropy$	Var_{1norm}	$Var_{diffnorm}$	$Entropy_{norm}$
0	FP1-F7	1.1523	0.5716	19.2057	0.4975	0.2422	0.1026
1	F7-T3	1.1313	0.5784	19.2419	0.1897	0.3073	0.6522
2	T3-T5	1.1447	0.6010	19.2484	0.3864	0.5219	0.7522
3	T5-O1	1.1501	0.6165	19.2561	0.4649	0.6706	0.8696
4	FP2-F8	1.1279	0.5508	19.2098	0.1397	0.0437	0.1646
5	F8-T4	1,1455	0.6018	19.2466	0.3976	0.5299	0.7244
6	T4-T6	1.1787	0.6400	19.2484	0.8846	0.8942	0.7520
7	T6-O2	1.1866	0.6511	19.2545	1.0000	1.0000	0.8444
8	T3-C3	1.1654	0.6246	19.2534	0.6887	0.7474	0.8418
9	C3-CZ	1.1184	0.5909	19.2588	0.0000	0.4256	0.9103
10	CZ-C4	1.1516	0.6288	19.2647	0.4867	0.7875	1.0000
11	C4-T4	1.1446	0.6125	19.2571	0.3845	0.6320	0.8840
12	FP1-F3	1.1548	0.5706	19.1990	0.5344	0.2323	0.0000
13	F3-C3	1.1405	0.5997	19.2498	0.3248	0.5100	0.7734
14	C3-P3	1.1233	0.5882	19.2530	0.0716	0.4000	0.8224
15	P3-O1	1.1473	0.6203	19.2625	0.4239	0.7062	0.9667
16	FP2-F4	1.1261	0.5462	19.2052	0.1137	0.0000	0.0945
17	F4-C4	1.1187	0.5841	19.2589	0.0047	0.3616	0.9121

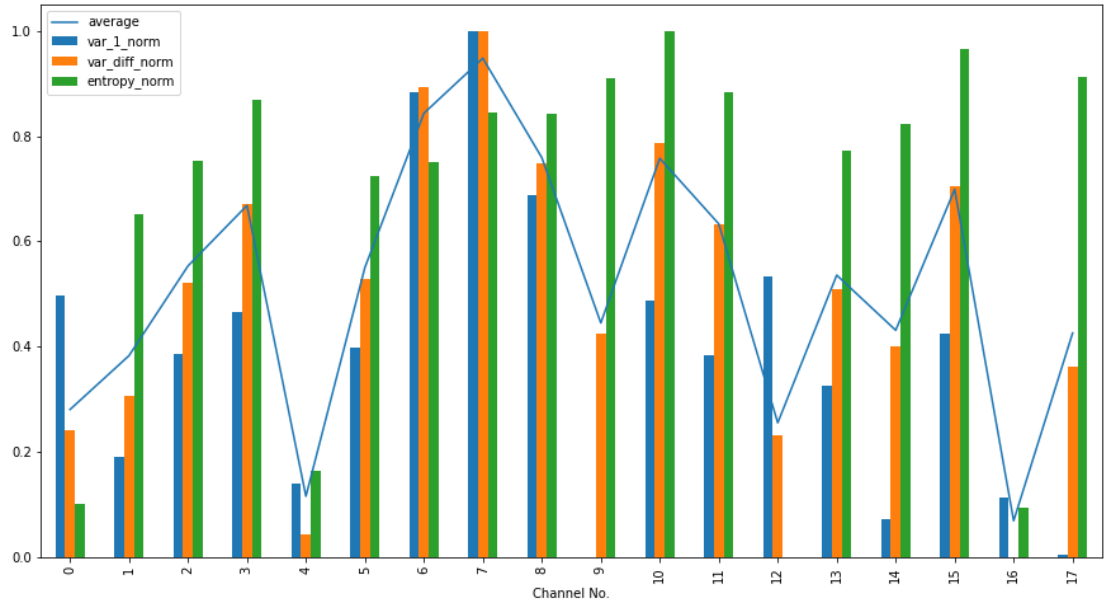


Figure 6.1: Channel selection based on statistics

6.1.2 Subset channel selection performance evaluation

After sorting the channels with the method mentioned in the previous section, the top 15, 12, 9, 6, and 3 channels are taken out in turn to form new subsets for model training. These subsets and the original set are used to train the model for 20 epochs. Post-processing steps are also applied for predictions of each model. And the performance of the final results is shown in Table 6.2.

Table 6.2: Performance of different subsets

Number of channels	Params	Acuuracy	Precision	Sensitivity	F1 score	FAs/24hrs
18	69537	0.8406	0.6667	0.6029	0.6331	17.27
15	66721	0.8289	0.6382	0.5599	0.5965	17.56
12	61089	0.8505	0.7689	0.5495	0.6410	9.13
9	52641	0.8463	0.7245	0.5680	0.6368	12.12
6	44193	0.8587	0.7621	0.5959	0.6688	10.63
3	35745	0.8137	0.6000	0.5225	0.5586	19.259

It's apparent from the table that the model performance is improved when a small number of channels are removed, but with the number increasing, the performance starts to decline. As mentioned in the previous chapter, the dataset is imbalanced greatly, so the accuracy is always high, which means it is not the metric that is of the utmost importance. In the seizure detection model, precision shows the ability to give precise alarms and sensitivity indicates the ability to detect all seizures. Both metrics are important, so F1 scores are used to evaluate the performance. It can be inferred from the results that the subsets with 6-12 channels perform best. And considering the model complexity, the subset consisting of 6 channels is chosen as the optimized dataset. The channels in this dataset are T6-O2, T4-T6, T3-C3, CZ-C4, P3-O1, and T5-O1. The montage of these six selected channels are illustrated in Figure 6.2. It can be seen that are all on the back of the head. Thus it provide a possibility to design a more simple to wear EEG measurement device with a small surface rather than a fully covered helmet.

6.2 Structure optimization

The classification ability of a machine learning model is related to the structure of the model, which includes the depth of the model, the type of layers used, and the number of neural units in every layer. In this section, these factors will be discussed and the optimal structure will be acquired by comparing the models with different structures.

6.2.1 Model depth

In the previous chapters, there are two BiLSTM layers in the multi-layer BiLSTM model. In order to verify if more layers can improve the classification ability of the model, a multi-layer BiLSTM model containing 3 BiLSTM layers are trained and compared with the two-BiLSTM-layer model. The comparison is presented in Table 6.3. It

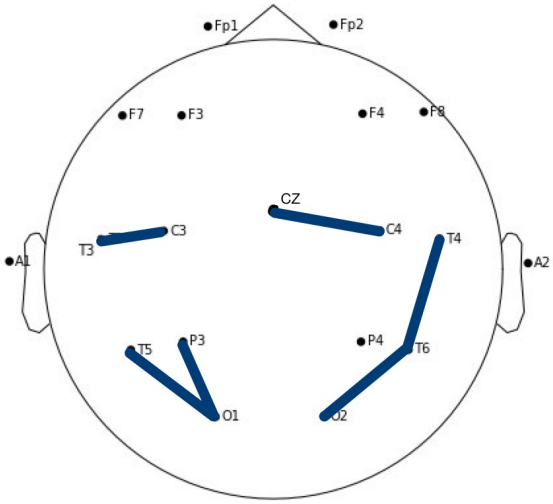


Figure 6.2: Montage of the 6 selected channels

shows that increasing the depth of the model will not improve the model's classification ability, and the performance of all evaluation metrics is even slightly worse than the two-layer model. It may because of the over-fitting as a result of the too complex model. This result suggests that two layers are enough for this problem.

Table 6.3: Comparison of models with different depth

BiLSTM layers	Params	Acuuracy	Precision	Sensitivity	F1 score	FAs/24hrs
2	35745	0.8540	0.7238	0.6017	0.6571	13.12
3	38353	0.8378	0.6974	0.5543	0.6176	13.61

6.2.2 Batch Normalization (BN) layer

During the training process of the neural network, the increasing number of layers will result in the overall distribution of the input values of the activation function gradually moving closer to the upper and lower limits of the activation function. This makes each layer of the input no longer independently distributed because the data of the previous layer needs to adapt to the new input distribution, and the parameters of each layer change optimally with training, which eventually leads to the disappearance of the gradient of the neural network in the lower layers during back-propagation. This phenomenon is called internal covariate shift (ICS), which is difficult to avoid in complex model training. The scaling factor of BN can effectively identify the neurons that do not contribute much to the network. After the activation function, some neurons can be weakened or eliminated automatically. In addition, due to normalization, the model with BN layer rarely has too much variation in parameter value due to different data distributions.

BN layer is widely used in classification models. And for this seizure detection model, a BN layer is added after each LSTM layer in order to improve the classification ability of the model. Table 6.4 compares the performance of the multi-layer BiLSTM model with and without BN layers. The result proves the BN layer's effectiveness, which shows that all evaluation metrics perform better. So two BN layers should be added to this seizure detection model.

Table 6.4: Comparison of models with and without BN layers

Model type	Params	Acuuracy	Precision	Sensitivity	F1 score	FAs/24hrs
Without BN layers	35745	0.8540	0.7238	0.6017	0.6571	13.12
With BN layers	36129	0.8561	0.7715	0.5869	0.6666	10.13

6.2.3 Number of hidden units

There are no rules for determining the number of hidden units in an LSTM layer. More hidden units can contain more details and richer expressiveness, but it also brings disadvantages such as over-fitting and longer computation time. The hidden size should be determined by the requirements of different problems. In general, the more complex the dataset is, the more hidden units are required. And the more complex data-generating process will also increase the number of needed units. The specific number can only be figured out by trials. So several combinations of different numbers of units in each layer are tested. The result obtained from the experiments is shown in Table 6.5. As can be seen from the table, the model is more effective when the first LSTM layer has twice as many hidden units as the second LSTM layer. Meanwhile, the performance of the model slightly improves as the complexity of the model decreases at first. However, after the 32/16 combination, the learning ability of the model becomes progressively worse as the number of units decreases. It can be concluded that 32/16 is the optimal combination of the number of hidden units.

Table 6.5: Comparison of models with different number of hidden units

Hidden units	Params	Acuuracy	Precision	Sensitivity	F1 score	FAs/24hrs
64/32	109121	0.8420	0.7336	0.5375	0.6205	10.79
32/32	50753	0.8459	0.7122	0.5789	0.6387	13.28
32/16	36129	0.8561	0.7715	0.5869	0.6667	10.13
32/8	30353	0.8778	0.6751	0.6512	0.6630	19.10
16/8	13457	0.8453	0.6421	0.6801	0.6606	23.41

6.3 Epoch

An epoch is a process by which a complete data set has passed through the neural network once and returned once. The number of epochs varies for different problems

and different sizes of training sets. Generally, the model is first trained with the training set, and then the performance of the model is evaluated on the validation set. As the number of epochs increases, the model performs better and better on the test set, but if the number is too large, it will cause the model to over-fit the training data and the performance of the test set will decline. Ideally, the inflection point where the model changes from good to bad is the optimal point, which is named as early stopping. It is the most popular strategy to decide whether to stop training. However, in this problem, the difference between the training set and the test set is significant, and the performance of the validation set can not fully represent the performance of the model on the test set. Therefore, the performance of the models trained with 50, 100, 150, 200, 250, 300, and 350 epochs are compared directly on the test set to determine whether the model is able to continue converging and whether the most appropriate number of epochs has reached. The performance is shown in Table 6.6.

Table 6.6: Models trained with different number of epochs

Epochs	Acuuracy	Precision	Sensitivity	F1 score	FAs/24hrs
50	0.8561	0.7715	0.5869	0.6667	10.13
100	0.8507	0.6883	0.6299	0.6578	16.77
150	0.8459	0.7431	0.5546	0.6351	10.79
200	0.8639	0.7857	0.5910	0.6746	8.97
250	0.8560	0.7082	0.6436	0.6744	15.94
300	0.8438	0.7336	0.5556	0.6323	11.45
350	0.8372	0.7045	0.5454	0.6149	12.95

It can be referred from the model that there is no obvious improvement as the number of epochs increases. And the model performs best when the number of epochs equals 200. So 200 is considered as the optimal number of epochs.

6.4 Result

In this chapter, three different methods are used to optimize the seizure detection model. Firstly, a filtered feature selection method was used to reduce the input data size. 6 channels were selected from the original 18-channel EEG dataset, which reduced the model's parameters from 69537 to 44193, 36.45% decrease. It not only reduced the complexity of the model itself but also saved the amount of time in data loading. Then, the structure of the model was optimized. First, we used a model with more layers to verify whether the depth of the adult model can further improve its performance of the model. The results proved that the two-layer model was sufficient to complete this experiment, and the maturity of the morning-home model did not help to improve the model prediction level. Next, we tried to add BN layers to the model to speed up the convergence, prevent over-fitting and prevent gradient explosion to improve the classification ability of the model. The use of BN layers led to a certain improvement in accuracy, precision, sensitivity, and F1 score of the model. At the same time, the number of model parameters increased from 35745 to 38353, only 7.3% increase. Therefore,

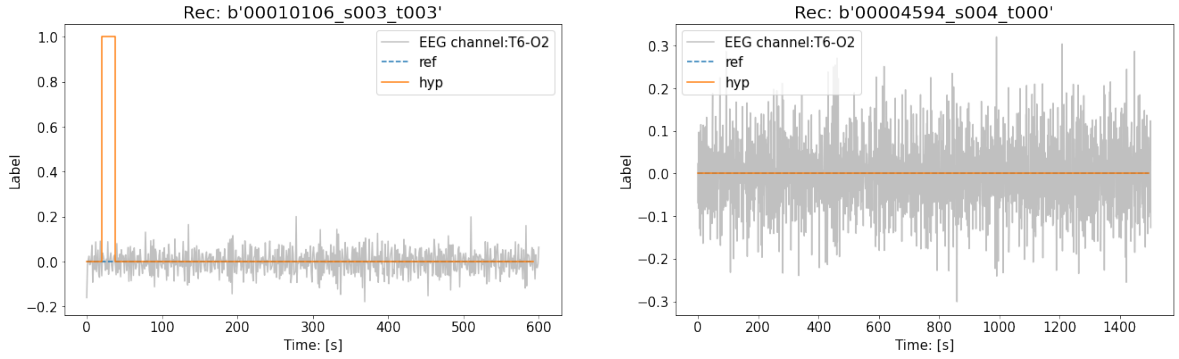
the BN layer can be added to the model as an improved structure for model optimization. Finally, we discussed the influence of the number of hidden units per layer on the model. By testing several sets of combinations according to the gradient design, it can be concluded that the model consisting of two LSTM layers with 32 and 16 hidden units, respectively, performs best. It should be noted that the BiLSTM layers were used in this model, which means each of these layers consists of two identical LSTM layers. So the two BiLSTM layers in this model have 64 and 32 neurons, respectively. Finally, we talked about the effect of the number of epochs during the training process. As it can be seen that the model has largely converged after 50 epochs, and there is no significant over-fitting appearing. Therefore, we finally used 200 epochs to train the optimized model.

7 Result

7.1 Real data visualization

This model is an event-based model, but it also has the ability to detect the start and end time for each event. It has different abilities to detect different types of seizures and seizures of different lengths. In this section, the predictions of different signals will be visualized and compared. And the performance of the model will be analyzed in the following sections. The blue dashed line is the label of the signal. And the orange line is the hypothesis(predictions) of the signal.

Figure 7.1 is the visualization of two normal signals. In Figure 7.1a, there is a false alarm exist. These false alarms in normal signals are the main source of all false alarms. And Figure 7.1b illustrates a correctly predicted normal signal.



(a) Normal signal with one false alarm

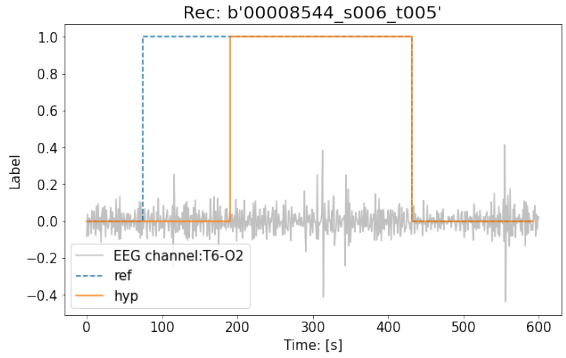
(b) Correctly predicted normal signal

Figure 7.1: Predictions for normal signals

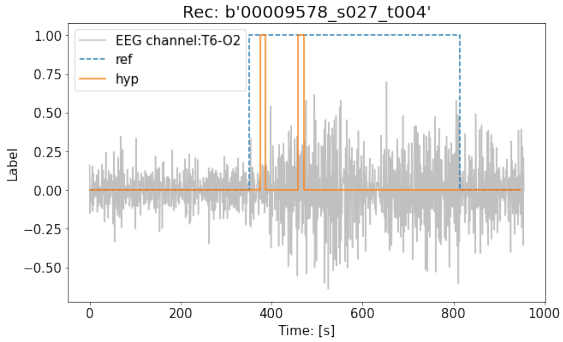
A seizure longer than 5 minutes will be classified as a long seizure, which is much more serious than short seizures. Figure 7.2 shows the predictions for two long seizure signals. According to the OVLP evaluation algorithm, both of these seizures are predicted correctly. However, in Figure 7.2a the hypothesis is still a long seizure, but in Figure 7.2b, the hypothesis shows that there are two short seizures. So, it's apparent that the first prediction is more in accordance with the actual circumstance.

There is another common scenario for seizure occurrence: multiple short seizures occur continuously in one recording. It is illustrated in Figure 7.3. In Figure 7.3a, all four seizures are correctly predicted. And in Figure 7.3b, only three of the five short seizures are truly predicted. The other two seizures are missed.

Compared to multiple short seizures occurring continuously, more short seizures happen alone. Figure 7.4 shows this scenario, one of which is correctly predicted and the other one is predicted as the normal signal. So in Figure 7.4a there is one TP and

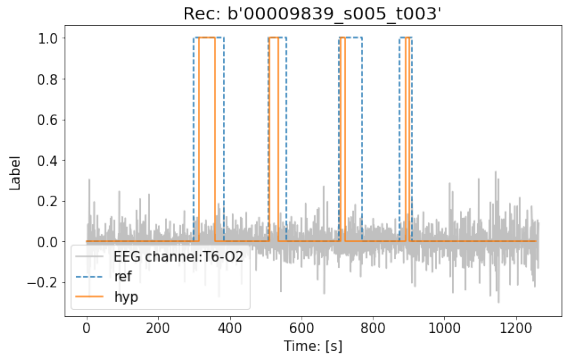


(a) Long seizure with 1 TP predcition

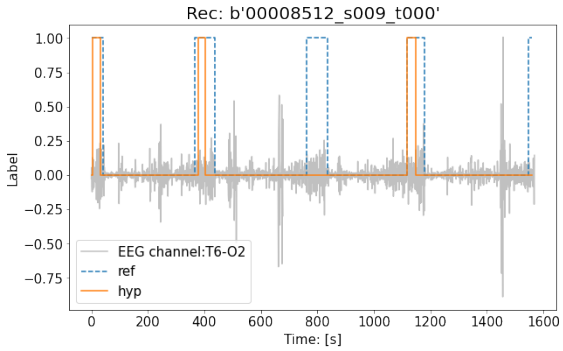


(b) long seizure with 2 TP predictions

Figure 7.2: Predictions for long seizure signals



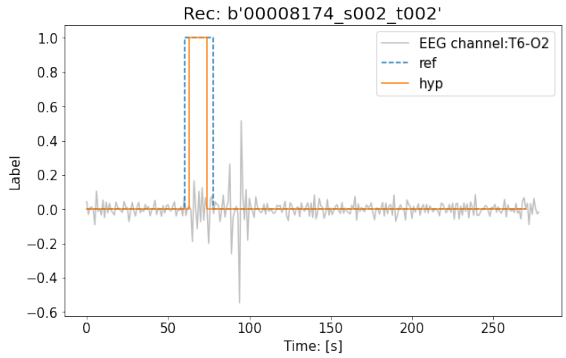
(a) Multiple short seizures with 4 TP predictions



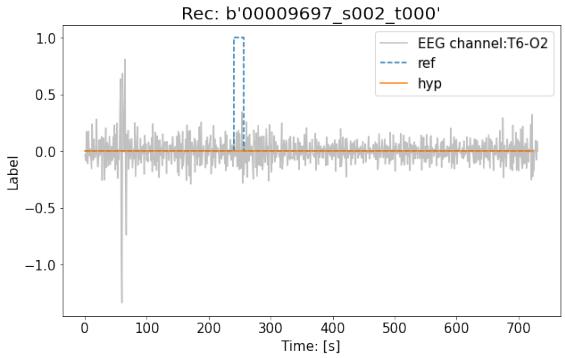
(b) Multiple short seizures with 3 TP predictions and two FN predictions

Figure 7.3: Predictions for signals with multiple short seizures

in Figure 7.4b there is one FN.



(a) Correctly predicted short seizure



(b) Short seizure not correctly predicted

Figure 7.4: Predictions for signals one short seizure

7.2 Seizure length

Seizures of different lengths have varying degrees of difficulty to detect. The reason is that the feature of flakes, long seizures, and seizures of normal length is not exactly the same. Besides, in this model, each recording is segmented into epochs and classified epoch by epoch. And even in the same recording, the signals in different phases are also different. Considering the model's different ability to detect these phases and the different percentages of these phases in each seizure, it's inevitable that the model performs differently in classifying seizures of different lengths. In order to analyze the influence of seizure length, the seizures were divided into four categories, flake, normal, long, and emergency. Seizures shorter than 30 seconds and longer than 10 seconds are considered as flakes. Most seizures last from 30 seconds to 200 seconds, which are classified as normal length seizures. And seizures longer than 200 seconds and short than 300 seconds are categorized as long seizures. The longer the seizure lasts, the more dangerous it is. It's difficult for a very long seizure to stop without any medication. So a seizure longer than 5 minutes is a medical emergency. In this section, these seizures are classified into the emergency group.f

To avoid interference between seizures of different lengths, the recordings are divided into two types: the recordings that include only one type of seizures and the recordings include more than one type of seizures. In this case, 111 flake recordings, 180 normal seizure recordings, 36 long seizure recordings, 22 emergency recordings, and 121 mixed recordings were selected. And there are 190 flakes, 332 normal seizures, 39 long seizures, 23 emergencies, and 532 non-seizure epochs in total. The performance of the model in different recording types is shown in Table 7.1. It can be seen that the model performs best in detecting long seizures. And its ability to detect flakes and normal seizures is also considerably good. But it performs badly in detecting emergencies. What's more, it can be inferred that the model is not able to analyze the recordings with more than one length type of seizures. It may be because the inter-phase between two seizures of different lengths is too misleading for the model to classify correctly.

Table 7.1: The performance of the model to detect different length seizures

Recording Type	Accuracy	Precision	Sensitivity	F1 score	FAs/24hrs
Flake	0.7657	0.9921	0.5952	0.7440	1.1791
Normal	0.7992	0.9818	0.6522	0.7837	2.3348
Long	0.8922	1.0	0.75	0.8571	0.0
Emergency	0.5952	0.9	0.36	0.5143	2.8930
Mixed	0.6246	0.8727	0.4571	0.5999	8.0067

7.3 Seizure number

As mentioned in the previous section, it's difficult for the model to correctly classify the inter-phase between two seizures. The number of seizures in one recording will also influence the model's accuracy. Since the number of inter-phases will increase as

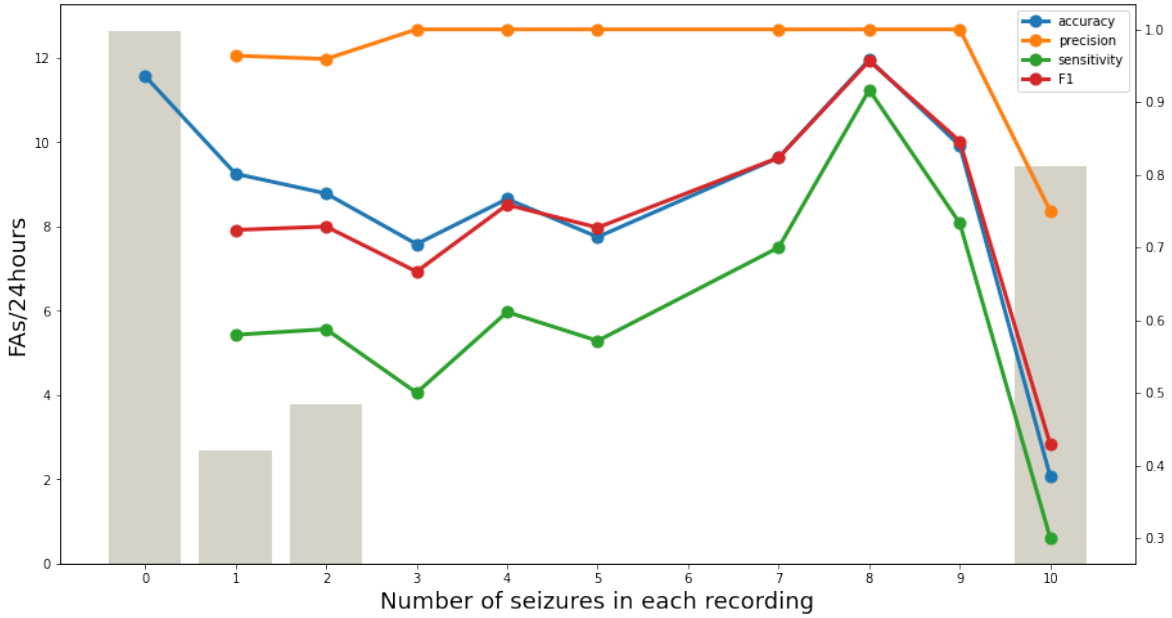


Figure 7.5: The performance of the model to detect recordings with different numbers of seizures

the number of seizures increases. So, in this section, the performance of the model to process recordings with a different number of seizures is evaluated. As shown in Figure 7.5, the model performs best in detecting the recordings with 8 seizures. There was no significant difference in the recordings, including less than 10 seizures. But for number greater than 10, the accuracy declines significantly. It can be deduced that the model can deal with the recordings with a limited number of seizures but is not able to process the recordings with overmany (more than 10) seizures.

7.4 Seizure type

This model is a binary classification model rather than a multi-class classification model. In the training process, all data are labeled as either seizure or non-seizure. And the predictions can only tell if there is a seizure occurs. However, the model itself has different abilities to detect different types of seizures since different seizures have different features. Thus, the performance of the model in detecting different types of seizures is discussed first. A summary of the number of different seizures is shown in Table 7.2. It shows that the FNSZ, GNSZ, CPSZ and TNSZ are four major types of seizures that account for 96% of all seizures. And the seizures of each remaining type are less than 20, which is too small to derive reliable conclusions. So four subsets with all recordings, which include specific seizure types, are created.

The performance of the model to detect these four types of seizures is shown in Table7.3. The model has a similar ability to detect CPSZ, FNSZ, and GNSZ. And it performs best in detecting GNSZ. However, for TNSZ, it performs very badly. It has high precision and low FAs/24hrs, but it can be concluded that it performs well.

Table 7.2: Number of different seizures in test set

Seizure Type	Events	Freq	Cum.
FNSZ	256	46%	46%
GNSZ	172	31%	78%
CPSZ	59	11%	88%
TNSZ	44	8%	96%
TCSZ	16	3%	99%
SPSZ	3	1%	100%
MYSZ	1	0%	100%
ABSZ	0	0%	100%
ATSZ	0	0%	100%
CNSZ	0	0%	100%
Total:	551	100%	

Although all the positive predictions it made are correct, most of the real positive events are missed. And as a result, both the accuracy and sensitivity are quite low. There are two possible reasons. First, the number of TNSZ is relatively small in the training set. In order to get higher accuracy, the model learns more features about other types of seizures and tends to predict more other seizures also. Second, the features of this type are different from the others. The number of CPSZ and TNSZ are similar, but the model is not less accurate in predicting CPSZ. So the reason may be that the selected features of TNSZ have a considerable value difference from other types of seizures. In conclusion, this model cannot predict TNSZ correctly.

Table 7.3: The performance of the model to detect different types of seizures

Seizure Type	Accuracy	Precision	Sensitivity	F1 score	FAs/24hrs
FNSZ	0.8283	0.8598	0.5860	0.6970	8.2291
GNSZ	0.8602	0.9167	0.7237	0.8088	8.8302
CPSZ	0.8298	0.8800	0.6286	0.7333	8.3867
TNSZ	0.5526	1.0000	0.2609	0.4138	0.0000

7.5 Noise Robustness

Noise performance is one of the most important methods to evaluate model robustness. In this experiment, different noise levels are determined by different SNR (signal-to-noise ratio) values. These different levels of noise are added to raw data. Then the noisy signals are used to check the performance of the model under different noise levels. The signals from the TUSZ dataset were collected in different environments with different types and intensities of noise, so the noise is different in each recording. To calculate the noise with the same SNR in each recording, the first 5 seconds are regarded as the background noise. And the equation of SNR is shown in Equation 7.1.

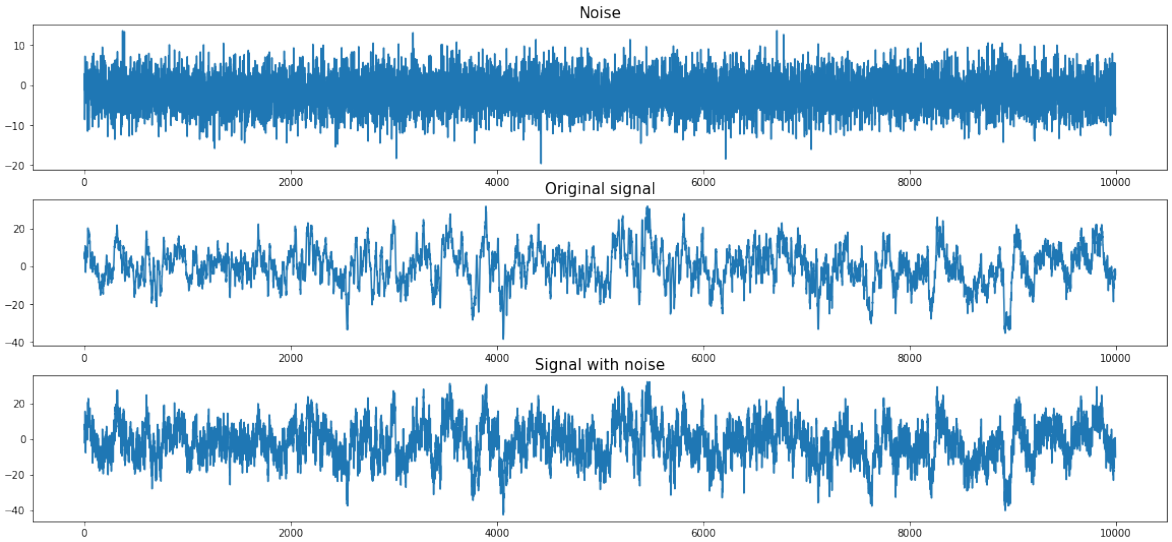


Figure 7.6: Rec: 00005479s004t000 added noise with the SNR = 10dB (T6-O2 channel)

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right) = 10\log_{10}\left[\left(\frac{A_{\mathrm{signal}}}{A_{\mathrm{noise}}}\right)^2\right] \tag{7.1}$$

The added noise is normally distributed Gaussian noise, whose mean and standard deviation is determined by the background noise in each recording. Figure 7.6 illustrated the noise, original signal, and noisy signal with 10dB SNR of recording 00005479s004t000 (T6-O2 channel).

The noise performance of the model is displayed in Figure 7.7. Running horizontally along the figure is SNR in dB, on which the different noise is added based. The left vertical axis is the value of FAs/24hrs, and the right vertical axis is the value of accuracy, precision, sensitivity, and F1 score. When the SNR is small, which means the noise is loud, the number of false alarms is extremely low. Since almost all recordings will be detected as the normal signal, the sensitivity is extremely low as well. As the SNR increases, the noise amplitude gradually decreases, and the model begins to predict the seizures correctly. As a result, the sensitivity also increased accordingly. The accuracy doesn't change significantly because the majority of the recordings are normal signals. It's worth noting that the change in precision is not significant. So F1 score represents the ability of the model to detect seizures correctly and effectively. It can be inferred from the figure that the F1 score stays the same after the SNR reaches 30dB. So the conclusion is that the model can maintain robustness when the SNR is greater than 30dB.

7.6 Computation time evaluate

The objective of this thesis is to provide an energy-efficient algorithmn for wearable EEG measurement device. It requires the device should make classification of the real

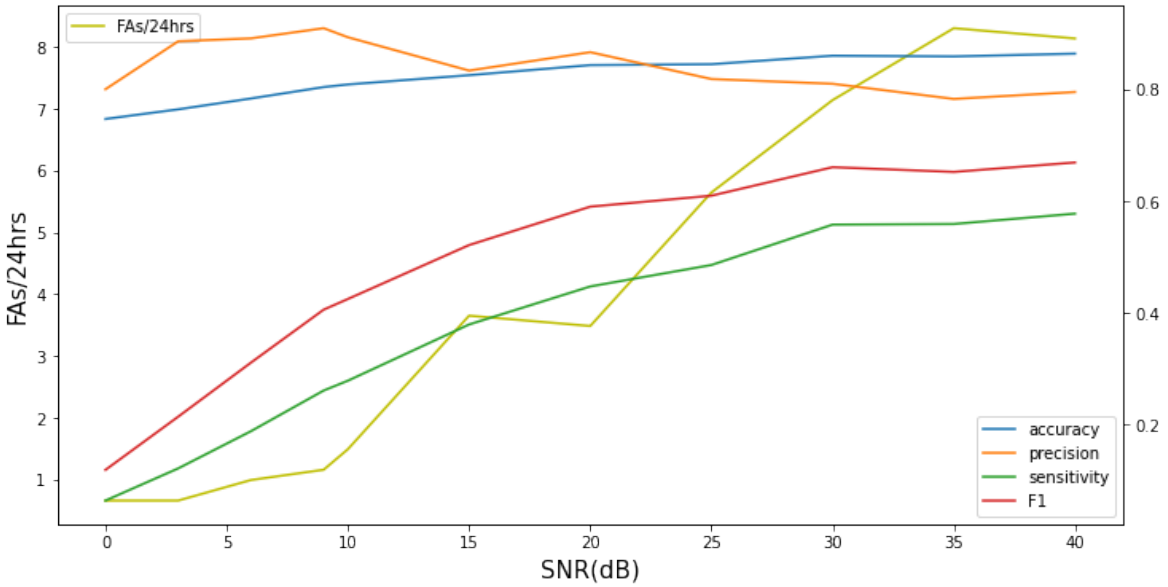


Figure 7.7: Performance of the model with different levels of noise

time data as promptly as possible. In order to achieve this goal, the computation speed of this model should be fast enough. Thus, in this section, the computational time is evaluated. And the computation time before and after optimization has been compared.

There are four major steps in the whole classification process, data loading, feature extraction, classification and post-processing. In the first step, the edf. format EEG data are loaded. And the second step, feature extraction, 11 features are extracted from each channel. In the third step, the ML model make raw classification for each time points. And the last step post-processing merges and filters the raw predictions to get the final predictions for seizure events in each recordings. There are 891 recordings in the test set and the total duration length is 528763s. The computation time of each step is shown in Table 7.4. After optimization, the computation time decreased from 2111.78s to 1063.72s, 49.63% time was saved. And ratio of the total computation time to the total duration length for the optimized model is only 0.2%, which is small enough for a real-time seizure detection device.

Table 7.4: Computation time for each step in the models before and after optimization

Number of channels	Data loading(s)	Feature extraction(s)	Classification(s)	Post-processing(s)	Total computation time(s)
18	673.51	1402.63	61.06	4.58	2111.78
6	252.44	748.50	58.28	4.50	1063.72

7.7 Comparison with other machine learning models

Many machine learning models that can learn spatial and temporal context efficiently can be used for seizure detection. In this section, five different machine learning-based

methods are used to make a comparison with the newly proposed LSTM-based method. All these model were trained with TUH EEG Corpus and evaluated by the OVLP scoring method[76]. HMM/SdA is a hybrid model that uses a hidden Markov model (HMM) as a decoder and a Stacked Denoising Autoencoder (SdA) as a postprocessor[77]. HMM/LSTM consists of an HMM decoder and an LSTM postprocessor[78]. IPCA/LSTM is an Incremental Principal Component Analysis (IPCA) followed by an LSTM decoder[77]. CNN/MLP and CNN/LSTM are both fully based on deep learning methods[79]. The former uses a CNN decoder and a Multi-Layer Perceptron (MLP) postprocessor. The latter uses CNN and LSTM layers together. The comparison is shown in Table 7.5. It shows that the newly proposed ML model has the highest accuracy, sensitivity, and F1 score. CNN/LSTM has the lowest FAs/24hrs, but this value of the BiLSTM model is very close to the CNN/LSTM's and much lower than the other four models. Consequently, this LSTM-based model has an outstanding ability to detect seizures.

Table 7.5: Comparison between the performance of different ML-based models

Model	Accuracy	Sensitivity	F1 score	FAs/24hrs
HMM/SdA	0.651	0.3535	0.31	77.39
HMM/LSTM	0.665	0.3005	0.33	60.92
IPCA/LSTM	0.656	0.3297	0.34	73.52
CNN/MLP	0.669	0.3909	0.38	77.19
CNN/LSTM	0.789	0.3083	0.45	6.75
BiLSTM	0.864	0.5910	0.67	8.97

8 Conclusion

8.1 Summary of the result

This thesis proposes a new machine learning method to detect seizures in EEG signals. This new method uses time-domain features extracted from the raw signals. And the classifier is a hybrid model, which is named as multi-layer BiLSTM model. It consists of two BiLSTM layers and a DNN layer. This model is optimized by experiments with different hyperparameters and structures. The scoring method is a term-based method, OVLP.

In Chapter 3, the unwanted data was filtered first. Then with feature extraction process, 11 time-domain features are selected for the model.

In Chapter 4, the post-processing process was discussed. The predictions should be filtered and merged first. The evaluation metrics used were accuracy, sensitivity, precision, F1 score and FAs/24hrs. The scoring metric used was Any-Overlap Method (OVLP) approach.

In Chapter 5, the length of the epoch and segment to extract features were decided based on the experiments. Also, multi-layer BiLSTM model was selected as the optimal model compared with other structures. The decision threshold of the model's output value is decided by the F1 curve of the post-processed result.

In Chapter 6, the model was optimized from different aspects. First, the input size was reduced by channel selection. The 6 most important channels were selected from all 18 channels. Second, the hyperparameters, including model depth and number of hidden units in each layer were decided by experiments. Also, the effects of using the BN layer was explored. Last, the number of epochs during the training process was talked about.

The problems proposed in Section 1.2 are as follows.

- **Reduce the input data size with some preprocessing strategies**
  The raw EEG data are 18-channel signals. With the 200 Hz sampling frequency, there are 36000 data points in 10 seconds raw signal. In this thesis, the raw signals are divided into epochs first. And 11 time-domain features are extracted from each single-channel epoch. So there are 198 features for each epoch. And the input vector is a feature sequence whose length is the number of epochs in each segment. The epoch is 0.5s, and the segment is 10s in this thesis. So with the preprocessing method, input data size is compressed from (2000, 18) to (20, 198).

- **Propose proper postprocessing steps to get the final result**
  The raw predictions are filtered and merged to form final predictions first. Then OVLP, a term-based scoring method, is selected to evaluate the predictions. Based

on this scoring method, the evaluation metrics, including accuracy, sensitivity, precision, F1 score and FAs/24hrs are used to evaluate the model's performance.

- **Design an LSTM-based model structure for seizure detection**
  In order to achieve the objective of this project, an improved LSTM structure, BiLSTM is used. The model proposed by this thesis is multi-layer BiLSTM model. It performs better than the other two LSTM models: BiLSTM-DNN model and DNN-BiLSTM model.

- **Optimize the model to decrease its complexity and improve its performance**
  From the experiments, 6 most important channels are selected from the 18 channels. The optimal structure of the model is the multi-layer BiLSTM model with two BiLSTM layers in total. Each BiLSTM layer is followed by a BN layer. The hidden units number of each BiLSTM layer is 64 and 32. Finally, the model is trained for 200 epochs.

8.2 Future work

In this thesis, a new energy-efficient seizure detection algorithm is proposed. There are still some possible future works for it to explore.

- Deep learning is also a powerful feature extraction approach, which can be used to replace the time-domain feature extractors.

- The parameters of the model trained by Tensorflow are floating numbers. Some quantization methods can be applied to minimize the memory usage in the hardware.

- The model proposed by this thesis is a binary classification model. And it's possible to be trained as mutli-class classification to detect not only the occurrence but also the type of the seizures.

- Some other advanced LSTM layers, such as stacked LSTM and Gated Recurrent Unit (GRU) may also be the possible choice for this problem.

- The model is designed for the TM circuit. Some hardware simulations can be made to evaluate its power consumption performance further.

Bibliography

[1] W. H. O. Who, "Epilepsy," *World Health Organization: WHO*, Jun 2019.

[2] M. M. Zack, "National and State Estimates of the Numbers of Adults and Children with Active Epilepsy — United States, 2015," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 66, 2017.

[3] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel, L. Forsgren, J. A. French, M. Glynn, D. C. Hesdorffer, B. I. Lee, G. W. Mathern, S. L. Moshé, E. Perucca, I. E. Scheffer, T. Tomson, M. Watanabe, and S. Wiebe, "ILAE Official Report: A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, pp. 475–482, Apr 2014.

[4] "About epilepsy — cdc." https://www.cdc.gov/epilepsy/about/index.htm. (Accessed on 05/25/2021).

[5] J. C. Henry, "Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Fifth Edition," *Neurology*, vol. 67, pp. 2092–2092–a, Dec 2006.

[6] F. Yuan, *CMOS Time-Mode Circuits and Systems*. Boca Raton, FL, USA: CRC Press, 2018.

[7] S. S. Haykin, *Neural Networks and Learning Machines, Volume 10*. Upper Saddle River, NJ, USA: Prentice Hall, 2009.

[8] S. Hochreiter and J. Schmidhuber, "Flat minima," *Neural Computation*, vol. 9, p. 1–42, Jan 1997.

[9] M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, and J. Xiang, "Epileptic Seizure Detection Based on EEG Signals and CNN," *Front. Neuroinf.*, vol. 0, 2018.

[10] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Integer convolutional neural network for seizure detection," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 4, pp. 849–857, 2018.

[11] M. Shahbazi and H. Aghajan, "A generalizable model for seizure prediction based on deep learning using cnn-lstm architecture," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 469–473, 2018.

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

[13] T. Gandhi, B. K. Panigrahi, M. Bhatia, and S. Anand, "Expert model for detection of epileptic activity in EEG signature," *Expert Syst. Appl.*, vol. 37, pp. 3513–3520, Apr. 2010.

[14] S. L. Moshé, E. Perucca, P. Ryvlin, and T. Tomson, "Epilepsy: new advances," *Lancet*, vol. 385, pp. 884–898, Mar. 2015.

[15] T. F. Hasan and W. O. Tatum, "When should we obtain a routine eeg while managing people with epilepsy?," *Epilepsy Behavior Reports*, vol. 16, p. 100454, 2021.

[16] E. Ben-Menachem, "Vagus-nerve stimulation for the treatment of epilepsy," *Lancet Neurol.*, vol. 1, pp. 477–482, Dec. 2002.

[17] B. Hunyadi, *Learning from structured EEG and fMRI data supporting the diagnosis of epilepsy*. PhD thesis, PhD thesis, 2014.

[18] L. A. van Graan, L. Lemieux, and U. J. Chaudhary, "Methods and utility of EEG-fMRI in epilepsy," *Quant. Imaging Med. Surg.*, vol. 5, p. 300, Apr 2015.

[19] V. Pravdich-Neminsky, *Ein Versuch der Registrierung der elektrischen Gehirnerscheinungen". Zentralblatt für Physiologie*. 1913.

[20] L. F. Haas, "Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography," *J. Neurol. Neurosurg. Psychiatry*, vol. 74, p. 9, Jan. 2003.

[21] N. Birbaumer, "Brain-computer-interface research: coming of age," *Clin. Neurophysiol.*, vol. 117, pp. 479–483, Mar. 2006.

[22] "Automatic recognition of epileptic seizures in the EEG," Nov. 1982. [Online; accessed 25. Aug. 2022].

[23] C. J. Park and S. B. Hong, "High Frequency Oscillations in Epilepsy: Detection Methods and Considerations in Clinical Application," *Journal of Epilepsy Research*, vol. 9, p. 1, June 2019.

[24] O. Mecarelli, "Electrode Placement Systems and Montages," in *Clinical Electroencephalography*, pp. 35–52, Cham, Switzerland: Springer, 2019.

[25] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 52, no. 3-6., p. ;, 1999.

[26] "iSyncWave | xn–2z1bo3huti4pc9wd [iMediSync]," Apr. 2022. [Online; accessed 15. Jul. 2022].

[27] E. N. V., "BrainLink EEG headset," July 2022. [Online; accessed 15. Jul. 2022].

[28] M. T. Knierim, C. Berger, and P. Reali, "Open-Source Concealed EEG Data Collection for Brain-Computer-Interfaces – Real-World Neural Observation Through OpenBCI Amplifiers with Around-the-Ear cEEGrid Electrodes," *ResearchGate*, Jan. 2021.

[29] S. Park, C.-H. Han, and C.-H. Im, "Design of Wearable EEG Devices Specialized for Passive Brain–Computer Interface Applications," *Sensors (Basel, Switzerland)*, vol. 20, Aug. 2020.

[30] D. Looney, P. Kidmose, and D. P. Mandic, *Ear-EEG: User-Centered and Wearable BCI*, pp. 41–50. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[31] Y. Gu, E. Cleeren, J. Dan, K. Claes, W. Van Paesschen, S. Van Huffel, and B. Hunyadi, "Comparison between Scalp EEG and Behind-the-Ear EEG for Development of a Wearable Seizure Detection System for Patients with Focal Epilepsy," *Sensors (Basel).*, vol. 18, p. 29., Dec. 2017.

[32] P. Kidmose, D. Looney, M. Ungstrup, M. L. Rank, and D. P. Mandic, "A Study of Evoked Potentials From Ear-EEG," *IEEE Trans. Biomed. Eng.*, vol. 60, pp. 2824–2830, May 2013.

[33] K. B. Mikkelsen, S. L. Kappel, D. P. Mandic, and P. Kidmose, "EEG Recorded from the Ear: Characterizing the Ear-EEG Method," *Front. Neurosci.*, vol. 9, Nov 2015.

[34] K. B. Mikkelsen, D. B. Villadsen, M. Otto, and P. Kidmose, "Automatic sleep staging using ear-EEG," *Biomed. Eng. Online*, vol. 16, pp. 1–15, Dec 2017.

[35] V. Goverdovsky, D. Looney, P. Kidmose, and D. P. Mandic, "In-ear eeg from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 271–277, 2016.

[36] V. Srinivasan, C. Eswaran, and N. Sriraam, "Artificial Neural Network Based Epileptic Detection Using Time-Domain and Frequency-Domain Features," *J. Med. Syst.*, vol. 29, pp. 647–660, Dec. 2005.

[37] V. K. Harpale and V. K. Bairagi, "Time and frequency domain analysis of EEG signals for seizure detection: A review," in *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, pp. 1–6, IEEE, Jan. 2016.

[38] L. Logesparan, A. J. Casson, and E. Rodriguez-Villegas, "Optimal features for online seizure detection," *Med. Biol. Eng. Comput.*, vol. 50, pp. 659–669, July 2012.

[39] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[40] W. R. S. Webber, R. P. Lesser, R. T. Richardson, and K. Wilson, "An approach to seizure detection using an artificial neural network (ANN)," *Electroencephalogr. Clin. Neurophysiol.*, vol. 98, pp. 250–272, Apr. 1996.

[41] L. Guo, D. Rivero, J. Dorado, J. R. Rabuñal, and A. Pazos, "Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks," *J. Neurosci. Methods*, vol. 191, pp. 101–109, Aug. 2010.

[42] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sept. 1995.

[43] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with Support Vector Machines," *Clin. Neurophysiol.*, vol. 122, pp. 464–473, Mar. 2011.

[44] Y. Liu, W. Zhou, Q. Yuan, and S. Chen, "Automatic Seizure Detection Using Wavelet Transform and SVM in Long-Term Intracranial EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, pp. 749–755, July 2012.

[45] T. Zhang and W. Chen, "LMD Based Features for the Automatic Seizure Detection of EEG Signals Using SVM," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, pp. 1100–1108, Sept. 2016.

[46] P. Ghaderyan, A. Abbasi, and M. H. Sedaaghi, "An efficient seizure prediction method using KNN-based undersampling and linear frequency measures," *J. Neurosci. Methods*, vol. 232, pp. 134–142, July 2014.

[47] J. Birjandtalab, M. Baran Pouyan, D. Cogan, M. Nourani, and J. Harvey, "Automated seizure detection using limited-channel EEG and non-linear dimension reduction," *Comput. Biol. Med.*, vol. 82, pp. 49–58, Mar. 2017.

[48] P. P. M. Shanir, K. A. Khan, Y. U. Khan, O. Farooq, and H. Adeli, "Automatic Seizure Detection Based on Morphological Features Using One-Dimensional Local Binary Pattern on Long-Term EEG," *Clin. EEG Neurosci.*, vol. 49, pp. 351–362, Sept. 2018.

[49] T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, pp. 278–282 vol.1, 1995.

[50] C. Donos, M. Dümpelmann, and A. Schulze-Bonhage, "Early Seizure Detection Algorithm Based on Intracranial EEG and Random Forest Classification," *Int. J. Neural Syst.*, vol. 25, p. 1550023, Apr. 2015.

[51] M. Mursalin, Y. Zhang, Y. Chen, and N. V. Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier," *Neurocomputing*, vol. 241, pp. 204–214, June 2017.

[52] T. Zhang, W. Chen, and M. Li, "AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier," *Biomed. Signal Process. Control*, vol. 31, pp. 550–559, Jan. 2017.

[53] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Sig. Process.*, vol. 151, p. 107398, Apr. 2021.

[54] G. R. Minasyan, J. B. Chatten, M. J. Chatten, and R. N. Harner, "Patient-Specific Early Seizure Detection from Scalp EEG," *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, vol. 27, p. 163, June 2010.

[55] L. Vidyaratne, A. Glandon, M. Alam, and K. M. Iftekharuddin, "Deep recurrent neural network for seizure detection," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 1202–1207, IEEE, July 2016.

[56] A. M. Abdelhameed, H. G. Daoud, and M. Bayoumi, "Deep Convolutional Bidirectional LSTM Recurrent Neural Network for Epileptic Seizure Detection," in *2018 16th IEEE International New Circuits and Systems Conference (NEWCAS)*, pp. 139–143, IEEE, June 2018.

[57] Y. Li, Z. Yu, Y. Chen, C. Yang, Y. Li, X. Allen Li, and B. Li, "Automatic Seizure Detection using Fully Convolutional Nested LSTM," *Int. J. Neural Syst.*, vol. 30, p. 2050019, Feb. 2020.

[58] X. Hu, S. Yuan, F. Xu, Y. Leng, K. Yuan, and Q. Yuan, "Scalp EEG classification using deep Bi-LSTM network for seizure detection," *Comput. Biol. Med.*, vol. 124, p. 103919, Sept. 2020.

[59] V. Shah, E. von Weltin, S. Lopez, J. R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, and J. Picone, "The Temple University Hospital Seizure Detection Corpus," *Front. Neuroinf.*, vol. 0, 2018.

[60] F. Lotte, "A new feature and associated optimal spatial filter for EEG signal classification: Waveform Length," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 1302–1305, IEEE, Nov. 2012.

[61] A. A. Torres-García, O. Mendoza-Montoya, M. Molinas, J. M. Antelis, L. A. Moctezuma, and T. Hernández-Del-Toro, "Pre-processing and feature extraction," in *Biosignal Processing and Classification Using Computational Learning and Intelligence*, pp. 59–91, Cambridge, MA, USA: Academic Press, Jan. 2022.

[62] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Chichester, England, UK: Wiley, 2013.

[63] I. Stancin, M. Cifrek, and A. Jovic, "A Review of EEG Signal Features and their Application in Driver Drowsiness Detection Systems," *Sensors (Basel).*, vol. 21, p. 3786., May 2021.

[64] V. Ravinuthula, *Time-mode circuits for analog computation*. PhD thesis, University of Florida, USA.

[65] R. J. D'Angelo and S. R. Sonkusale, "A time-mode translinear principle for nonlinear analog computation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, p. 2187–2195, Sep 2015.

[66] J.-M. Muller, "Elementary functions and approximate computing," *Proceedings of the IEEE*, vol. 108, no. 12, pp. 2136–2149, 2020.

[67] W. Himmelbauer and A. Andreou, "Log-domain circuits in subthreshold mos," in *Proceedings of 40th Midwest Symposium on Circuits and Systems. Dedicated to the Memory of Professor Mac Van Valkenburg*, vol. 1, pp. 26–30 vol.1, 1997.

[68] D. Greenberg, M. Aminoff, and R. Simon, *Clinical Neurology 8/E*. New York, NY, USA: McGraw-Hill Education, 2012.

[69] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognit. Lett.*, vol. 30, pp. 27–38, Jan. 2009.

[70] S. Ziyabari, V. Shah, M. Golmohammadi, I. Obeid, and J. Picone, "Objective evaluation metrics for automatic classification of EEG events," *arXiv*, Dec. 2017.

[71] M. Varsavsky, *Epileptic Seizures and the EEG:Measurement, Models, Detection and Prediction*. Andover, England, UK: Taylor & Francis, Jan. 2011.

[72] M. Vavrečka and L. Lhotská, "Eeg feature selection based on time series classification," *Machine Learning and Data Mining in Pattern Recognition*, p. 520–527, 2013.

[73] P. Geethanjali, Y. K. Mohan, and J. Sen, "Time domain Feature extraction and classification of EEG data for Brain Computer Interface," in *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 1136–1139, IEEE, May 2012.

[74] A. Shoeibi, M. Khodatars, N. Ghassemi, M. Jafari, P. Moridian, R. Alizadehsani, M. Panahiazar, F. Khozeimeh, A. Zare, H. Hosseini-Nejad, A. Khosravi, A. F. Atiya, D. Aminshahidi, S. Hussain, M. Rouhani, S. Nahavandi, and U. R. Acharya, "Epileptic Seizures Detection Using Deep Learning Techniques: A Review," *Int. J. Environ. Res. Public Health*, vol. 18, p. 5780, May 2021.

[75] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, pp. 1–21, Dec. 2015.

[76] V. Shah, M. Golmohammadi, I. Obeid, and J. Picone, "Objective evaluation metrics for automatic classification of eeg events," *Biomedical Signal Processing*, p. 223–255, 2021.

[77] M. Golmohammadi, A. H. Harati Nejad Torbati, S. Lopez de Diego, I. Obeid, and J. Picone, "Automatic Analysis of EEGs Using Big Data and Hybrid Deep Learning Architectures," *Front. Hum. Neurosci.*, vol. 0, 2019.

[78] M. Golmohammadi, S. Ziyabari, V. Shah, S. L. de Diego, I. Obeid, and J. Picone, "Deep Architectures for Automated Seizure Detection in Scalp EEGs," *arXiv*, Dec. 2017.

[79] M. Golmohammadi, V. Shah, I. Obeid, and J. Picone, "Deep Learning Approaches for Automated Seizure Detection from Scalp Electroencephalograms," in *Signal Processing in Medicine and Biology*, pp. 235–276, Cham, Switzerland: Springer International Publishing, 2020.

[80] A. K. Geim and H. A. M. S. ter Tisha, "Detection of earth rotation with a diamagnetically levitating gyroscope," *Physica B: Condensed Matter*, vol. 294–295, pp. 736–739, 2001.

[81] A. K. Another and H. A. M. S. ter Tisha, "Detection of earth rotation with a diamagnetically levitating gyroscope," *Physica B: Condensed Matter*, vol. 294–295, pp. 736–739, 2001.

[82] "EPILEPSY IN THE WHO EUROPEAN REGION: Fostering Epilepsy Care in Europe," Jan 2022.

[83] M. Shahbazi and H. Aghajan, "A GENERALIZABLE MODEL FOR SEIZURE PREDICTION BASED ON DEEP LEARNING USING CNN-LSTM ARCHITECTURE," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 26–29, IEEE.

[84] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, pp. 1735–1780, Nov. 1997.