# Automatic detection of epileptic seizure onset and offset in scalp EEG

by

**Poomipat Boonyakitanont**

Advisor

**Assistant Professor Jitkomut Songsiri, Ph.D.**
**Assistant Professor Apiwat Lekutai, Ph.D.**
**Assistant Professor Krisnachai Chomtho, M.D.**

A thesis proposal presented for the degree of
Doctor of Philosophy in Electrical Engineering

Department of Electrical Engineering
Chulalongkorn University
Thailand
November 19, 2019

# Contents

# 1 Introduction

Defined by the International League Against Epilepsy (ILAE), an epileptic seizure is a transitory occurrence of symptoms due to abnormal excessive or synchronous neuronal activity in the brain [FAA+14]. It was reported that 65 million people of all ages are affected the epilepsy [TBB+11]. Consequences of epilepsy are dependent on types of seizures and areas that the seizures appear. For instance, a tonic-clonic seizure can initiate from one side or both sides of the brain. People affected by tonic-clonic seizures have uncontrollable, stiffening, and jerking muscles that may cause the people fall down or bite their tongue [BR07]. An absence seizure which is a generalized onset seizure affects patient's awareness. Absence seizures usually have effects in a short period, less than 10 seconds, but there are also absence seizures that last longer [RT03]. Due to the impacts of epileptic seizures, which can lead to neuronal and physical injuries, patients with recurrent or prolonged seizures should be reviewed by neurologists for a prompt diagnosis and treatment. Neurologists usually monitor the patients with continuous video-EEG monitoring [SS97, MFF+13] for those having refractory status epilepticus that are unresponsive to therapy. This is a combination of electroencephalography (EEG) and video, recorded simultaneously to observe brain activities in relation with a clinical change. Nevertheless, this task is still a time-consuming process for the neurologists to review the continuous EEG. Therefore, automated epileptic seizure detections using EEG signals have been developed to facilitate the interpretation of long-term monitoring.

Seizure onset detection has an important role in situations that need immediate treatments, especially in cases when patients do not respond to the medication. There are two types of seizure oasdft that can be inspected from the scalp EEG signals regarding to the spatial distribution of the seizure activity. When a seizure originates at some point rapidly distributing the whole networks, causing EEG changes apparently on the whole brain, it is called a generalized-onset seizure. In contrast, a seizure is focal-onset when originating within networks limited to one hemisphere, making the changes in EEG restricted in a particular brain region [SRS+19, FCD+17]. Some patient who requires a treatment to reduce a seizure effect after the seizure starts needs a seizure detection system that alarms immediately, or a few seconds later, after the seizure onset. A detection delay from the actual seizure onset can cause wrong localization of an epileptogenic focus and late therapy [NFBA13]. For instance, a responsive neurostimulation system is a device that is implanted in the brain to observe and stimulate brain activities [MR18, Gel18]. This device releases electricity to reduce an impact of seizure after the seizure onset occurs. Hence, a nearly correct indication of the seizure onset is needed for the proper treatment.

Moreover, detecting seizure offset is also important. Seizure offset recognition can help reduce the side effects in postictal states by a prompt treatment [VRB10]. Providing a period of an epileptic seizure in an EEG record to neurologists instead of only an occurrence of the seizure can better assist the neurologists to consequently analyze and diagnose types of seizure so that the patients receive antiepileptic drug (AED) therapy properly [Gol10]. For instance, it is highly possible that the seizure still maintains if a patient affected by epilepsy longer than five minutes does not receive therapy properly. In this case, lack of treatment can considerably damage the human brain. However, it is not easy to indicate the seizure offset following the seizure activity. There are several possibilities of transitions from the seizure activities to their terminations; the seizure offset cannot be directly observed from the channels where the seizure initiates [SB10]. It is possible that a focal-onset seizure is still localized or developed to the whole brain. For example, the focal-onset seizure can be evolved to a secondary generalized seizure, a seizure activity spreading from the focal area to the whole brain. Generalized-onset seizures can also end with focal or generalized activities.

From the needs and importance of the automatic detection of epileptic seizures and the starting and ending points, this work mainly concentrates on detecting the seizure events and determining their onsets and offsets. We aims to develop methods of the automatic epileptic seizure detection and of seizure onset-offset localization using only EEG signals. We divide the whole project into two main steps: epoch-based classification and onset-offset detection. The epoch-based classification is to classify epochs from long EEG signals, and the onset-offset detection adopts the epoch-based results to improve the classification performance and indicate the seizure onsets and offsets.

In Section 2, the proposal overview is described, including objective, scope of work, and benefit and outcome. Section 3 reviews backgrounds which includes EEG and montages, characteristics of EEG, and convolutional neural network (CNN). Studies related to the detection of seizure events, onsets and offsets are discussed in Section 4. The problem statement and research methodology are stated in Sections 5 and 6, respectively. Moreover, Section 7 describes a proposed model including classification in epoch-based seizure detection and onset-offset detection technique. The data sets of scalp EEG signals that can be downloaded only are explained in details in Section 8. In addition, All experimental settings including data modification and hyperparameter tuning are clarified in Section 9. Finally, Section 10 summarizes the conclusion, limitations, future work of this thesis.

# 2 Proposal overview

In this section, we present an overview of this proposal. This overview contains the objective, scopes, benefits, and outcomes of this work.

## 2.1 Objectives

This study aims to provide an offline detection method of seizure activities and the identification of their starting and ending points in multi-channel scalp EEG signals. The seizure onsets and offsets can be used to infer when and how long the seizures appear. Furthermore, this method can also be applied to EEG records that have been continuously collected from a subject being monitored. Neurologists can use these results as a guide to further dispense and treat the subject appropriately. Moreover, the results which include the seizure onset and offset can be exploited as a pre-annotation of the data. EEG signals with pre-annotation can reduce time spent by physicians on inspecting types and characteristics of seizures, and labeling a new data set for further research.

## 2.2 Scopes of work

- The proposed system needs to be early trained before it will be used for a specific patient. So data for training and testing must be collected from the same patient.
- Multi-channel scalp EEG signals are used to detect the seizures activities, not intracranial EEG signals. An annotation of each record must contain seizure onsets and offsets of individual seizure events. All training and testing data must be acquired from the same montage, and the data are collected from an online open source.
- The training and testing stages are conducted offline.
- Types of seizures are not specified, and we do not discriminate the types in this work.
- Results of the proposed method and previous methods are compared.

## 2.3 Benefits and outcomes

**Benefits.** With our proposed method, less efforts than usual from the neurologists are required to review the continuous EEG, and a little background knowledge about epilepsy is needed. Moreover, no other modality, *e.g.*, electrocardiogram (ECG), electromyogram (EMG), is included. This means no other equipment is required when collecting the data. In addition, a pre-annotation of an originally unlabeled data for the neurologists to further analyze seizure characteristics is a result from the method. Finally, the method can be used to first label the data of a new data set to enhance an application of machine learning in this research field.

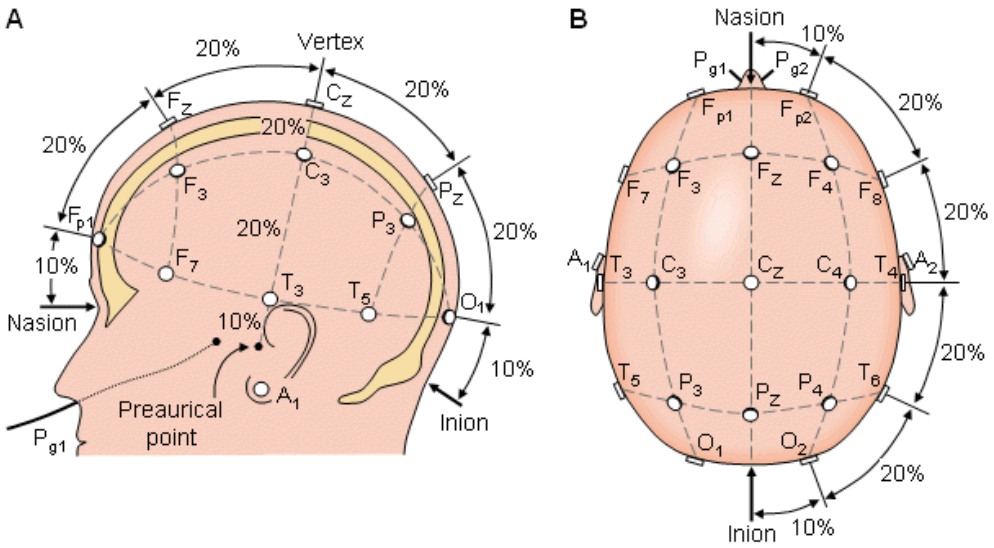



Figure 1: Illustration of the international 10-20 system from (A) left and (B) top views of the head; 'A' stands for an ear lope, from [MP95].

**Outcomes.** First, we provide a method for automatic detection of epileptic seizures and their seizure onsets and offsets using multi-channel EEG signals. The algorithm associated with the proposed method is also given to train a model using EEG signals that contain seizure activities for a specific subject.

# 3 Background

To appropriately design a method for detecting seizures and their onsets and offsets, some background knowledge about characteristics of EEG signals and epileptic seizures is required. This section is divided into three parts: EEG and montages, EEG characteristics, and CNN.

## 3.1 EEG and montages

EEG is one clinical way of recording and studying electric potentials involved with the brain's electrical activities. The study of the electrical activities in the brain using EEG records is one of the most essential tools for diagnosing diseases in neuroscience, for instance, epilepsy, brain tumors, head injury, and sleep disorders. There are two types of EEGs, scalp and intraccranial EEGs, depending on where signals are obtained. The scalp EEG signals are recorded by placing small disks called electrodes in different positions on the scalp surface with liquid gel. For the intracranial EEG (iEEG), or so-called electrocorticogram (ECoG), the subdural electrodes are implanted directly in the brain during the surgery to measure the electrical signals directly from the cortical cortex.

Locations of electrodes on the scalp are critical because the measured signals spatially vary on the position of the scalp; thus, this causes difficulties in interpretations. One of the standard placements of electrodes is the international 10-20 electrode system. As shown in Figure 1, electrodes are placed with 10% or 20% of actual distances between adjacent electrodes in all three directions. The reference points of the system are nasion, the depressed area between the eyes, and inion, the prominent bone locating on the middle line of the skull. Each location is assigned by a letter to specify a lobe and by a number to specify the location of each lobe. The letters F, T, C, P and O are used in the positions of Frontal, Temporal, Central, Parietal and Occipital lopes, respectively. A 'z' is indicated the midline of the brain. Even numbers identify electrodes on the right hemisphere, whereas odd numbers identify those on the left hemisphere.

Because an EEG signal is a difference of electrical signals obtained from two electrodes, the

electrical signals are amplified using differential amplifiers. The EEG signal can be monitored in the various way according to a type of montages, the placement of the electrodes. Two popularized montages that are currently used are bipolar and referential montages. In the bipolar montage, a pair of adjacent electrodes are inputs to a differential amplifier resulting a waveform of each channel displayed on the monitor. The referential montage is a montage that the output of each channel is the voltage difference between a certain electrode and a common reference electrode. Generally, there is no standard position for the reference; however, the linked ears, referring to the positions A1 and A2, and midline positions are often used as a reference. When the common reference is an voltage averaged over the brain, the montage is called an average reference montage.

### 3.2 EEG characteristics

Since this proposal aims to detect ictal patterns in long EEG signals, it is important to understand normal behaviors of the EEG signals in order to comprehend the abnormal one. Clinically, neurologists use the knowledge of the normal activities to visually identify the epileptic seizures from the long EEG signals. There are four main rhythms of the normal EEG, namely alpha, beta, theta, and delta, that need to be primarily described [RT03]. Alpha rhythm occurs in a frequency range of 8–13 Hz. This rhythm is considered as the principal background of the normal EEG and discovered when the patient is relaxed, waking state, and eyes closed. It is usually maximum in the occipital area and spreads asymmetrically to the adjacent regions, *e.g.*, parietal and temporal regions. Beta rhythm (14–30 Hz or higher) appears with longer duration than muscle action potentials. Asymmetric amplitude between both sides of the brain commonly refers to the pathological hemisphere. Theta rhythm is defined as an activity in a frequency band of 4–7 Hz. It is typically dominant in the midline and the temporal region. This rhythm indicates a waking and drowsiness state and should be symmetrically diffused. If the theta activity appears only in one area or one hemisphere, this may refers to structural disease. Delta rhythm is a slow wave that its frequency distributes in 0.5–4 Hz. This wave usually has high amplitudes and reliably indicates localized brain diseases. An occurrence of this wave is also prominent to implications of cerebral dysfunction and sleeping in adults.

On the other hand, epileptiform patterns in EEG signals are abnormal patterns used to indicate epileptic seizures in the long EEG signals. By definition, the epileptiform patterns are *spikes* and *spike-wave complexes*; however, other abnormal patterns such as *sharp waves* are also practically significant to the detection of the epileptic seizures [BYL84]. The definition of the *spike* is an abrupt change of temporal potential from the background where its decline slope is lower than that of the incline. The spike duration ranges from 20–80 milliseconds and the spike is often followed by a slow wave with the duration of approximately 200 milliseconds. The *spike-wave complex*, also called a *spike-slow wave*, contains the spike and a following slow wave containing relatively high amplitudes. The spike-slow wave is in $3 \pm 0.5$ Hz and the amplitude of the spike is usually lower than that of the slow wave. The *sharp wave* is practically essential in determining the epileptic seizure even though it is not demonstrated as epileptic patterns. The sharp wave is defined as a wave with a frequency of 5–12.5 Hz. A sequence of spike, sharp, and spike-slow wave is referred to ictal patterns of EEG when seizures occur. By the morphology of these three patterns, *i.e.*, spikes, spike-slow waves, and sharp waves, changes in amplitudes, frequencies, and rhythms continuously happen relative to the background [PPCE92]. First, amplitudes of EEG signal during epileptic seizure activities tend to be higher than those of normal periods. Second, a frequency shift appears when brain activities transit from normal events, *e.g.*, drowsiness, eye blink, to the seizure activities. Third, rhythms or patterns in EEG signals change from normal activities to specific patterns. However, some change seems to be an occurrence of epileptic seizures even though this change is referred to an artifact. For instance, EEG signals interfered by main electricity have evolution of amplitudes from low to high and then still maintain the amplitudes at this level for a course of time. Moreover, periodic epileptiform discharges (PED) are also uncommon EEG characteristics similar to seizure activities but determined as non-seizure activities. This makes seizure detection challenging in discriminating the ictal patterns from EEG signals.

### 3.3 Convolutional neural network (CNN)

CNN is a type of neural networks that has been intensively and widely used in various applications: image processing, object detection, face recognition, natural language processing, and video processing [LBH15]. For example, VGG16net is a deep CNN that achieves top-5 accuracy in the ImageNet data set [SZ15]. The CNN is biologically inspired by the idea of animal vision that concentrates on a specific area of an image, called receptive field, instead of focusing on the whole image. The main advantages of this network are that it has spatial invariance property and less computational complexity because of the weight-sharing architecture of convolutional layers [ATY+19]. The CNN structure mainly consists of convolutional, activation, pooling, and fully connected layers stacked deeply. The computations of the convolutional, activation, and pooling layers are visualized in Figure 2. Some regularization technique such as dropout is also added to reduce the effect of an overfiting problem [SHK+14], and a batch normalization layer is used to enhance the learning speed [IS15].

The convolutional layer is a layer in which each neuron is locally connected to some area in the previous layer. This layer is mainly designed to extract and collect low-level and high-level features from each layer [ATY+19]. The result of each neuron is obtained by multiplying the local input by weights of filters. As shown in Figure 2a, the convolutional layer is a result of convolution of the input and the weights. The result can be visually interpreted as a feature map extracted on the receptive field. So, to extract many features simultaneously in the same layer, independent filters stacked in depth are used instead of only one filter.

The activation layer also called an activation map is a layer that visualizes activation nodes by using an activation function. The output of every node in the previous layer is independently passed to the activation function. Additionally, the activation function can also be physically interpreted as a function that activates and deactivates each neuron in the layer. An example of using activation functions transforming a feature map is illustrated in Figure 2b. Common activation functions are listed with their benefits and drawbacks as follow:

- *Identity function* is a function that the output and input are the same:

$$f(z) = z, \quad \frac{d}{dz} f(z) = 1. \tag{1}$$

  The identity function is put in the output layer when a regression problem is considered. However, it is well-known that the activation function in hidden layers should not be the identity function because if all activation functions are the identify function, the output is only a linear transformation of the input.

- *Sigmoid function* ($\sigma$), or logistic function, is a common activation function used in neural networks. The output of the function is known to be the conditional probability given the input or to be the smooth function of the step function:

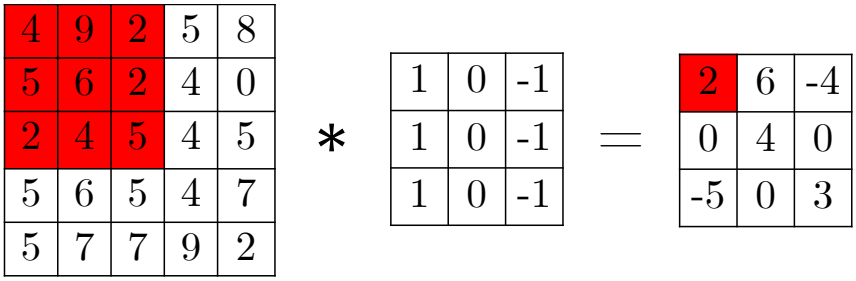
$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad \frac{d}{dz}\sigma(z) = \sigma(z)(1 - \sigma(z)). \tag{2}$$

  The advantages of this function are that it is differentiable at every point, bounded, and monotonic. However, when $z$ is largely positive and negative, the slope of the curve becomes to small, increasing training time; this problem is called a vanishing gradient. The sigmoid function also has a shift bias, causing the network to learn slow [XHL16].

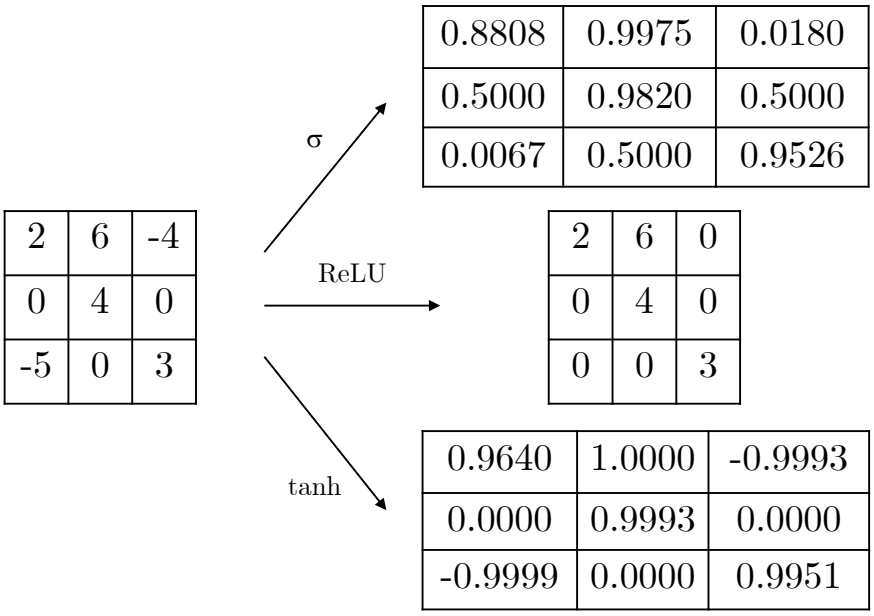
- *Hyperbolic tangent* (*tanh*) *function* is a function that is similar to the sigmoid function that it is bounded. Unlike the sigmoid function, the output of the tanh function is in the range of $(-1, 1)$:

$$\tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}, \quad \frac{d}{dz}\tanh(z) = 1 - \tanh^2(z). \tag{3}$$

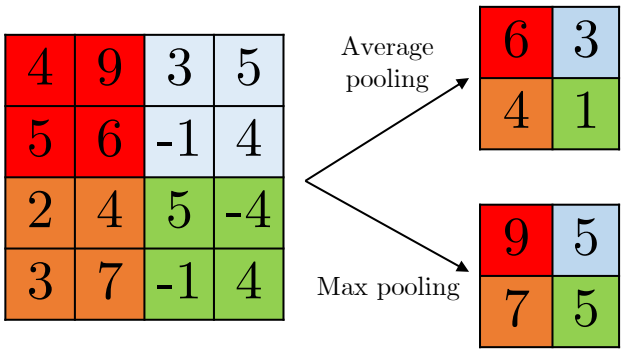
  The tanh function is used to overcome the shifted bias problem; however, the vanishing gradient problem still occurs.




(a) Convolutional layer.




(b) Activation layer.




(c) Pooling layer.

Figure 2: Computation of each layer in CNN.

- *Rectified linear unit* (*ReLU*) *function* is a piece-wise linear function that provides zero output when the input is negative, and passes the input to the output when the input is positive:

$$\mathrm{ReLU}(z) = \begin{cases} 0, & z \leq 0, \\ z, & z > 0. \end{cases}, \quad \frac{d}{dz}\mathrm{ReLU}(z) = \begin{cases} 0, & z < 0, \\ 1, & z > 0. \end{cases} \tag{4}$$

The main advantage of using the ReLU function is its computational efficiency for both forward and backward propagation [NH10]. Moreover, the ReLU function overcomes the vanishing gradient problem when $z$ is large since its derivative is always one. It has also been shown that, in practice, using the ReLU function provides greater convergence performance

than using the sigmoid function. However, the function is not always differentiable, and the network learning is prohibited when there are several dead neurons, the neurons that initially give zero outputs always provide zero outputs.

The pooling layer is a layer used extract some appropriate features from the previous layer. When an input is two-dimensional, an image for example, this can be interpreted as performing downsampling along the width and the height, the first and second dimensions, of the input. It can be intuitively considered as collecting useful information from the previous layer and filtering out some spatially unnecessary parts. Two common pooling strategies are max pooling and average pooling. As depicted in Figure 2c, the max pooling passes the highest value from the receptive field, while the average pooling does average the values in the window.

The batch normalization layer normalizes each input features independently at each mini-batch so that the mean of features is zero and the variance of features closes to one [IS15]. According to the ability to extract features in each layer, each neuron in the feature map possibly has different mean and variance. Moreover, the distribution of the activations is also changed during training since the weights are adapted continuously. This problem is called *Internal Covariate Shift* and it affects the learning speed. This layer is added to enhance the network to converge faster and prevent the network from the internal covariate shift. Considering a mini-batch $\mathcal{B} = \{x_1, x_2, \ldots, x_k\}$, the process of the batch normalization is demonstrated in Algorithm 1 where $\epsilon$ is a positive constant preventing numerical instability.

**Algorithm 1:** Batch normalization

**Input:** $x$ over a mini-batch: $\mathcal{B} = \{x_1, x_2, \ldots, x_k\}$
**Parameter:** $\gamma, \beta$
**Output:** $\{y_i = \gamma x_i + \beta\}$

1 $\mu_{\mathcal{B}} \leftarrow \frac{1}{k} \sum_{i=1}^{k} x_i$ // mean of mini-batch

2 $\sigma^2_{\mathcal{B}} \leftarrow \frac{1}{k} \sum_{i=1}^{k} (x_i - \mu_{\mathcal{B}})^2$ // variance of mini-batch

3 $\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma^2_{\mathcal{B}} + \epsilon}}$ // normalization

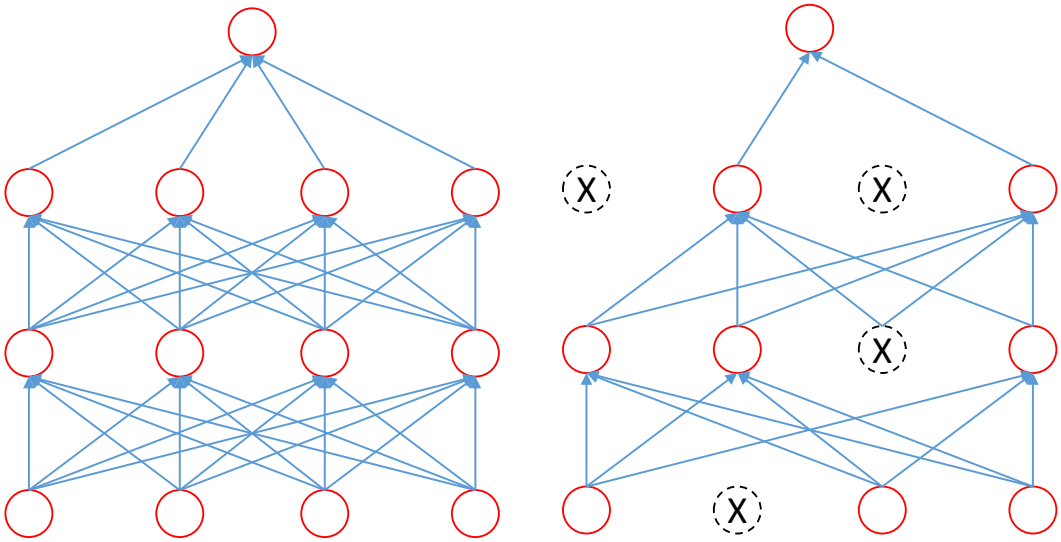
4 $y_i \leftarrow \gamma x_i + \beta$ // scale and shift

The dropout layer is added to randomly and temporarily removes some neurons in the input layer [SHK+14], as pictorially depicted in Figure 3. In Figure 3, the dropout technique is applied to both hidden layers to temporarily set to neurons to be inactive with a fixed probability. The dropout can be interpreted as a regularization technique for preventing the network from an overfitting problem The dead neurons in the layer are untrainable so the weights that need to be train are only the remaining connections. Furthermore, the dropout is also claimed to be superior over other regularization techniques [SHK+14].

The fully-connected layer is a layer containing neurons that are all connected to every neuron in the adjacent layers as visualized in Figure 4. This layer acting like a traditional multilayer perceptron that receives features as an input and produces a real value as an output. Each connection presents a weight that links two neurons. In deep learning, the fully-connected layer is usually added in the last layer because of its capability of classifying features from the input.

## 4 Literature review

From past literature, there have been a lot of researchers aiming to detect epileptic seizure activities in long EEG signals. Focusing on using scalp EEG signals, many studies mainly developed the automatic epileptic seizure detection based on epochs from the long EEG signals [SEC+04, SG10b, TYK16], while some research was designed to detect the seizure activities in the long EEG signals without any segmentation process [SLUC15]. Previously, the automatic detection of epileptic



(a) Standard neural network. (b) After employing dropout.

Figure 3: Dropout in neural network. By randomly dropping some neurons, the standard neural network is altered to a network containing less neurons. The neurons with a cross sign are temporarily removed from the network.

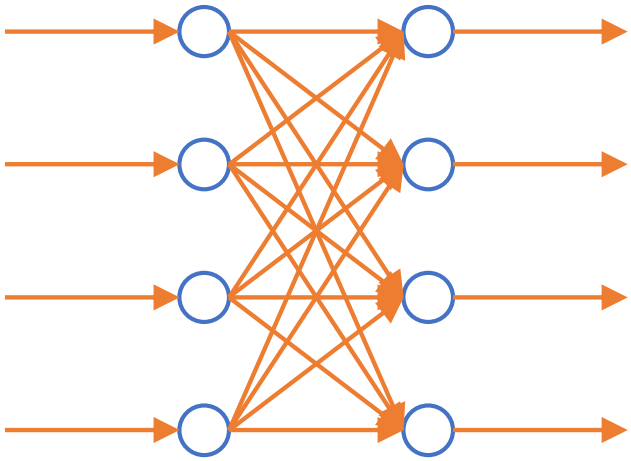


Figure 4: Illustration of a fully-connected layer.

seizures normally contained processes of signal transformation or decomposition, feature extraction, and classification. Sometimes, artifact or noise rejection was also optionally added at the beginning of the detection process [AKS18]. In addition, a channel selection technique was considered when multi-channel EEG signals were used [AESAA15], and feature dimension reduction or selection was taken into account when inputs have a considerably large magnitude [AWG06].

In our opinion, three aspects: characterization of the seizure activities via feature extraction, methods of the automatic detection, and the determination of the onsets and offsets, are fundamental to the automatic detection of epileptic seizure onset and offset. Therefore, we review features commonly used in the automated epileptic seizure detection in Section 4.1. Section 4.2 describes methods of automatic epileptic seizure detection using scalp EEG signals. However, there were only a few developments in identifying seizure onset and offset, as opposed to determining seizure occurrences. So all of these studies are summarized intensely in Section 4.3.

### 4.1 Feature extraction

Features are observable quantities used to determine characteristics or properties of events. In a classification problem, features should be chosen appropriately to be distinguishable between classes. Many features have been employed to discriminate ictal patterns from normal activities in EEG [ASS+13, ASSK16, BLuCS19b]. These features were categorized according to the purpose of the work. Some studies employed a group of features according to their meanings and interpretations [Got82, GRD+10, OLC+09], while others used features according to the domain from which they were extracted [TTM+11a, ASSK16]. For example, entropy-based features were applied to measure the fluctuation of the signal [AMS+12, AFS+15, LYLO14, TYK16]. Using amplitude-related features including nonlinear energy [AG99] and variance has shown a significant performance of detecting seizure activities with high amplitudes [Sho09, CODL15, SLUC15]. Different responses of features are demonstrated in Figure 5. On the other hand, features were also categorized into time, frequency, and time-frequency-domain features. Time-domain features were computed on raw or decomposed signals, intrincsic mode functions (IMFs) from EMD for example, in time domain, whereas frequency-domain features were calculated discrete-Fourier transform (DFT) or power spectral density (PSD) coefficients of raw EEG signals. On the other hand, time-frequency-domain attributes were obtained from transformed EEG signals containing both time and frequency information. For example, coefficients of short-time Fourier transform (STFT) or discrete-wavelet transform (DWT) were used in feature extraction. From our experimental results in [BLuCS19b], statistical parameters, energy and entropies were common features in those three domains to capture information about distributions, amplitudes, and uncertainties. It was concluded that statistical parameters such as mean, variance, skewness, and kurtosis were always applied jointly. Features, including the energy and entropies, relevant to amplitude and uncertainty were sometimes used independently. It was evident that the energy was the most promising feature to capture changes of amplitude in EEG signals. Eventually, the experiments conducted in [BLuCS19b] showed that variance and energy calculated from the DWT coefficients were recommended as features based on the Bayesian method and correlation-based feature selection (CFS) [HS97].

### 4.2 Automated epileptic seizure detection

In this section, we discuss applications of the automatic detection of epileptic seizure using the CHB-MIT Scalp EEG database since the data in this database are multi-channel scalp EEG signals. As previously mentioned above, there have been several studies focusing on the developments of the automatic epileptic seizure detection. Tables 1 and 2 summarize the performances of methods using features extracted from a specific domain and multiple domains, respectively.

There were many studies using single-domain features to detect seizures in EEG signals. Some works aimed to use only a single feature to detect seizures. Raw EEG signals were purely used as inputs of an artificial neural network (ANN) [CCS+18]. It was reported that this method accomplished 100% accuracy. However, the data were specified to contain simple and complex partial epileptic seizures in the frontal area collected from only female subjects. Amplitude-integrated EEG (aEEG) was exploited to identify occurrences of high-amplitude seizures [SLUC15]. By using an adaptive thresholding method, the method obtained the sensitivity of 88.50% and false positive rate per hour (FPR/h) of 0.18. Nevertheless, this method also responded to artifacts with high amplitudes and required EEG signal that began with normal activities. An energy computed in frequency domain using filter bank analysis and a radial basis function (RBF) SVM were jointly employed to characterize the epileptic seizures. As a result, the energies from seizure samples were higher than that of the normal ones. Moreover, the logarithm of variance of DWT coefficients in each sub-band from a selected channel was used to determine a seizure epoch with a thresholding [Jan17a]. According to the best result of each patient, the method obtained the average performances of 93.24% accuracy, 83.34% sensitivity, and 95.53% specificity. Similarly, the author also conducted an experiment using a smaller data set, including only 12 subjects. The results showed that using those features with SVM outperformed a feature combination of line length, nonlinear energy (NE), variance, power, and maximum value of raw EEG signals with the average

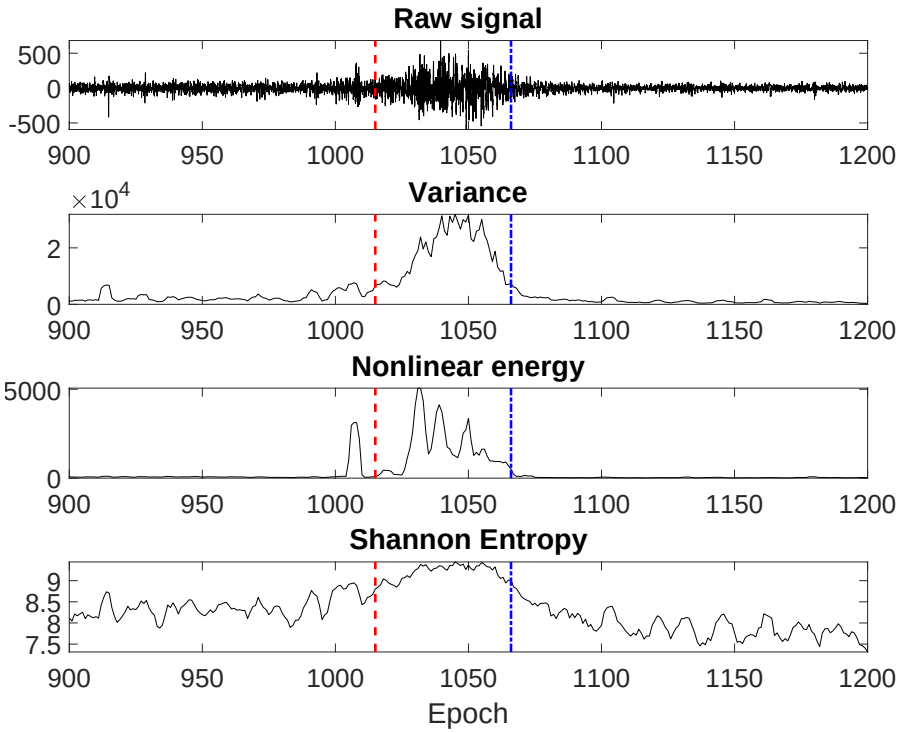



Figure 5: Features responding to changes in EEG signals. Each feature is calculated from 4-second EEG epochs and the sliding window is one second. This displayed signal is collected from the record `chb01_16` in the CHB-MIT Scalp EEG database [GAG+00] on the channel *FP1-F7*. Dash line indicates the seizure onset and dashdotted line shows the seizure offset.

accuracy, sensitivity, and specificity of 96.87%, 72.99%, and 98.13%, respectively. Furthermore, the STFT spectrogram was used with a modified stacked sparse denoising autoencoder (mSSDA) to detect an epileptic seizure in individual epochs [YXJZ17]. It reported that this method outperformed the other methods conducted in the experiment and obtained the accuracy of 93.82%.

On the other hand, a combination of features in a single domain was proposed to capture ictal patterns in many aspects. Fractal dimension called a box-counting dimension ($D_B$) and energy were exploited to observe complexity and amplitude of the EEG signal [VI17]. The records included in [VI17] were chosen to have the same bipolar montage, and the subject `chb16` was excluded because of the short seizure duration. Eventually, the authors showed that using relevant vector machine (RVM) with these features computed on harmonic wavelet packet transform (HWPT) coefficients potentially achieved the sensitivity of 97.00% and FPR/h of 0.10. Mean, ratio of variance, standard deviation (SD), skewness, kurtosis, mean frequency, and peak frequency were extracted from DWT coefficients [AS16]. An extreme learning machine (ELM) was employed to classify EEG epochs into a specific class. Due to its effectiveness and efficiency, this combination could accomplish the accuracy of 94.83%. The work in [AKS18] compared the detection performance of using different transformations and different classifiers via the accuracy (Acc). First, multi-channel EEG signals were filtered by multi-scale principal component analysis (MSPCA) to remove artifacts. Then, the features –absolute mean value, average power, SD, ratio of absolute mean values, skewness, and kurtosis– computed on decomposed signals by EMD, DWT and wavelet packet decomposition (WPD) were applied to many classifier: random forest (RF), SVM, ANN, and k-NN. Finally, it was concluded that the methods using DWT and WPD obtain 100% accuracy. However, only 2,000 eight-second EEG epochs, 1,000 samples for each group, were selected.

Moreover, several features in many domains were also exploited to obtain information in different domains. The work in [FHH+16] employed many classifiers: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), polynomial classifier, logistic regression, k-nearest neighbor (k-NN), decision tree, Parzen classifier, and support vector machine (SVM) with the same features.

Table 1: Summary of automated epileptic seizure detection using the *CHB-MIT* Scalp EEG database when *single-domain* features were used.

| Domain | Features | Method | Performance | Ref. |
|---|---|---|---|---|
| Time | Raw signal | ANN | Acc = 100% | [CCS+18] |
| | aEEG | Thresholding | Sen = 88.50%, FPR/h = 0.18 | [SLUC15] |
| | Line length, NE, variance, average power, max | RBF SVM | Acc = 95.17%, Sen = 66.35%, Spec = 96.91% | [Jan17a] |
| | Absolute mean values, average power, SD, ratio of absolute mean values, skewness, kurtosis | MSPCA + EMD + RF | Acc = 96.90% | [AKS18] |
| | | MSPCA + EMD + SVM | Acc = 97.50% | [AKS18] |
| | | MSPCA + EMD + ANN | Acc = 96.90% | [AKS18] |
| | | MSPCA + EMD + k-NN | Acc = 94.90% | [AKS18] |
| | $D_B$[4] | RVM | Sen = 97.00%, FPR/h = 0.24 | [VI17] |
| Frequency | Energy | RBF SVM | Sen = 96.00%, FPR/h = 0.08 | [SG10a]* |
| Time-frequency | Spectrogram | STFT + mSSDA | Acc = 93.82% | [YXJZ17] |
| | Mean, ratio of variance, SD, skewness, kurtosis, mean frequency, peak frequency | DWT + ELM | Acc = 94.83% | [AS16] |
| | Log of variance | DWT + thresholding | Acc = 93.24%, Sen = 83.34%, Spec = 93.53% | [Jan17b] |
| | | DWT + RBF SVM | Acc = 96.87%, Sen = 72.99%, Spec = 98.13% | [Jan17a]* |
| | Absolute mean, average power, SD, ratio of absolute mean, skewness, kurtosis | MSPCA[1] + DWT + RF | Acc = 100% | [AKS18] |
| | | MSPCA + DWT + SVM | Acc = 100% | [AKS18] |
| | | MSPCA + DWT + ANN | Acc = 100% | [AKS18] |
| | | MSPCA + DWT + k-NN | Acc = 100% | [AKS18] |
| | | MSPCA + WPD[2] + RF | Acc = 100% | [AKS18] |
| | | MSPCA + WPD + SVM | Acc = 100% | [AKS18] |
| | | MSPCA + WPD + ANN | Acc = 100% | [AKS18] |
| | | MSPCA + WPD + k-NN | Acc = 100% | [AKS18] |
| | Energy | HWPT[3] + RVM | Sen = 97.00%, FPR/h = 0.25 | [VI17] |
| | Energy, $D_B$ | HWPT + RVM | Sen = 97.00%, FPR/h = 0.10 | [VI17] |

Acc = accuracy, Sen = sensitivity, Spec = specificity, FPR/h = false positive rate per hour
[1] Multi-scale principal component analysis, [2] wavelet packet decomposition, [3] harmonic wavelet packet transform, [4] box-counting dimension
* Use all data records

Table 2: Summary of automated epileptic seizure detection using the *CHB-MIT* Scalp EEG database when *multi-domain* features were used.

| Time | Frequency | Time-frequency | Method | Performance | Ref. |
|---|---|---|---|---|---|
| Variance, RMS, skewness, kurtosis, SampEn | Peak frequency, median frequency | | LDA | Sen = 70.00%, Spec = 83.00% | [FHH+16] |
| | | | QDA | Sen = 65.00%, Spec = 92.00% | [FHH+16] |
| | | | Polynomial classifier | Sen = 70.00%, Spec = 83.00% | [FHH+16] |
| | | | Logistic regression | Sen = 79.00%, Spec = 86.00% | [FHH+16] |
| | | | k-NN | Sen = 84.00%, Spec = 85.00% | [FHH+16] |
| | | | Decision tree | Sen = 78.00%, Spec = 80.00% | [FHH+16] |
| | | | Parzen classifier | Sen = 61.00%, Spec = 86.00% | [FHH+16] |
| | | | SVM | Sen = 79.00%, Spec = 86.00% | [FHH+16] |

Variance, root mean squared value (RMS), skewness, kurtosis, and sample entropy (SampEn) were used as time-domain features, and peak frequency and median frequency computed from PSD were exploited to extract information in frequency domain. Combined with a feature selection call LDA with a backward search, the k-NN outperformed the other classifier with the sensitivity of 84.00% and specificity of 85.00%. However, the authors chose only records that contained seizures activities in this study.

### 4.3 Applications of Seizure onset and offset detection

There have been only a few attempts that aim to develop seizure onset and offset detection. One of the first automated seizure offset detection was designed by Shoeb et al. [SKS+11]. The researchers proposed both patient specific and non-specific algorithms using multi-channel scalp EEG signals. Long EEG signals of patients in the CHB-MIT Scalp EEG database were analyzed by segmenting the signals into five second epochs and advancing each epoch by one second. Both patient specific and non-specific methods used signal energies of 25 contiguous frequency bands spanning 0–25 Hz from each channel independently to observe spatial and spectral properties in the epoch. In the patient non-specific setting, a feature vector was constructed from the signal energy averaged over channels of the frequency bands. For the patient-specific case, each feature was a weighted average of the energy of each frequency band over all channels. The weights were calculated based on the

differences between the signal energies in ictal and postictal states. Each feature vector was then fed to SVM to classify the epoch as ictal or postictal. A linear SVM was used in the patient-specific case whereas a radial basis function SVM (RBF SVM) was exploited in the other case. Once the seizure onset had been recognized by the algorithm from their previous study [SG10a], the end of seizure was declared when five consecutive epochs were recognized as postictal. It was reported that the patient non-specific method was able to detect all seizure ends with an average accuracy of 84% and an average absolute offset latency of $8.9 \pm 2.3$ seconds while the patient-specific algorithm detected 132 out of 133 seizure offsets with an accuracy of 90% and an averaged absolute latency of $10.3 \pm 5.5$ seconds over patients. However, seizures that slowly changed from the ictal to the postictal periods led to a large delay of seizure offset detection. In contrary, seizure ends were so early detected when the seizure activities were corrupted by artifacts. Additionally, this method requires an onset detection system to alarm the seizure onset first.

Orosco et al. [OCDL16] applied stationary wavelet transform (SWT)-based feature extraction in detecting seizures and their onset and offset. Eighteen subjects from the CHB-MIT Scalp EEG database were used to perform patient-specific and patient non-specific scenarios. Non-overlapping two second epochs were decomposed by SWT in each channel individually and coefficients of 4 sub-bands corresponding to normal EEG rhythms were used to extract features. In each channel, mean frequency and peak frequency were calculated on the power spectral density (PSD) of all selected sub-bands coefficients and a relative energy of each frequency band, an energy of each band normalized by the total energy, was extracted. The features were then spatially averaged over left anterior, right anterior, left posterior, right posterior, and central areas. By feature selection based on the statistical parameter called Lambda of Wilks, 26 features left were applied to LDA and artificial neural network (ANN). The results showed that, in the patient-specific case, LDA outperformed ANN with overall specificity of 99.99%, sensitivity of 92.6%, false positive rate per hour of 0.3, and onset and offset latencies of 0.2 and 4 seconds after and before the annotation. For the patient non-specific case, LDA also achieved 99.9% specificity, 87.5% sensitivity, 0.9 false positive rate per hour (FPR/h), and onset and offset latencies of 1.3 and 3.7 seconds respectively on average. In this paper, the positive latency was observed when the algorithm detected a seizure before an annotation. Nevertheless, ranges of seizure onset and offset were very wide in both patient specific and non-specific cases. Ranges of the seizure onset and offset in the patient-specific case were 42.4 and 84.4 seconds, while the ranges of the onset and offset in the other case were 248 and 81.3 seconds, respectively. Moreover, in [OCDL16], the sensitivity was calculated based on seizure events, while the specificity was an epoch-based metric. Due to high FPR/h obtained from each subject, it was possible that an small amount of epochs during seizure activities were detected so that the event-based sensitivity was that high.

Another approach focusing on the patient-specific detection of seizure onset and offset that used the CHB-MIT Scalp EEG database was found in [CUFK19]. EEG records from 18 patients were analyzed from a 1-second sliding window by exploiting an orthonormal triadic wavelet transform. Each EEG epoch was decomposed into specific frequency ranges using triadic wavelets. Statistics-based features were extracted each channel individually from selected frequency bands corresponding to normal EEG rhythms. Then the features of each channel were classified by LDA and k-nearest neighbor (k-NN) independently. Segments which were recognized as seizure for at least 6 channel were marked as 1 representing seizure EEG epochs. The results from the channel-based detection were post-processed by centered moving average (CMA) of length 15 to reduce a false alarm. Eventually, the output from CMA of each epoch was compared to a threshold of 0.4 to determine the final decision. The first epoch detected as seizure was determined as a seizure onset and a seizure end was observed when the final decision changed from 1 to 0, representing transition from a seizure stage to a normal stage. As a result, the method using k-NN achieved 99.62% accuracy, 98.36% sensitivity, 99.62% specificity, 0.80 FPR/h, 6.32 seconds for seizure onset lantacy, and $-1.17$ seconds for seizure offset latency respectively on average. On the other hand, averaged classification performance measurements evaluated by LDA were 98% accuracy, 100% sensitivity, 98.05% specificity, 4.02 FPR/h, and 1.41 and 8.19 second onset and offset latencies, respectively. This study denoted the positive latency as a time delay that a predicted time point

was after an actual time point. However, these methods were also not robust across patients; a seizure offset of some patients was announced 20 seconds after the annotation whereas a seizure end of other patients was detect 20 seconds prematurely. Furthermore, the 100% sensitivity was accomplished when the FPR/h was extremely high. Specifically, the FPR/h of some subjects were higher than 10, meaning that there were repeated false alarms about every six minutes.

In addition, Correa et al. [CODL15] used the iEEG databased recorded at the Epilepsy Center of the University Hospital of Freiburg. The data set contained 196 one-hour six-channel iEEG segments from 21 patients, where 89 records contained seizure events. Every record with seizures had only one seizure activity. In the pre-processing, the authors applied a bi-directional Butterworth second-order filter with the frequency range of 0.5–60 Hz the useful information for detecting the epileptic seizure contained in the frequency range [GG05]. In each window and each channel, the PSD was calculated from one-second window and 0.5-second overlapping. Subsequently, the authors computed the relative powers of the normal bands (theta, alpha, and beta), and applied a median filter with a window of 30 seconds to smooth the sequences of the relative powers. The derivative of each sequence was then computed by the difference quotient to observe changes in the sequence. Finally, the final sequence was obtained by averaging the derivative sequences of every channel and every frequency band followed by the median filter. An iEEG segment was declared as it contained a seizure when the amplitude of the final sequence was three times higher than the average power of the final sequence, and the exceeding period was longer than 30 seconds. When the seizure event was detected, a discrete-wavelet transform (DWT) was exploited to decompose the iEEG into five sub-bands. Windows of 30 seconds before and after the detected seizure event were considered to determine the onset and offset. Energies computed from the detail coefficients of levels 3, 4, and 5 of each channel were used to detect the onset and offset. The 18 sequences of the energies (from three sub-bands and six channels) were filtered using the median filter. The onset and offset from each sequence were determined from the first and last points that the sequence was two times above its median. Eventually, the final onset and offset were obtained by averaging 18 onsets and offsets. As a result, average event-based sensitivity and specificity were 85.39% and 83.17%, respectively. Onset and offset latencies reported from each subject and each segment were mostly less than 30 seconds. However, this work is not practical in clinic because of the data. It is possible that there are more than one seizure, as in the CHB-MIT Scalp EEG database, in a one-hour record. Furthermore, this method is heuristic; there are needs for parameter settings from experts since many types of seizures may occur in one patient. Even though the authors also reported epilepsy types and showed that there were a small amount of detection error, the data from each subject was too small to conclude that it was practical.

# 5 Problem statement

The problem of epileptic seizure detection and seizure onset-offset determination can be divided into two crucial steps in sequential order: epoch-based seizure detection and onset-offset identification, as shown in Figure 6. In the process of the seizure detection, a seizure detector receives inputs as information and produces the probability of a seizure occurrence as the output. A multi-channel EEG epoch windowed from a long multi-channel EEG signal is considered as a sample, and the output is the probability that a seizure occurs in the epoch. When all EEG epochs from the long EEG signal are applied to the seizure detection algorithm, the output is the sequence of seizure probabilities of individual epochs. Subsequently, the probability sequence is fed to the onset-offset detector to indicate the seizure onset and offset of each individual seizure in the long EEG signal.

## 5.1 Classifier

For a binary classification problem, let $\mathcal{D} = \mathcal{X} \times \mathcal{Y}$ be a space of pairs $(x_i, y_i)$ where $\mathcal{X}$ and $\mathcal{Y} = \{0, 1\}$ are vector spaces of all inputs and outputs, respectively. Formally, in the probabilistic view point, there is a joint probability distribution $f_{xy}(x, y)$ over $\mathcal{D}$, and $(x_i, y_i)$ is drawn from the distribution $f_{xy}$. In machine learning, there exists an actual function that maps every input sample $x_i \in \mathcal{X}$

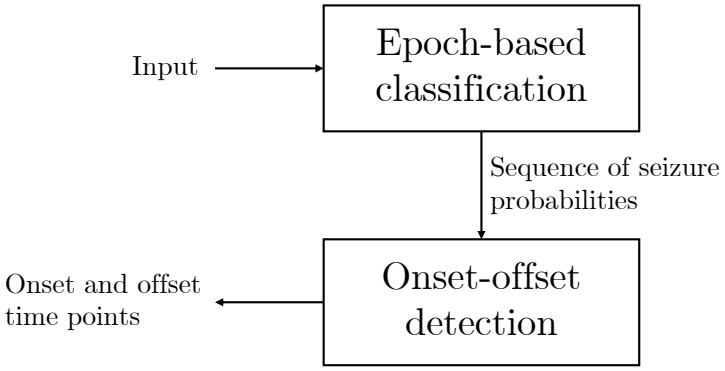



Figure 6: Scheme of the problem containing two statements: epoch-based seizure detection and onset-offset detection.

to its label $y_i \in \mathcal{Y}$. So, the major goal is to find a mapping function called a classifier $h$, also called a hypothesis or a learner, in a hypothesis space $\mathcal{H}$ that approximately behaves like the actual function: $h(x_i) \approx y_i, \forall (x_i, y_i) \in \mathcal{D}$ [FHT01].

A loss function $L$ is a non-negative-valued function that is used to observe how accurate the classifier is from the difference between the predicted and the actual values. For instance, a 0-1 loss function, which disregards a correct classification but absolutely focuses on an incorrect result, is defined as

$$L(h(x_i), y_i) = \begin{cases} 0, & h(x_i) = y_i, \\ 1, & \text{otherwise.} \end{cases} \tag{5}$$

The *true error*, also called the *expected risk* and the *Bayes risk*, is defined as the expected value of the loss function to measure the overall error of the results from the classifiers:

$$R_{\text{true}}(h) = \mathbf{E}[L(h(x), y)]. \tag{6}$$

Since $\mathcal{Y}$ contains only discrete elements, the *true error* is

$$R_{\text{true}}(h) = \int \sum_{y \in \mathcal{Y}} f_{xy}(x, y) L(h(x), y) dx. \tag{7}$$

The main problem is to find the optimal learner $h^*$ in the hypothesis space $\mathcal{H}$ such that it minimizes $R_{\text{true}}(h)$:

$$h^* = \underset{h \in \mathcal{H}}{\operatorname{argmin}} R_{\text{true}}(h). \tag{8}$$

The optimal hypothesis $h^*$ is formally called the *Bayes optimal classifier*, and the minimum error $R_{\text{true}}(h^*)$ is named as the *Bayes error rate*. In addition, it is well-known that, by exploiting the Bayes' theorem, the best decision for the 0-1 loss function is made from the class of which the posterior probability is highest, meaning that

$$h^*(x) = \begin{cases} 1, & P(y = 1|x) > P(y = 0|x), \\ 0, & P(y = 1|x) < P(y = 0|x). \end{cases} \tag{9}$$

Note that $P(y = 1|x_i) = 1 - P(y = 0|x_i)$. However, $R_{\text{true}}(h)$ cannot be directly obtained from (7) and it cannot be minimized since $f_{xy}(x, y)$ is practically unknown. Hence, the *empirical error* as the measure of the *true risk* using data in $\mathcal{D}$ is employed as the estimation of $R_{\text{true}}(h)$:

$$R_{\text{emp}}(h) = \sum_{(x,y) \in \mathcal{D}} P(x, y) L(h(x), y), \tag{10}$$

where $P(x, y)$ is the hypothetical joint probability. Nevertheless, $P(x, y)$ is also generally unknown. It is, therefore, assumed to $1/|\mathcal{D}|$ where $|\mathcal{D}|$ is the number of samples in set $\mathcal{D}$:

$$R_{\text{emp}}(h) = \frac{1}{|D|} \sum_{(x,y) \in \mathcal{D}} L(h(x), y). \tag{11}$$

The optimal learner $h^*$ is, therefore, obtained by minimizing $R_{\text{emp}}(h)$:

$$h^* = \underset{h \in \mathcal{H}}{\operatorname{argmin}} R_{\text{emp}}(h). \tag{12}$$

In what follows, we omit to use $^*$ for more convenience of comprehension.

In the binary classification, there are many classifiers to mimic the actual function and loss functions to evaluate the classifiers. Several hypothesis spaces are also employed in the binary classification problem. For example, a linear classifier is the simplest classifier that works well when input features are linearly separable. A polynomial classifier is extended from the linear classifier to classify samples with a nonlinear decision boundary. SVM is commonly used to classify samples into groups using separating hyperplanes. Moreover, by exploiting a kernel function, RBF for example, the SVM can be improved to classify data that are not linearly separable. Recently, the use of neural networks has been popularized, especially in classification problems, because of their ability to universally approximate any real-valued function [HSW89]. Furthermore, approaches of deep learning are also widely developed and explored since deep learning networks are able to extract low-level and high-level features by themselves [LBH15]. In this research, we mainly concentrate on designing deep learning models to differentiate ictal patterns from raw EEG signals.

Moreover, many loss functions have been used to suit specific purposes. For instance, a binary cross entropy is one of the most popular loss function for this problem. The binary cross entropy is the measure of dissimilarity between the probability distributions of the label and the predicted output. Simply derived from the log of the probability mass function of a Bernoulli distribution, it is defined as

$$L(h(x), y) = -y \log f(x) - (1 - y) \log(1 - f(x)), \tag{13}$$

where $f(x)$ is the output of the model, which is the probability that $y = 1$ given the input $x$ in the case of neural networks for example, and $h(x) = \Theta(f(x) - 0.5)$ where $\Theta(x)$ is the Heaviside step function. In other words, $h(x) = 1$ when $f(x) > 0.5$ and $h(x) = 0$ when $f(x) < 0.5$. The square loss function finding the difference between $y$ and $h(x)$ is also used in the classification and regression problems. It is defined as

$$L(h(x), y) = (y - h(x))^2 \tag{14}$$

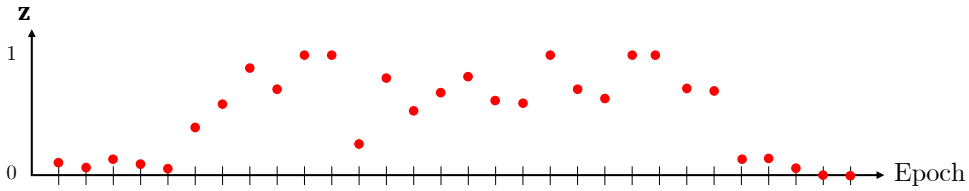
This study does not specifically decide yet which loss function is mainly used because we will see in Section 9 that a common loss function like the binary cross entropy is unsuitable when the data are extremely imbalanced.

### 5.2 Onset-offset detector

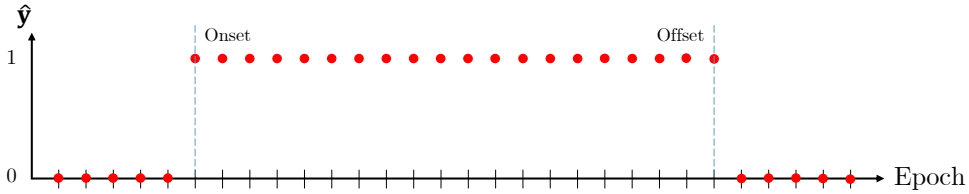
Seizure onset is a time point at which the seizure begins and seizure offset is time when the seizure terminates. A seizure onset-offset detection is the process of determining the beginning and the ending of a seizure. Therefore, the main purpose of this process is to imply when the seizure starts and ends in a long EEG signal from all detection outputs from the classifier. Since an epileptic seizure activity should appear with some period, a classification result of a single EEG epoch cannot, however, sufficiently imply an occurrence of the seizure. In fact, it requires a sequence of classification results from adjacent, both before and after, epochs in identifying the seizure event. Therefore, the sequence of classification results is required to determine the seizure onset and offset.

Suppose that $z_i = h(x_i)$ is the result of classification when $x_i$ is the input. We denote $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$ and $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ as the vectors of predicted class, and classification output of all sequential epochs, respectively, where each element refers to the result of each epoch and $n$ is the number of epochs in the long EEG signal as visualized in Figure 7. This process initially uses a function denoted as $g : [0, 1]^n \to \{0, 1\}^n$ to modify the vector of classification output $\mathbf{z}$, depicted in Figure 7a, to obtain the new classification vector $\hat{\mathbf{y}}$, shown in Figure 7b, that is then used to determine the seizure onset and offset:

$$\hat{\mathbf{y}} = g(\mathbf{z}) \tag{15}$$




(a) Output from the epoch-based seizure detection.




(b) Output of the onset-offset detection.

Figure 7: Illustration of determining the onset and offset. A onset-offset detector is a function $g$ that transforms $\mathbf{z}$ to $\hat{\mathbf{y}}$.

The seizure onset is determined from the time value of index $k$ for which $\hat{y}_k = 1$ (referred to ictal) and $\hat{y}_{k-1} = 0$ (referred to normal). Similarly, the index $k$ implies the seizure offset when $\hat{y}_k = 1$ and $\hat{y}_{k+1} = 0$.

# 6 Research methodology

This section explains the study plan depicted in Table 3 and the methodology of this research by items as follows:

Table 3: Study plan.

| Item | Semester | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Review literature | ■ | ■ | ■ | | | | | |
| Collect online data | | ■ | ■ | | | | | |
| Write and submit a review journal | | | ■ | ■ | | | | |
| Propose and verify method for the proposal | | | ■ | ■ | | | | |
| Prepare proposal examination | | | | ■ | ■ | | | |
| Present method to detect seizure onset and offset | | | | | ■ | ■ | ■ | |
| Study abroad | | | | | | ■ | | |
| Conclude the thesis and prepare the examination | | | | | | | | ■ |

- Review literature on data collection, pre-processing, feature extraction, classification, and process of determining the seizure onset and offset in EEG signals.
- Propose a method to detect seizure activities based on each epoch using a machine learning tool and present a technique to indicate the seizure onset and offset.
- Collect data from several subjects where each subject has many records. There must be at least one record of each subject containing at least one seizure activity.
- Train a classifier on EEG segments where the training set must contain at least one seizure event. Verify results from classification on a test set collected from the same patient.

- Apply an onset-offset detector model to the classification results of EEG epochs to determine starting and ending points of seizure events. Compare the results of the onset-offset detection and the classification results by using the same metrics and the same practically reasonable conditions.
- Conclude the detection performances, limitations, and future work.

# 7 Proposed method

This section discusses the proposed method for the automated epileptic seizure detection, and the seizure onset and offset determination. The entire process consisted of 3 steps, including epoch-based seizure detection, onset-offset determination, and evaluation, as illustrated in Figure 8. As explained above, the classifier was used to determine an occurrence of seizure in each small epoch from a long EEG signal. In this proposal, a deep CNN was developed as the epoch-based seizure detector and the raw EEG segment was considered as an input to the model. Subsequently, results of the CNN model were applied to the onset-offset detector to identify the seizure onset and offset. The outcomes of the onset-offset detector were then compared to the results of the CNN model. The comparisons were assessed by common types of metrics: epoch-based metrics, event-based metrics, and the onset and offset latency.

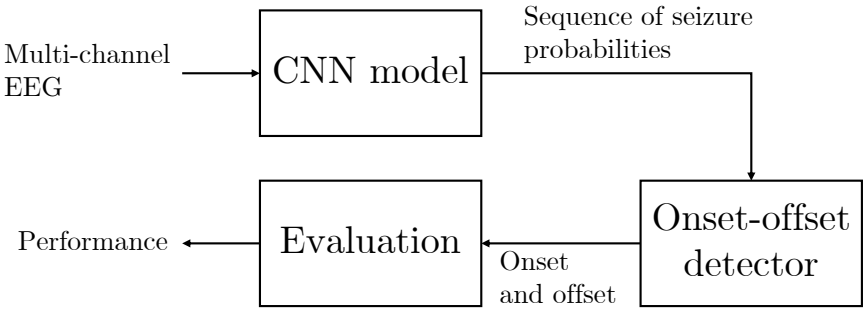
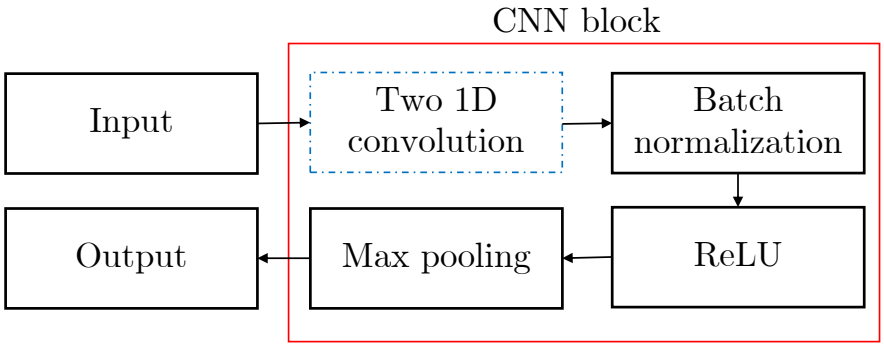



Figure 8: Scheme of the proposed method consisting of three steps: epoch-based classification, onset-offset detection, and evaluation.

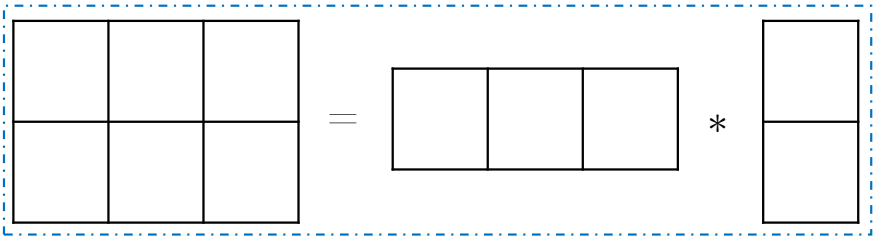
## 7.1 Classification

We employed a deep CNN model to extract features instead of handful-engineering features, and to classify a raw EEG epoch. Figure 9 illustrates a design of CNN block. The deep CNN model contained blocks of layers including convolutional, normalization, activation, and max pooling layers as shown in Figure 9a. Every block had the same sequence of layers but hyperparameters of some layer were changed to serve a physical meaning. For example, some block had a one-dimensional max-pooling layer to down sampling feature maps in the temporal domain only, whereas a two-dimensional max-pooling layer was used to reduce the dimensions temporally and spatially.

In the design of the convolutional layer to suit this problem, the size of EEG epoch was taken into consideration. Suppose that a raw EEG epoch was expressed as a matrix of size $m \times N$, where $m$ is a number of channels, $N$ is a number of temporal samples in the epoch, and, practically, $m \ll N$. So, in this problem, the convolutional layer was designed to capture temporal information, EEG pattern, rather than spatial characteristics and dispersion of electric field. Therefore, the width of the filter was larger than its height. Moreover, we exploited the concept of filter decomposition to reduce a model complexity and to overcome an overfitting problem [SVI+16]. A two-dimensional filter was decomposed into two one-dimensional filters as shown in Figure 9b. The first filter in Figure 9b could be physically interpreted as a feature extractor in temporal domain, and the other was to find a relationship of a feature between channels. Next, a batch normalization layer was added to reduce an *internal covariate shift* [IS15]. Following the normalization layer, the ReLU




(a) Block of CNN containing convolutional, one batch normalization, one activation, and one pooling layers.




(b) Example of filter factorization from a three-by-two filter into three-by-one and one-by-two filters. The first filter aims to extract a temporal feature and the second filter indicates a spatial relationship.

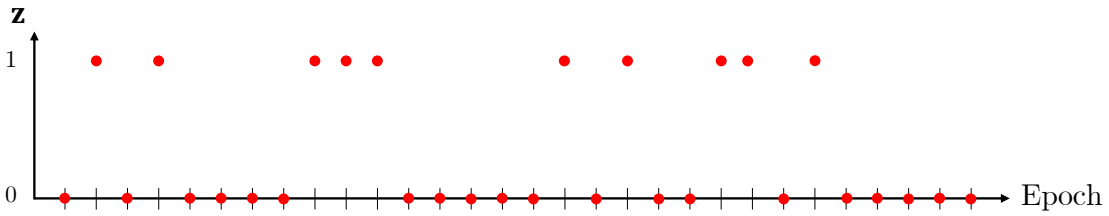
Figure 9: Design of CNN block. In the blue box, the two-dimensional filter is factorized into two one-dimensional filters.

function was used as an activation function to fasten the learning procedure [NH10]. Subsequently, a max-pooling layer was used to draw the most active values of features. The number of blocks was set to appropriately extract high-level useful features. Finally, dropout layer were applied to reduce overfitting problems, and fully-connected layers were exploited in the last layers to classify each EEG epoch into a specific class (normal/seizure).

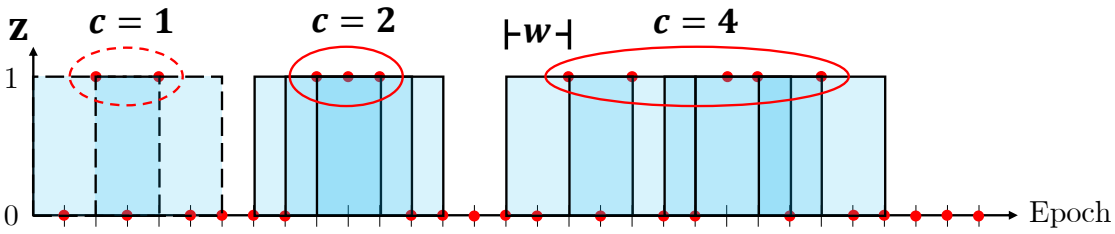
### 7.2 Seizure onset-offset determination

The method in the indication of seizure onset and offset is important and can further improve the classification performances. As mentioned, the seizure onset is a time point at which the seizure begins and the seizure offset is time when the seizure terminates. However, it is not practical to ingenuously use the above statements in the epoch-based detection. Particularly, the seizure activities do not occur for only a few seconds and then suddenly vanish [RT03]. This means that classifying each epoch independently as epileptic or normal is unreasonable since consecutive epochs are dependent. We will show in Section 9 that detecting each epoch separately can unfortunately produce considerable false positive rates and numerous declarations of seizure onset and offset. Moreover, it is practical to combine some close adjacent epileptic seizure events into one and ignore the gap of normal activity between them. In this case, the seizure onset and offset are reported only once. To certainly handle the above issues, we simply used a criteria-based method to modify the epoch-based classification outputs so that the final result is practically more reasonable.

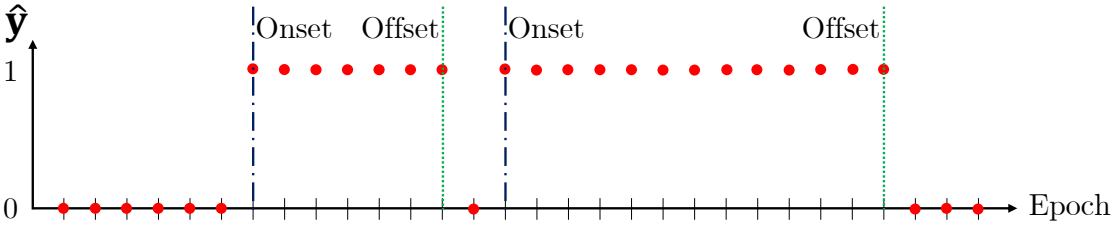
In this step, the outputs from classification are processed to identify the seizure onset and offset if available in the long EEG signal. Figure 10 illustrates an example of the onset-offset detection. Consider the sequence of epochs that are obtained from the classification step as shown in Figure 10a. All epochs that had been predicted as epileptic, denoted as $z = 1$, are covered with a rectangular window of size $2l + 1$ where the epoch is located at the window center visually interpreted in Figure 10b. All epochs in those windows are pre-labeled as epileptic and the window is named a seizure window. Subsequently, if there are at least $p$ consecutive overlaps or contacts from adjacent seizure windows, these seizure windows are finally declared as seizure. On the other




(a) Output of the CNN model.




(b) Rectangular windows covering seizure epochs.




(c) Results of the onset-offset detector. The onset and offset are the first and last epoch of the predicted event.

Figure 10: Example of the seizure onset-offset detection process where $l = 2$ and $p = 3$. The solid windows are finally treated as ictal ($\hat{y} = 1$), and the dashed windows are regarded as normal ($\hat{y}$=0).

hand, the other seizure windows that consecutively have overlaps or contacts less than $p$ epochs are eventually labeled as normal. In other words, if there exists at least $p$ consecutive seizure epochs that any two adjacent epochs are apart less than $2l$ epochs, all epochs between those epochs including $l$ epochs before the first seizure epoch and $l$ epochs after the last seizure epoch are combined to be a seizure event. The other epochs that do not meet this condition are treated as normal ($\hat{y} = 0$). Finally, the seizure onset is declared as the first epoch of the predicted event, and the seizure offset is determined from the last epoch of the event. The outcome of the onset-offset detector is displayed in Figure 10c.

### 7.3 Evaluation

In the problem of binary classification, detection performances are calculated from a confusion matrix containing the numbers of true positive (TP), false positive (FP), false negative (FN), and true negative (FN). With these values, many metrics are established for specific purposes. For example, common metrics such as accuracy (Acc), sensitivity (Sen), and specificity (Spec) are defined as

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\%, \tag{16}$$

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \tag{17}$$

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%. \tag{18}$$

The accuracy is used to indicate the overall performance of the classification, while the sensitivity and specificity are indicators determining the performance of correctly classifying outputs as ictal

and normal, respectively. Moreover, $F_1$, also known as F-measure is the measure of classification performance that takes an imbalance of the data into account [Pow11]. It is calculated from a harmonic mean of precision, or positive predictive value, and recall, or sensitivity. In other words, $F_1$ can also be calculated as follows:

$$F_1 = \frac{2TP}{2TP + FN + FP} \times 100\%. \tag{19}$$

Recently, Two groups of metrics, namely epoch-based and event-based metrics, have been used in evaluating the automatic epileptic seizure detection [TTM+11b]. Moreover, a latency, a time delay between the predicted and actual time points, is normally applied as a time-based indicator.

Epoch-based metrics are used to perform an evaluation of the detection performance when each epoch is regarded as a data sample. The calculations of the epoch-based metrics are related to the confusion matrix evaluated on all samples. For instance, many studies has reported the performance as accuracy, ssensitivity, and specificity [ASS+13, GRD+10, AWG06]. The epoch-based metrics can also imply how well the classifier is when a duration is concerned. However, it is hardly said that high values of the epoch-based metrics are clinically referred to good detection performance. For example, the epoch-based metrics is still incredibly high even though the detector misses one whole short seizure activity since other epochs are correctly classified.

Event-based metrics, on the other hand, are used to evaluate a classifier based on seizure events in long EEG signals. In this case, the true positive is counted when there is an overlap between detected epoch as ictal and the annotation, the false positive is declared when a detected period of EEG signal does not overlap an actual seizure period, and the false negative is indicated when there is no detected epoch as ictal during a seizure activity. Note that there is no true negative for the evaluation by an event. Two common metrics, good detection rate (GDR) and false positive rate per hour (FPR/h) calculated based on the intersection of detection results and annotations are also used in this application [VI17, SG10a, SLUC15]. Here, GDR, or event-based sensitivity, is also defined as (17). FPR/h, also called false detection rate per hour, is the proportion of events declared as a seizure without any intersection with the annotations in one hour:

$$\text{FPR/h} = \frac{\text{FP}}{\text{record duration}} \tag{20}$$

A higher GDR indicates a higher number of correctly detected seizure events, while a small FPR/h refers to having a lower number of wrongly recognized seizure events. However, care is required with these high event-based metrics to avoid being misled into a conclusion of a correct detection when a duration is considered. For example, declaring an occurrence of seizure at the last second of an actual seizure event is still counted as good detection even though the detection system nearly misses the whole seizure event.

A latency is a measure of identifying the difference between actual and detected time points. Unfortunately, there is no exact calculation of the latency since many studies have previously defined the latency differently [OCDL16, CUFK19]. Therefore, in this study, the latency is defined as a time delay of a detected seizure when an actual seizure is set to be a reference. Positive and negative onset/offset latencies refer to the declarations of onset/offset after and before the annotation, respectively.

The means of evaluating the automatic epileptic seizure detection is essential to compare the performances of each model. Since our purpose is to detect seizure events and their onset and offset, using a validation scheme that supposes that each epoch is an individual sample is not suitable because we cannot determine the onset and offset if the results is not sequential. In this case, leave-one-record-out cross-validation [SEC+04], as illustrated in Figure 11, was used to validate the proposed method. Suppose that each subject has $k$ records divided into two groups: training set and validation set. The training set contains $k - 1$ records, and the excluded one is in the validation set. In particular, the training set must include seizure and non-seizure activities so that the network can potentially learn to differentiate ictal and non-ictal patterns. The model is then trained on the training set, and validated on the validation set. This process repeats until every record was in the validation set.

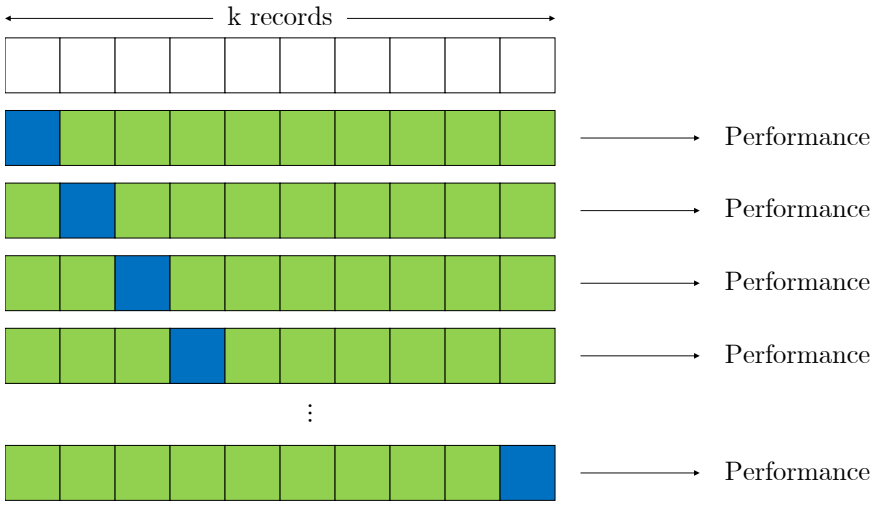



Figure 11: Scheme of leave-one-record out cross validation. The green records are for training, and the blue record for testing.

In this proposal, to reveal the detection performances of every aspects, we used accuracy, sensitivity, specificity, and $F_1$ as epoch-based metrics, FPR/h and GDR for event-based metrics, and seizure onset and offset latencies. We also computed absolute latencies to ignore the sign and obtain the actual delay. In addition, if the onset-offset detector detected many seizure events during only one actual seizure activity, the onset latency was defined as the latency from the first seizure event, and the offset latency was determined from the last event. For each patient, the average of each performance metric was collected. However, using only the mean value, which is influenced by outliers, is sometimes misleading. Therefore, we also reported the median of each performance metric of each patient to overcome the problem. In addition, we compared the differences of the metrics between before and after the onset-offset detection.

# 8 Data collection

This section describes scalp EEG databases that are publicly available online. As stated in Section 2.2, we focus on using multi-channel scalp EEG signals. Furthermore, the scalp EEG signals annotated with all seizure onset and offset are required to train and test the proposed model. Hence, there are currently two online databases which are the CHB-MIT Scalp EEG and Temple University Hospital EEG Seizure (TUSZ) databases that have the desirable requirements. An overview of the databases are illustrated in Table 5. The descriptions of these databases are issued in the following sections.

Table 4: Summary of the CHB-MIT Scalp EEG and TUSZ databases.

| Information | **CHB-MIT** | **TUSZ** |
|---|---|---|
| Number of cases | 24 | 314 |
| Number of files | 686 | 2,997 |
| Number of seizures | 198 | 2,012 |
| Number of files containing seizures | 129 | 703 |
| Record length per file | 1-4 hours | less than one hour |
| Total duration (hour) | 982.37 | 500.02 |
| Total seizure duration (hour) | 3.28 | 42.08 |
| Electrode placement | 10-20 international system | 10-20 international system |
| Montage | bipolar montage | referential montage |
| Sampling frequency (Hz) | 256 | 250 |

### 8.1 CHB-MIT Scalp EEG database

The database comprises of EEG recordings of 24 cases collected from 23 subjects at the Children's Hospital Boston [GAG+00]. Every signal was recorded at the sampling frequency of 256 Hz with resolution of 16 bit. The international 10-20 system was exploited to locate electrodes on the scalp and both referential and bipolar montages were used. In summary, there are 686 long EEG records which include 129 records containing 198 seizures in this database. Total duration and numbers of seizure activities from each case are concluded in Table 5. All records are publicly and freely downloaded from PhysioNet (https://physionet.org/physiobank/database/chbmit/).

Table 5: Summary of the CHB-MIT Scalp EEG database.

| Cases | Number of records | Total duration (sec) | Number of seizures | Total seizure duration (sec) |
|---|---|---|---|---|
| chb01 | 42 | 145,988 | 7 | 449 |
| chb02 | 36 | 126,959 | 3 | 175 |
| chb03 | 38 | 136,806 | 7 | 409 |
| chb04 | 42 | 561,834 | 4 | 382 |
| chb05 | 39 | 140,410 | 5 | 563 |
| chb06 | 18 | 240,246 | 10 | 163 |
| chb07 | 19 | 241,388 | 3 | 328 |
| chb08 | 20 | 72,023 | 5 | 924 |
| chb09 | 19 | 244,338 | 4 | 280 |
| chb10 | 25 | 180,084 | 7 | 454 |
| chb11 | 35 | 123,257 | 3 | 809 |
| chb12 | 24 | 85,300 | 40 | 1,515 |
| chb13 | 33 | 118,800 | 12 | 547 |
| chb14 | 26 | 93,600 | 8 | 177 |
| chb15 | 40 | 144,036 | 20 | 2,012 |
| chb16 | 19 | 68,400 | 10 | 94 |
| chb17 | 21 | 75,624 | 3 | 296 |
| chb18 | 36 | 128,285 | 6 | 323 |
| chb19 | 30 | 107,746 | 3 | 239 |
| chb20 | 29 | 99,366 | 8 | 302 |
| chb21 | 33 | 118,189 | 4 | 203 |
| chb22 | 31 | 111,611 | 3 | 207 |
| chb23 | 9 | 95,610 | 7 | 431 |
| chb24 | 22 | 76,640 | 16 | 527 |
| **sum** | **686** | **3,536,540** | **198** | **11,809** |

### 8.2 Temple University Hospital (TUH) EEG Seizure database

The TUH EEG Seizure Corpus [SvWL+18] is part of the TUH EEG Corpus [OP16] containing several EEG recordings for specific purposes. This TUH EEG Seizure Corpus contains EEG recordings of training and evaluation sets similarly distributed in terms of gender and age of subjects, and the duration of records to reinforce research in artificial intelligence. In total, there are 2,997 recordings pruned to be less than one hour with 2,012 seizure events. The total duration of all records is 500 hours, and the total seizure duration is approximately 42 hours. Every signal was recorded using the international 10-20 system with a sampling frequency of 250 Hz. Referential montages using two different references –averaged reference and linked ear– were applied to collect the data. However, none of patients are included in both training and evaluation sets. The summary of this database is demonstrated in Table 6 The full database is available on TUH EEG resources (https://www.isip.piconepress.com/projects/tuh_eeg/).

Table 6: Summary of the TUSZ database.

| Information | Train | Test | Total |
|---|---|---|---|
| Number of files | 1,984 | 1,013 | 2,997 |
| Number of sessions | 579 | 238 | 817 |
| Number of patients | 264 | 50 | 314 |
| Number of files with seizure | 417 | 286 | 703 |
| Number of sessions with seizure | 197 | 108 | 305 |
| Number of patient with seizure | 130 | 39 | 169 |
| Number of seizure | 1,327 | 685 | 2,012 |
| Total seizure duration (sec) | 90,464.09 | 61,036.84 | 151,500.9 |
| Total duration (sec) | 1,186,842 | 613,232 | 1,800,074 |

# 9 Experiment

In this section, we provide the detailed description of the experiment. This experiment was designed to evaluate and compare the performances of two seizure detection models using (i) only a classifier, and (ii) the same classifier followed by an additional onset-offset detector. According to the scope, the CHB-MIT Scalp EEG database was used in the experiment since, in the TUSZ database, there is no subject from the training set included in the development set. In the experimental setting, we describe the CNN network configuration and the intuition behind it. We also report the performances of each subject by mean and median for the quantitative and quantitative analysis.

## 9.1 Experimental setting

In this proposal, all EEG records from every subject in the CHB-MIT Scalp EEG database were applied in this proposal. Since a montage of each long EEG signal was not consistent, *i.e.*, both referential and bipolar montages were employed even though those EEG signals were from the same patient, all EEG signals were initially modified so that all montages were bipolar. The channels of the modified signals were sequentially listed as *FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-T3, T3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ*, and *CZ-PZ*. Then the long modified signals of every channel were jointly segmented into small epochs where each epoch was defined as one sample to be classified. The epoch size was chosen to be one second without overlap to reduce model complexity and redundancy between adjacent epochs. Since the loss was fluctuated while training, we set a stopping criteria based on a number of iteration instead of the decay of the loss. So the training process was repeated 100 iterations from no considerable change in a confusion matrix, and the batch size was set to be 100 samples to train the CNN model.

We designed a deep CNN model as illustrated in Figure 12. The model input was a raw EEG epoch, and the model output was a seizure probability. In Figure 12, each rectangular box represents a layer, and the description in the box explains the type of the layer. In this case, Conv$(h, w, f)$ is a convolutional layer containing $f$ $h$-by-$w$ filters, BN stands for a batch normalization layer, ReLU is an activation layer using the ReLU function as the activation function, Max$(h, w)$ is an $h$-by-$w$ max-pooling layer, Dropout$(\alpha)$ is a dropout layer with the disconnection fraction of $\alpha$ to the input nodes, and FC$(a)$ is a fully-connected layer with $a$ neurons. The number of filters in each block and the number of block was modified from our previous work in [BLuCS19a] reducing the number of parameter while the CNN model tested on the records of the subject `chb24` still gave high GDR. The optimizer called ADADELTA was exploited to train the model because it was robust to noise and had an adaptive learning rate [Zei12]. Furthermore, the loss function was the binary cross entropy and the sample was denoted as ictal when the seizure probability was higher than 0.5.

For parameter setting of the onset-offset detection, we chose the window width to be $l = 2$ epochs. Moreover, there must be $p = 3$ consecutive epochs that their seizure window intersected or contacted. These choices were selected based on the shortest seizure activity, which is seven

seconds long, so that the proposed detection could suitably capture other seizure activities.

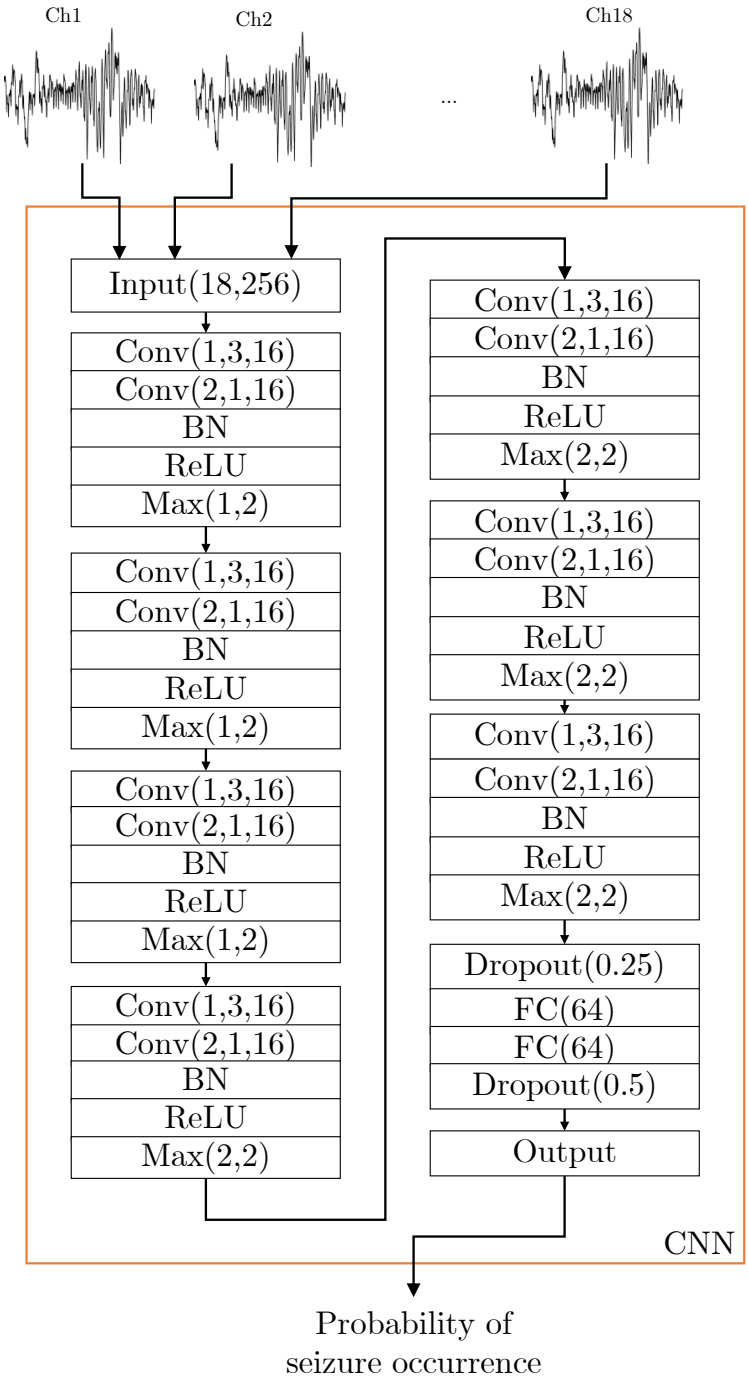



Figure 12: Deep CNN structure used in this proposal. Raw EEG signals from the chosen channels are together fed to the deep CNN to produce the seizure probability.

### 9.2 Preliminary results

Overall, Tables 7 and 8 summarize each performance metric evaluated on each individual subject before and after the onset-offset detection, including the average, minimum, and maximum values over the cases. Table 9 shows an improvement on the classification performances of the onset-offset detector from when the mean and median of each patient were compared. In this case, a positive improvement indicates a better performance, a negative improvement determines a worse performance, and '-' means the performances of using only the CNN is originally zero. In addition, Figures 13 to 16 display epoch-based metrics: accuracy, specificity, sensitivity, and $F_1$, and Fig-

ures 17 and 18 show event-based metrics: GDR and FPR/h to easily compare the differences of each case. Particularly, Figures 19 to 24 visualize results of using the CNN and the combination of the CNN and the onset-offset detector. Generally, we can see that seizure probabilities in seizure events were relatively high compared to probabilities during normal periods.

Table 7: **Mean** of performance of each subject evaluated on the CHB-MIT Scalp EEG database. All performance metrics but FPR/h are represented in percentage.

| Cases | Before onset-offset detection | | | | | | After onset-offset detection | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Event-based | | Epoch-based | | | | Event-based | | Epoch-based | | | |
| | FPR/h | GDR | Acc | Sen | Spec | $F_1$ | FPR/h | GDR | Acc | Sen | Spec | $F_1$ |
| chb01 | 0.65 | 100.00 | 99.83 | 43.12 | 99.98 | 35.51 | 0.00 | 100.00 | 99.91 | 64.79 | 100.00 | 75.90 |
| chb02 | 0.00 | 66.67 | 99.81 | 19.74 | 100.00 | 29.07 | 0.00 | 66.67 | 99.88 | 31.81 | 100.00 | 39.95 |
| chb03 | 0.24 | 100.00 | 99.85 | 51.53 | 99.99 | 40.79 | 0.00 | 100.00 | 99.93 | 76.67 | 99.99 | 83.57 |
| chb04 | 0.00 | 33.33 | 99.93 | 1.33 | 100.00 | 2.56 | 0.00 | 0.00 | 99.93 | 0.00 | 100.00 | 0.00 |
| chb05 | 0.33 | 100.00 | 99.90 | 77.96 | 99.99 | 48.20 | 0.00 | 100.00 | 99.95 | 89.21 | 100.00 | 93.05 |
| chb06 | 0.19 | 100.00 | 99.96 | 54.09 | 99.99 | 39.57 | 0.03 | 100.00 | 99.97 | 87.48 | 99.98 | 70.05 |
| chb07 | 0.30 | 100.00 | 99.88 | 39.82 | 99.99 | 31.82 | 0.05 | 100.00 | 99.93 | 64.69 | 99.98 | 73.53 |
| chb08 | 2.20 | 100.00 | 99.12 | 37.91 | 99.92 | 22.24 | 0.25 | 100.00 | 99.36 | 55.92 | 99.93 | 64.15 |
| chb09 | 1.14 | 100.00 | 99.91 | 79.48 | 99.94 | 53.18 | 0.24 | 100.00 | 99.89 | 91.21 | 99.90 | 73.43 |
| chb10 | 1.74 | 100.00 | 99.86 | 82.83 | 99.90 | 38.36 | 0.10 | 100.00 | 99.85 | 95.77 | 99.86 | 87.36 |
| chb11 | 0.03 | 100.00 | 99.61 | 64.47 | 100.00 | 57.66 | 0.00 | 100.00 | 99.78 | 90.75 | 99.99 | 92.85 |
| chb12 | 17.36 | 96.92 | 98.41 | 53.86 | 99.34 | 33.90 | 0.79 | 92.31 | 98.92 | 72.11 | 99.52 | 59.26 |
| chb13 | 2.36 | 95.83 | 99.57 | 22.33 | 99.92 | 15.80 | 0.24 | 70.83 | 99.62 | 37.80 | 99.91 | 36.49 |
| chb14 | 0.77 | 100.00 | 99.87 | 42.53 | 99.98 | 25.85 | 0.00 | 85.71 | 99.94 | 68.39 | 100.00 | 73.03 |
| chb15 | 2.87 | 100.00 | 99.05 | 57.17 | 99.65 | 35.53 | 0.20 | 92.86 | 99.29 | 72.79 | 99.66 | 65.17 |
| chb16 | 0.21 | 83.33 | 99.88 | 22.76 | 99.99 | 24.77 | 0.05 | 45.83 | 99.89 | 36.39 | 99.99 | 39.17 |
| chb17 | 2.48 | 66.67 | 99.63 | 20.39 | 99.93 | 7.74 | 0.00 | 66.67 | 99.77 | 36.96 | 100.00 | 46.62 |
| chb18 | 0.81 | 100.00 | 99.79 | 36.80 | 99.96 | 37.63 | 0.06 | 100.00 | 99.83 | 60.14 | 99.95 | 60.27 |
| chb19 | 0.33 | 100.00 | 99.89 | 54.42 | 99.99 | 34.67 | 0.00 | 100.00 | 99.94 | 73.99 | 100.00 | 84.56 |
| chb20 | 0.68 | 58.33 | 99.66 | 14.15 | 99.93 | 11.00 | 0.07 | 50.00 | 99.69 | 22.20 | 99.94 | 22.75 |
| chb21 | 2.98 | 100.00 | 99.75 | 6.83 | 99.92 | 4.84 | 0.06 | 25.00 | 99.82 | 6.40 | 99.98 | 8.16 |
| chb22 | 0.42 | 100.00 | 99.90 | 52.50 | 99.99 | 24.89 | 0.03 | 100.00 | 99.94 | 74.07 | 99.99 | 81.10 |
| chb23 | 1.18 | 100.00 | 99.68 | 43.56 | 99.93 | 26.45 | 0.11 | 100.00 | 99.77 | 71.48 | 99.91 | 71.14 |
| chb24 | 8.45 | 91.67 | 99.23 | 36.50 | 99.68 | 25.17 | 0.27 | 91.67 | 99.56 | 58.50 | 99.86 | 57.48 |
| **max** | 17.36 | 100.00 | 99.96 | 82.83 | 100.00 | 57.66 | 0.79 | 100.00 | 99.97 | 95.77 | 100.00 | 93.05 |
| **min** | 0.00 | 33.33 | 98.41 | 1.33 | 99.34 | 2.56 | 0.00 | 0.00 | 98.92 | 0.00 | 99.52 | 0.00 |
| **mean** | 1.99 | 91.36 | 99.66 | 42.34 | 99.91 | 29.47 | 0.11 | 82.81 | 99.77 | 59.98 | 99.93 | 60.79 |

Focusing on using a mean to evaluate the performances in Table 7, we found that accuracy and specificity achieved by the CNN model were almost 100% from every case. Good detection rates from all cases except `chb04` were also high, which were 91.36% on average from all subjects. However, without the onset-offset detector, the CNN model obtained low sensitivity and $F_1$. The CNN model obtained average sensitivity of 42.34%, and the minimum of 1.33%, and the maximum of 82.83%. The average, minimum, and maximum of $F_1$ were 29.47%, 2.56%, and 57.66%, respectively. This means that the CNN model yields bad epoch-based classification results. Furthermore, the CNN model produced several false positives in many cases, resulting the FPR/h of 1.99 on average and of 17.36 in the extreme case. In the case of using a median, as shown in Table 8, accuracy and specificity were also high, and both the averages and ranges of sensitivity and $F_1$ were low. On the other hand, GDRs were higher in many cases, and the average was 94.79%. Additionally, the CNN model obtained zero FPR/h in almost all cases but a high FPR/h was achieved in only the `chb12` case. As a result, it means that many FP occur in a few records, and there is normally no FP in other records; the CNN model alone is, therefore, inconsistent across patients. With high FPR/h and Spec, this implies that some individual and separated normal epochs are unfortunately detected as ictal. Moreover, the CNN model may effectively find a seizure event in the record but is ineffective in determining the whole seizure event because the GDR was high but the sensitivity was intermediate.

Table 8: **Median** of performances of each subject evaluated on the CHB-MIT Scalp EEG database. All performance metrics but FPR/h are represented in percentage.

| Cases | Before onset-offset detection | | | | | | After onset-offset detection | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Event-based | | Epoch-based | | | | Event-based | | Epoch-based | | | |
| | FPR/h | GDR | Acc | Sen | Spec | $F_1$ | FPR/h | GDR | Acc | Sen | Spec | $F_1$ |
| chb01 | 0.00 | 100.00 | 100.00 | 31.71 | 100.00 | 42.42 | 0.00 | 100.00 | 100.00 | 60.98 | 100.00 | 75.76 |
| chb02 | 0.00 | 100.00 | 100.00 | 14.63 | 100.00 | 25.53 | 0.00 | 100.00 | 100.00 | 20.73 | 100.00 | 34.34 |
| chb03 | 0.00 | 100.00 | 100.00 | 46.15 | 100.00 | 43.48 | 0.00 | 100.00 | 100.00 | 83.33 | 100.00 | 90.91 |
| chb04 | 0.00 | 0.00 | 100.00 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 | 100.00 | 0.00 |
| chb05 | 0.00 | 100.00 | 100.00 | 85.34 | 100.00 | 67.04 | 0.00 | 100.00 | 100.00 | 95.04 | 100.00 | 97.46 |
| chb06 | 0.00 | 100.00 | 99.97 | 61.54 | 100.00 | 43.96 | 0.00 | 100.00 | 99.99 | 85.71 | 100.00 | 88.00 |
| chb07 | 0.00 | 100.00 | 100.00 | 39.58 | 100.00 | 26.00 | 0.00 | 100.00 | 100.00 | 70.14 | 100.00 | 82.45 |
| chb08 | 0.50 | 100.00 | 99.97 | 34.30 | 99.99 | 0.00 | 0.00 | 100.00 | 100.00 | 64.91 | 100.00 | 74.46 |
| chb09 | 0.00 | 100.00 | 100.00 | 83.08 | 100.00 | 60.27 | 0.00 | 100.00 | 100.00 | 89.23 | 100.00 | 92.31 |
| chb10 | 0.50 | 100.00 | 99.99 | 83.33 | 99.99 | 0.00 | 0.00 | 100.00 | 100.00 | 98.59 | 100.00 | 96.30 |
| chb11 | 0.00 | 100.00 | 100.00 | 52.17 | 100.00 | 67.71 | 0.00 | 100.00 | 100.00 | 100.00 | 100.00 | 97.06 |
| chb12 | 8.98 | 100.00 | 99.36 | 65.07 | 99.72 | 40.74 | 0.00 | 100.00 | 99.60 | 88.65 | 100.00 | 68.25 |
| chb13 | 0.00 | 100.00 | 100.00 | 20.14 | 100.00 | 6.06 | 0.00 | 83.33 | 100.00 | 26.39 | 100.00 | 29.55 |
| chb14 | 0.00 | 100.00 | 99.97 | 47.06 | 100.00 | 0.00 | 0.00 | 100.00 | 100.00 | 88.24 | 100.00 | 93.75 |
| chb15 | 0.50 | 100.00 | 99.92 | 73.08 | 99.99 | 16.67 | 0.00 | 100.00 | 100.00 | 86.59 | 100.00 | 80.06 |
| chb16 | 0.00 | 100.00 | 100.00 | 26.13 | 100.00 | 29.17 | 0.00 | 37.50 | 100.00 | 29.16 | 100.00 | 42.29 |
| chb17 | 0.00 | 100.00 | 99.97 | 14.61 | 100.00 | 0.00 | 0.00 | 100.00 | 100.00 | 39.33 | 100.00 | 56.45 |
| chb18 | 0.00 | 100.00 | 100.00 | 38.92 | 100.00 | 42.95 | 0.00 | 100.00 | 100.00 | 66.90 | 100.00 | 67.50 |
| chb19 | 0.00 | 100.00 | 100.00 | 53.85 | 100.00 | 34.40 | 0.00 | 100.00 | 100.00 | 75.61 | 100.00 | 86.11 |
| chb20 | 0.00 | 75.00 | 100.00 | 10.16 | 100.00 | 0.00 | 0.00 | 50.00 | 100.00 | 15.00 | 100.00 | 0.00 |
| chb21 | 0.00 | 100.00 | 100.00 | 5.81 | 100.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 | 100.00 | 0.00 |
| chb22 | 0.00 | 100.00 | 100.00 | 46.58 | 100.00 | 0.00 | 0.00 | 100.00 | 100.00 | 65.33 | 100.00 | 79.03 |
| chb23 | 0.50 | 100.00 | 99.98 | 39.13 | 99.99 | 18.24 | 0.00 | 100.00 | 100.00 | 72.81 | 100.00 | 73.71 |
| chb24 | 1.00 | 100.00 | 99.43 | 35.08 | 99.97 | 32.00 | 0.00 | 100.00 | 99.72 | 62.41 | 100.00 | 73.44 |
| **max** | 8.98 | 100.00 | 100.00 | 85.34 | 100.00 | 67.71 | 0.00 | 100.00 | 100.00 | 100.00 | 100.00 | 97.46 |
| **min** | 0.00 | 0.00 | 99.36 | 0.00 | 99.72 | 0.00 | 0.00 | 0.00 | 99.60 | 0.00 | 100.00 | 0.00 |
| **mean** | 0.50 | 94.79 | 99.94 | 41.98 | 99.98 | 24.86 | 0.00 | 86.28 | 99.97 | 61.88 | 100.00 | 65.80 |

When the onset-offset detector was exploited, sensitivity and $F_1$ substantially increased, and FPR/h significantly decreased. As demonstrated in Table 9, the onset-offset detector could potentially improve the sensitivity, $F_1$, and FPR/h. When using the mean value to observe the overall performance, we discovered that the sensitivity positively increased at least 14% and the $F_1$ also grew up more than 37%. This means the onset-offset detector can considerably improve the performances by 137.68% in $F_1$ and 50.75% in sensitivity on average. For instance, Figure 19 demonstrates the seizure probability, epoch-based decision, and output from the onset-offset detector tested on the sample `chb01_04`, where Figures 19a and 19b show the results of the whole record and during the seizure activity, respectively. Another example validated on the sample `chb05_06` with a longer epileptic seizure is depicted in Figure 20. As we expected, the CNN did not provide high seizure probabilities to all epochs in the duration. Therefore, the decision made based on the probability of individual epochs is not sufficiently reasonable, and an onset-offset detector is needed. Figures 19b and 20b show that our proposed onset-offset detector could potentially prevent incorrectly detected epochs during the seizure activities. We found that the epochs classified as normal during the activities were changed to be ictal. In addition, FPR/h was absolutely reduced by at least 75% in every case except in `chb02` and `chb04` cases because there was no FPR/h in these cases. As shown in Figure 21, all false positives were annihilated so that the FPR/h was reduced.

Nevertheless, applying the onset-offset detection does not always provide favorable results. We discovered that these undesirable results are caused by two stages: epoch-based classification and onset-offset detection. For example, if the CNN produces too many consecutive false positives or few true positives, the final results can be misinterpreted. Figure 22 shows that consecutive false positives cannot be eliminated by the onset-offset detection and those false positives still occur in

Table 9: Overall improvements of the onset-offset detector. A positive sign presents an improvement of using the onset-offset detector, a negative value means the onset-offset detector decreases the detection performances, and '-' means the reference is originally zero. Every metrics is expressed in percentage.

| Cases | Mean | | | | | | Median | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Event-based | | Epoch-based | | | | Event-based | | Epoch-based | | | |
| | FPR/h | GDR | Acc | Sen | Spec | $F_1$ | FPR/h | GDR | Acc | Sen | Spec | $F_1$ |
| chb01 | 100.00 | 0.00 | 0.08 | 50.25 | 0.02 | 113.75 | - | 0.00 | 0.00 | 92.31 | 0.00 | 78.57 |
| chb02 | - | 0.00 | 0.08 | 61.17 | 0.00 | 37.46 | - | 0.00 | 0.00 | 41.67 | 0.00 | 34.51 |
| chb03 | 100.00 | 0.00 | 0.08 | 48.78 | 0.00 | 104.86 | - | 0.00 | 0.00 | 80.56 | 0.00 | 109.09 |
| chb04 | - | -100.00 | 0.00 | -100.00 | 0.00 | -100.00 | - | - | 0.00 | - | 0.00 | - |
| chb05 | 100.00 | 0.00 | 0.05 | 14.44 | 0.01 | 93.03 | - | 0.00 | 0.00 | 11.36 | 0.00 | 45.37 |
| chb06 | 85.46 | 0.00 | 0.02 | 61.73 | 0.00 | 77.02 | - | 0.00 | 0.02 | 39.29 | 0.00 | 100.18 |
| chb07 | 82.94 | 0.00 | 0.05 | 62.44 | -0.01 | 131.11 | - | 0.00 | 0.00 | 77.19 | 0.00 | 217.11 |
| chb08 | 88.64 | 0.00 | 0.25 | 47.53 | 0.01 | 188.42 | 100.00 | 0.00 | 0.03 | 89.22 | 0.01 | - |
| chb09 | 79.31 | 0.00 | -0.02 | 14.76 | -0.04 | 38.06 | - | 0.00 | 0.00 | 7.41 | 0.00 | 53.16 |
| chb10 | 94.25 | 0.00 | -0.01 | 15.63 | -0.04 | 127.76 | 100.00 | 0.00 | 0.01 | 18.31 | 0.01 | - |
| chb11 | 100.00 | 0.00 | 0.17 | 40.76 | 0.00 | 61.03 | - | 0.00 | 0.00 | 91.67 | 0.00 | 43.35 |
| chb12 | 95.45 | -4.76 | 0.52 | 33.88 | 0.18 | 74.78 | 100.00 | 0.00 | 0.24 | 36.24 | 0.28 | 67.53 |
| chb13 | 89.74 | -26.09 | 0.05 | 69.23 | -0.01 | 130.95 | - | -16.67 | 0.00 | 31.03 | 0.00 | 387.61 |
| chb14 | 100.00 | -14.29 | 0.07 | 60.79 | 0.02 | 182.50 | - | 0.00 | 0.03 | 87.50 | 0.00 | - |
| chb15 | 93.03 | -7.14 | 0.24 | 27.32 | 0.01 | 83.45 | 100.00 | 0.00 | 0.08 | 18.50 | 0.01 | 380.35 |
| chb16 | 75.00 | -45.00 | 0.01 | 59.88 | 0.00 | 58.14 | - | -62.50 | 0.00 | 11.58 | 0.00 | 44.98 |
| chb17 | 100.00 | 0.00 | 0.14 | 81.30 | 0.07 | 502.23 | - | 0.00 | 0.03 | 169.23 | 0.00 | - |
| chb18 | 93.10 | 0.00 | 0.04 | 63.41 | -0.02 | 60.15 | - | 0.00 | 0.00 | 71.86 | 0.00 | 57.14 |
| chb19 | 100.00 | 0.00 | 0.05 | 35.97 | 0.01 | 143.89 | - | 0.00 | 0.00 | 40.42 | 0.00 | 150.32 |
| chb20 | 89.83 | -14.29 | 0.03 | 56.88 | 0.00 | 106.90 | - | -33.33 | 0.00 | 47.66 | 0.00 | - |
| chb21 | 97.97 | -75.00 | 0.07 | -6.25 | 0.07 | 68.37 | - | -100.00 | 0.00 | -100.00 | 0.00 | - |
| chb22 | 92.31 | 0.00 | 0.04 | 41.08 | 0.00 | 225.80 | - | 0.00 | 0.00 | 40.27 | 0.00 | - |
| chb23 | 91.09 | 0.00 | 0.09 | 64.10 | -0.02 | 168.96 | 100.00 | 0.00 | 0.02 | 86.06 | 0.01 | 304.04 |
| chb24 | 96.77 | 0.00 | 0.33 | 60.25 | 0.18 | 128.37 | 100.00 | 0.00 | 0.29 | 77.90 | 0.03 | 129.49 |
| **max** | 100.00 | 0.00 | 0.52 | 81.30 | 0.18 | 502.23 | 100.00 | 0.00 | 0.29 | 169.23 | 0.28 | 387.61 |
| **min** | 75.00 | -100.00 | -0.02 | -100.00 | -0.04 | -100.00 | 100.00 | -100.00 | 0.00 | -100.00 | 0.00 | 34.51 |
| **mean** | 92.95 | -11.94 | 0.10 | 40.22 | 0.02 | 116.96 | 100.00 | -9.24 | 0.03 | 50.75 | 0.02 | 137.68 |

the final result. So the onset-offset detector can only group individual false positives into one and reduce FPR/h intermediately. Another case of a few true positives is depicted in Figure 23. In this case, only a few epochs detected as ictal were ignored, and consequently the GDR was decreased. As shown in 18, the GDRs evaluated on the subjects `chb04` and `chb16` were significantly reduced from 33.33% and 83.33% to 0% and 45.83%, respectively. Therefore, it is obvious that if epoch-based results are obviously unacceptable, the onset-offset detector cannot give satisfactory outcomes as well. On the other hand, the onset-offset detection sometimes unfortunately gives undesirable results even though the epoch-based classification results seem to be useful. For instance, Figure 24 visualizes the result assessed on the record `chb11_99` was not reliable since the onset-offset detector predicts that there are many seizure activities during one actual event. Figure 24b shows that several epochs during the actual seizure event are detected as seizure. However, they are inadequately close so that the onset-offset detector groups them into the separate seizure events.

Since the CNN model can not provide promising results when the duration is considered, it is unreasonable to compare onset and offset latencies between our two models. Moreover, as reviewed in Section 4.3, Orosco et al. [OCDL16] and Chandel et al.[CUFK19] computed means of positive and negative latencies, the final results were definitely unclear whether the reported latencies were potentially usable. Additionally, the data used in their works were not clearly described. For example, these two works both declared to apply the same criteria, *i.e.*, all records must have the same montage, to choose EEG records. However, these studies eventually selected different subjects and different records. Therefore, we can only compare and discuss the ranges of the latencies including the positive and negative cases. In addition, the absolute onset and offset latencies were

included to validate the variation of the latencies of the onset and offset.

Table 10: Comparisons of onset and offset latencies. Absolute latencies from our work are also reported. All latencies are represented in second. '-' in other works indicates that those records were excluded in the experiments.

| Case | Onset | | | Offset | | | Absolute onset | Absolute offset |
|---|---|---|---|---|---|---|---|---|
| | Our work | [OCDL16] | [CUFK19] | Our work | [OCDL16] | [CUFK19] | Our work | Our work |
| chb01 | 7.43 | 6.70 | 0.29 | -12.00 | 14.00 | 0.71 | 7.71 | 12.00 |
| chb02 | 5.50 | - | 1.33 | -37.00 | - | 5.33 | 5.50 | 37.00 |
| chb03 | 2.86 | 10.70 | 1.29 | -5.71 | 2.40 | 29.86 | 4.86 | 6.29 |
| chb04 | - | - | 8.00 | - | - | 17.50 | - | - |
| chb05 | 0.60 | 14.40 | 7.60 | -5.60 | 22.60 | 0.80 | 1.80 | 6.40 |
| chb06 | 1.10 | - | - | -0.20 | - | - | 2.10 | 0.80 |
| chb07 | 5.33 | 2.00 | 2.00 | -30.00 | -2.00 | -6.00 | 6.00 | 30.00 |
| chb08 | 7.20 | -11.60 | 3.00 | -62.80 | 13.30 | 5.80 | 7.20 | 62.80 |
| chb09 | 1.75 | 5.60 | 1.50 | -2.25 | -0.50 | -1.50 | 2.75 | 2.25 |
| chb10 | -1.71 | - | 2.86 | -2.29 | - | -5.71 | 1.71 | 2.57 |
| chb11 | -2.00 | -9.30 | -9.67 | -8.00 | -9.40 | 0.67 | 2.00 | 8.67 |
| chb12 | 0.83 | -2.00 | - | -5.94 | 0.60 | - | 3.51 | 7.66 |
| chb13 | 4.38 | 8.20 | - | -16.00 | 60.60 | - | 4.38 | 16.50 |
| chb14 | 0.43 | 5.30 | - | -5.29 | 7.60 | - | 0.43 | 5.86 |
| chb15 | 4.32 | - | - | -13.63 | - | - | 5.68 | 14.79 |
| chb16 | -0.50 | 2.00 | - | -2.00 | -3.50 | - | 0.50 | 2.00 |
| chb17 | 20.00 | 7.20 | 0.33 | -23.50 | -1.20 | 7.00 | 20.00 | 23.50 |
| chb18 | 7.67 | -28.00 | -4.50 | -8.17 | -13.70 | 52.67 | 9.00 | 8.17 |
| chb19 | 10.67 | 3.40 | 9.67 | -8.33 | -24.00 | -1.33 | 11.33 | 8.33 |
| chb20 | 10.50 | -7.80 | 7.63 | -11.25 | -24.20 | 10.38 | 10.50 | 11.25 |
| chb21 | 1.00 | -22.00 | -4.50 | -36.00 | 11.30 | 13.50 | 1.00 | 36.00 |
| chb22 | 18.33 | 6.30 | 3.00 | -0.33 | -4.40 | 2.00 | 18.33 | 0.33 |
| chb23 | 8.14 | -6.10 | -10.00 | -8.86 | 43.30 | 16.57 | 8.43 | 9.14 |
| chb24 | 5.00 | - | 5.50 | -8.27 | - | -0.88 | 5.93 | 8.53 |
| **mean** | 5.17 | -0.20 | 1.41 | -13.63 | 4.00 | 8.19 | 6.12 | 13.95 |
| **min** | -2 | -28.00 | -10.00 | -62.8 | -24.20 | -6.00 | 0.43 | 0.33 |
| **max** | 20 | 14.40 | 9.67 | -0.2 | 60.20 | 52.67 | 20 | 62.8 |

Table 10 shows a comparison of the average onset and offset latencies in our and the best methods of others studies where the cells filled with '-' means the study did not perform on those cases. Since our onset-offset detector could not detect any seizure activities from the case `chb04`, the onset and offset latencies were not found. In our study, onset latencies from almost all subjects were positive ranging from $-2.00$ to 20.00 seconds, and offset latencies were negative in the range of $-62.80$ to 0.20 seconds, respectively, when using our proposed onset-offset detector. Moreover, the onset and offset latencies were close to the absolute onset and offset latencies regardless the sign. This indicates that there is only a small amount of cancellation between positive and negative latencies. Consequently, it implies that the onset-offset detector mostly predicts the seizure activities inside the actual seizure. On the other hand, onset and offset latencies from other works were highly fluctuated. Many positive and negative latencies occurred in each subject independently. For example, the maximum and minimum onset latencies from [OCDL16] were 14.40 and -28.00 seconds, where as the maximum and minimum offset latencies were 60.20 and -24.20 seconds, respectively. It is possible that, in each case, some positive and negative latencies are neutralized. Although the latencies from other studies were unclear, ranges of the onset and offset latencies in our work were not far the ranges in other works. Moreover, we used all records from every patient instead of chosen records, making the data more imbalanced. It is well-known that highly imbalanced data can reduce classification performances significantly [JK19]. Therefore, it can be concluded that our proposed model can potentially detect onsets and offsets better than the existing models.

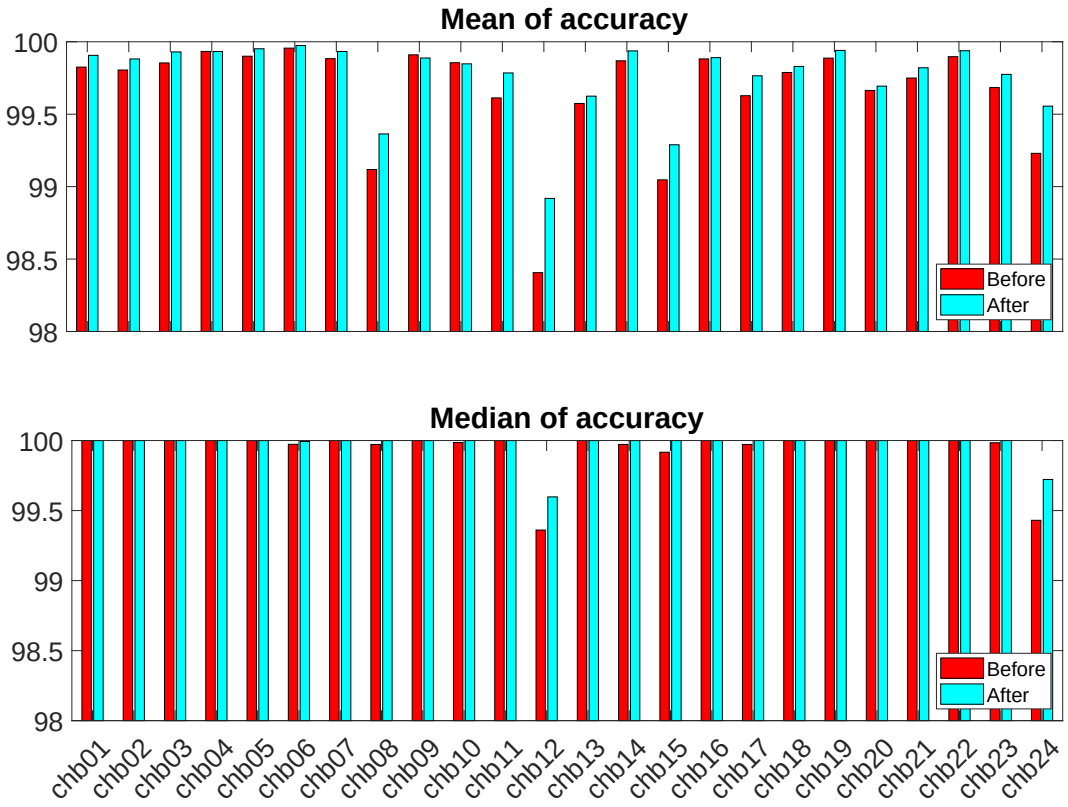



Figure 13: Mean and median of accuracy obtained from before and after the onset-offset detection.

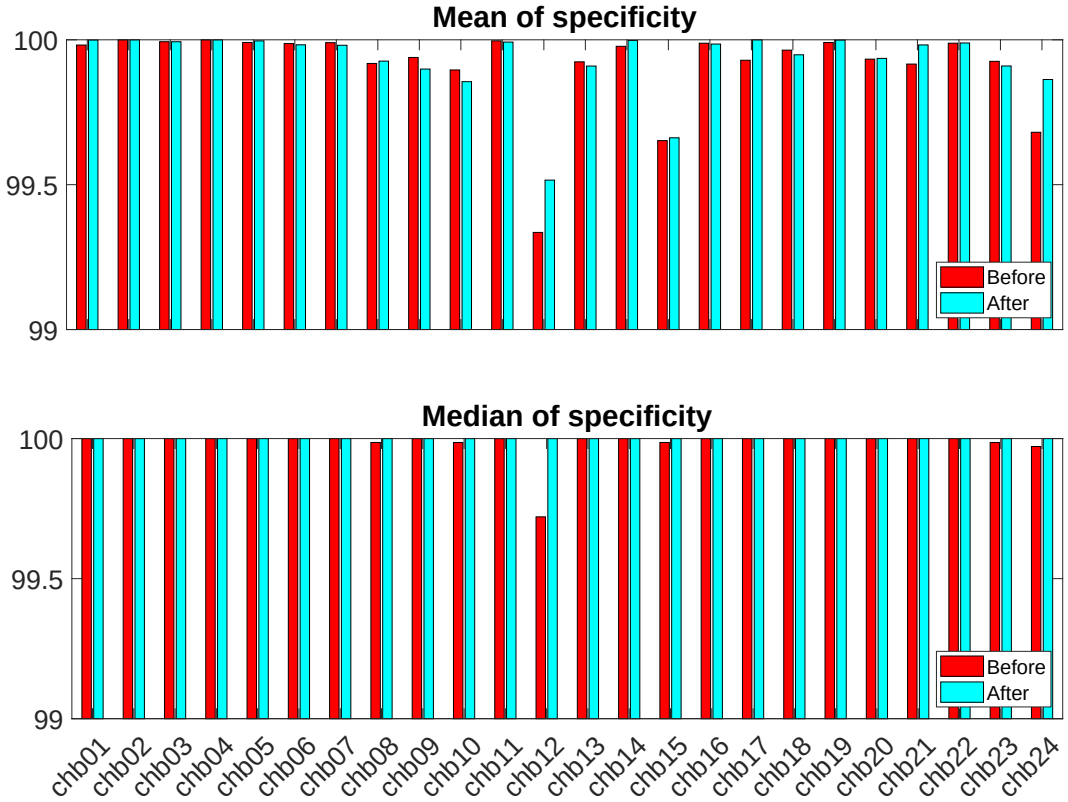



Figure 14: Mean and median of specificity obtained from before and after onset-offset detection.

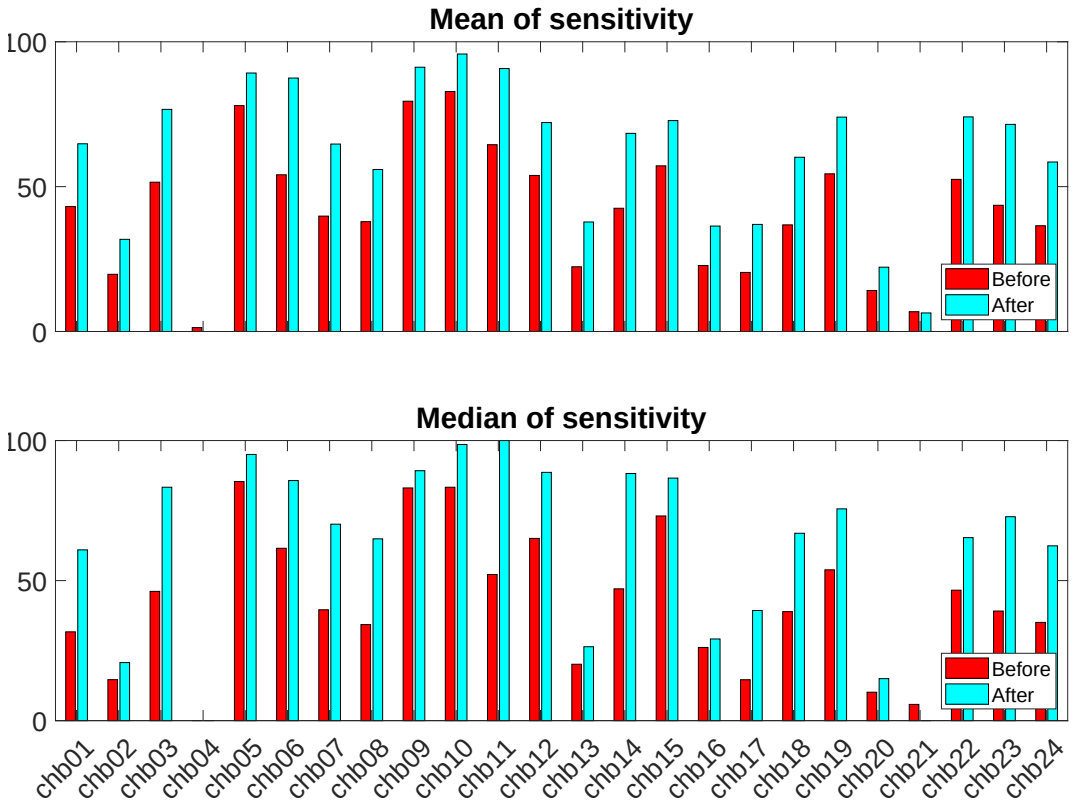



Figure 15: Mean and median of sensitivity obtained from before and after the onset-offset detection.

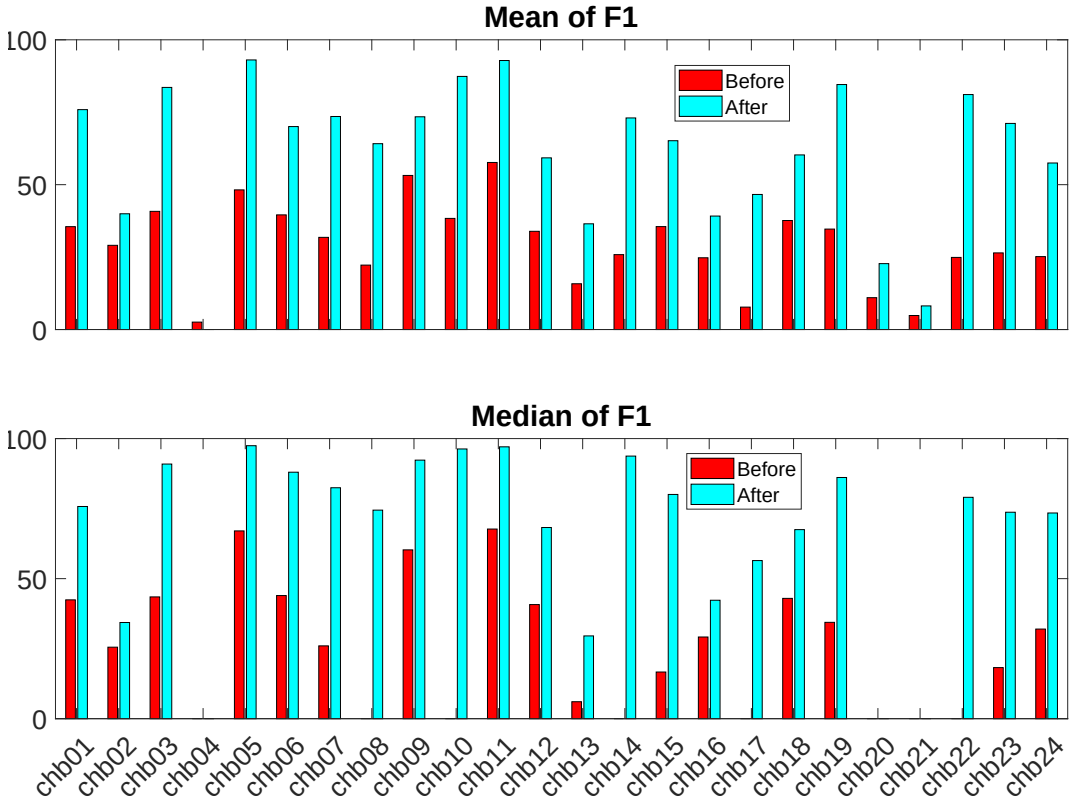



Figure 16: Mean and median of $F_1$ obtained from before and after the onset-offset detection.

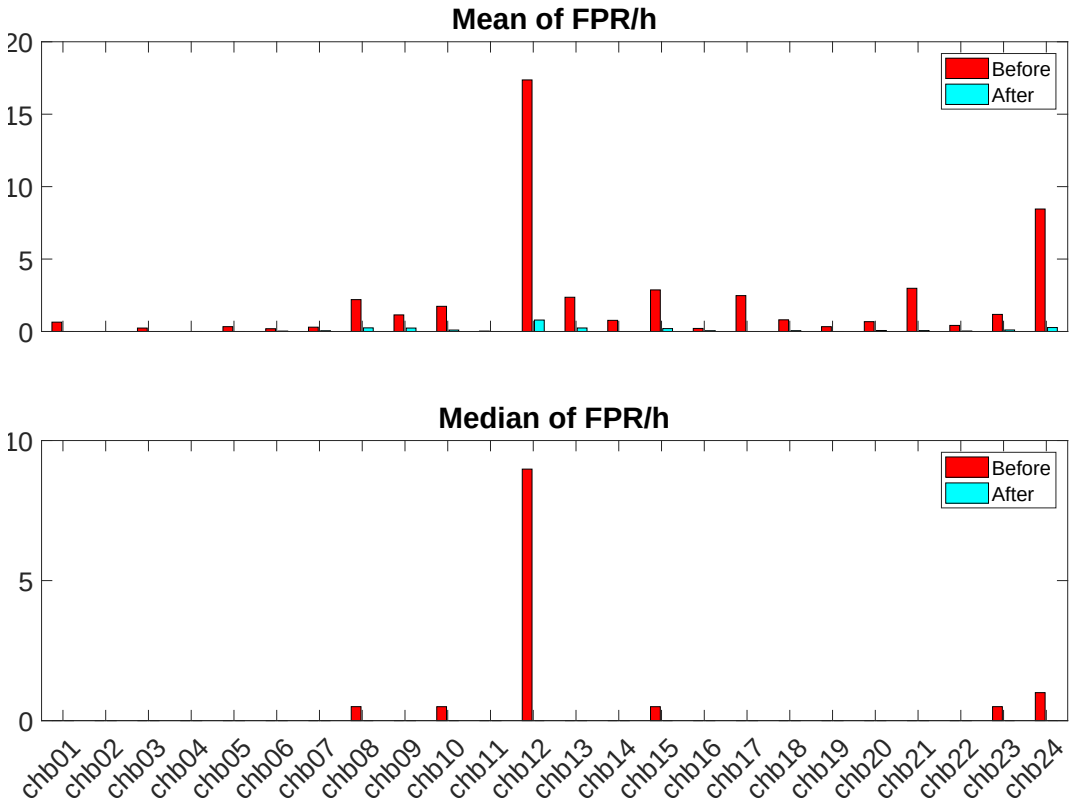



Figure 17: FPR/h obtained from before and after the onset-offset detection.

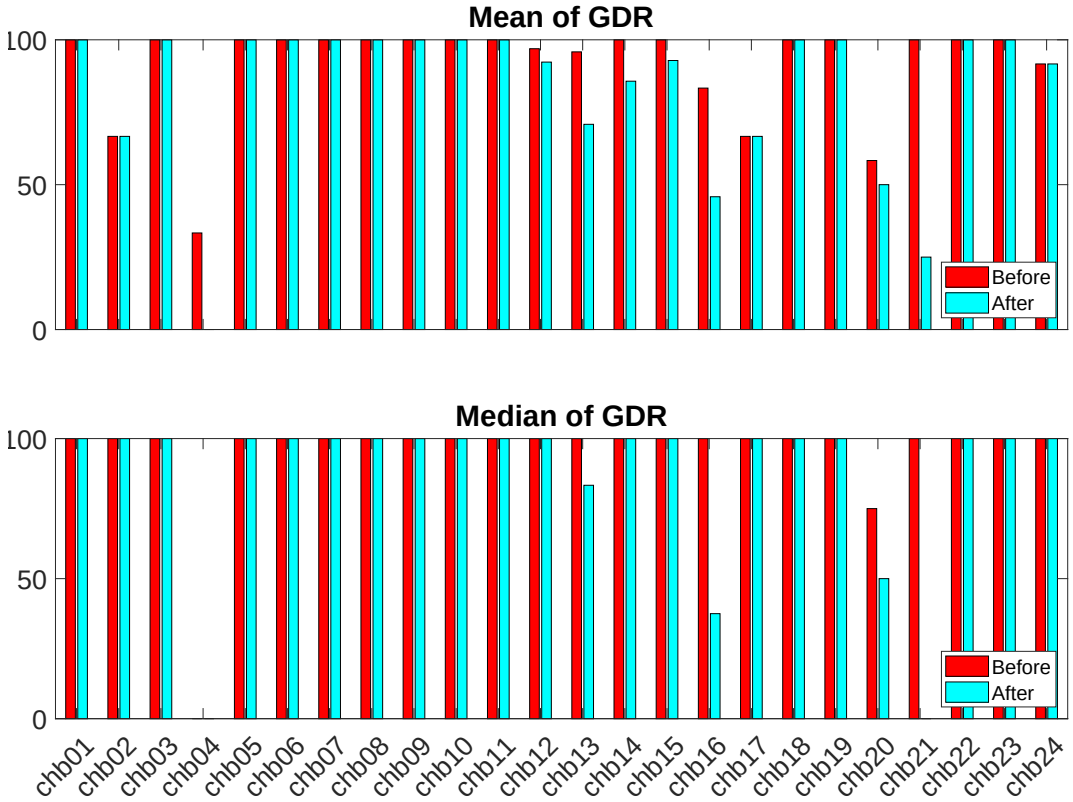
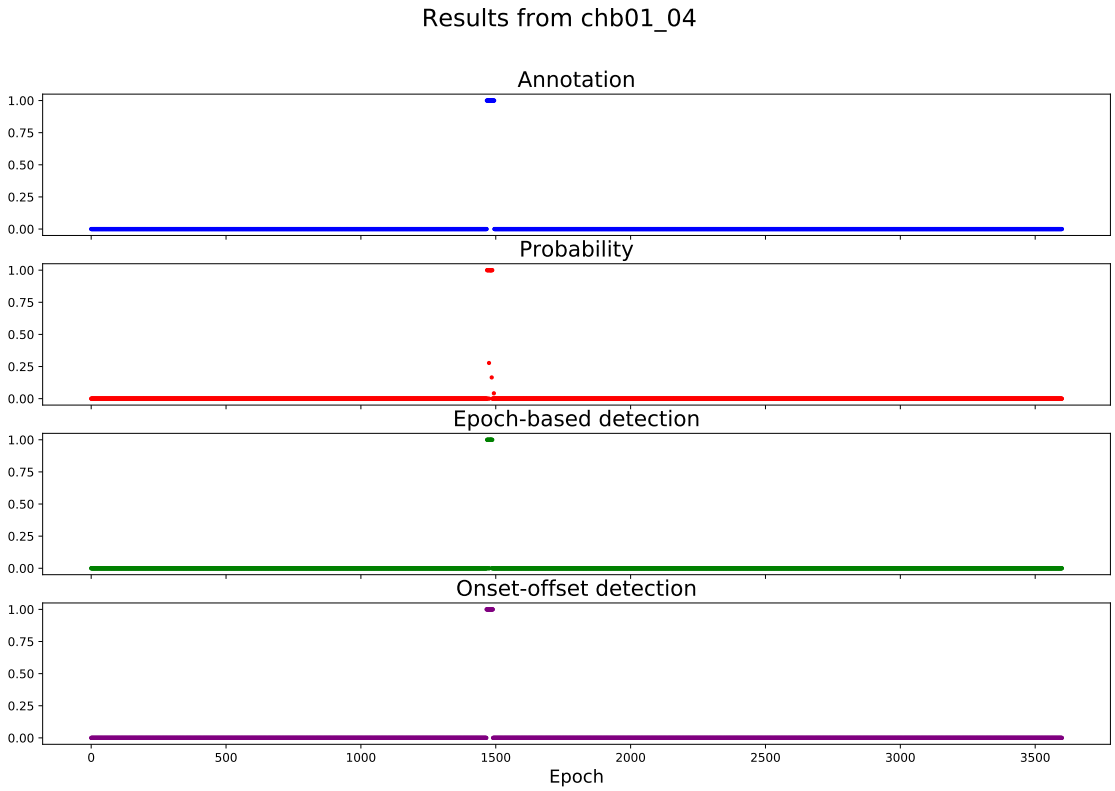
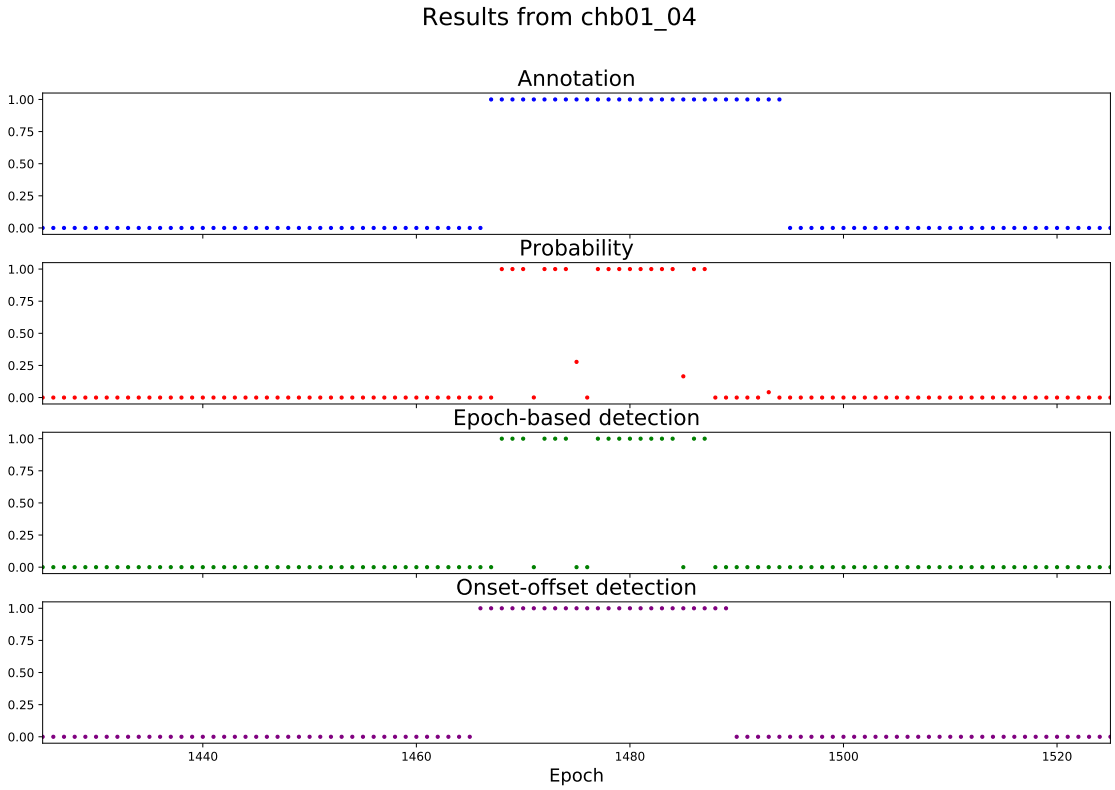



Figure 18: GDR obtained from before and after the onset-offset detection.

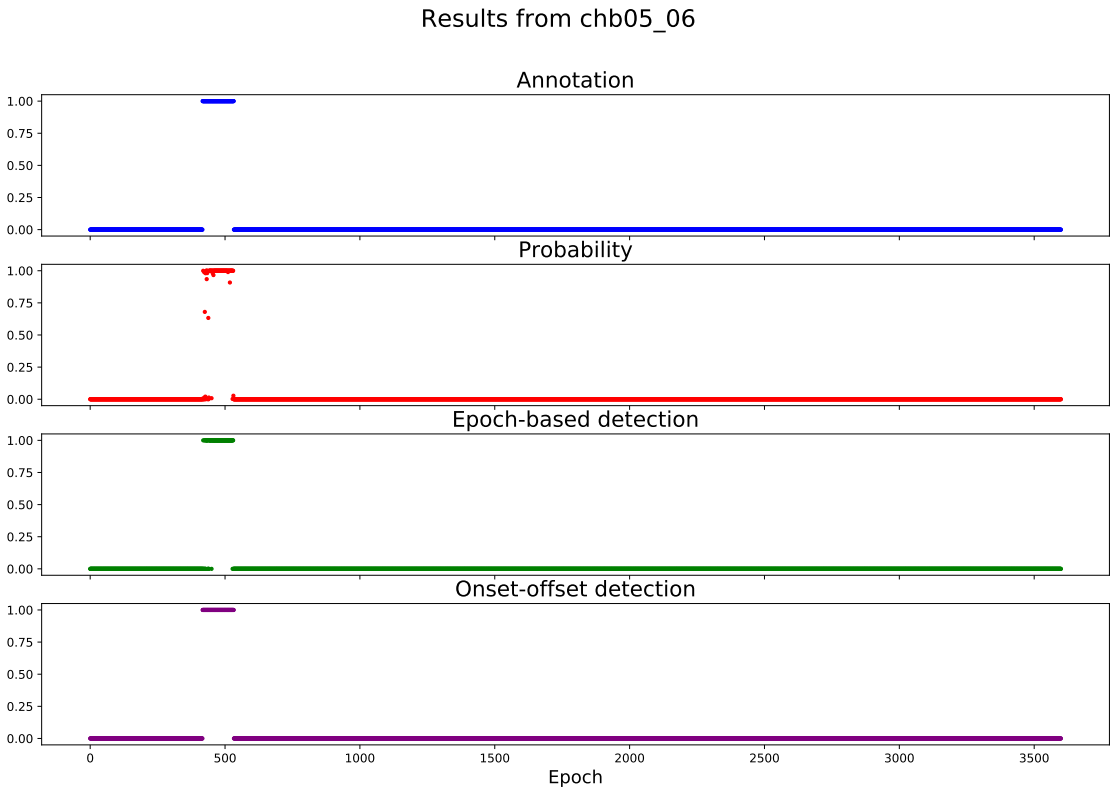



(a) Full record.

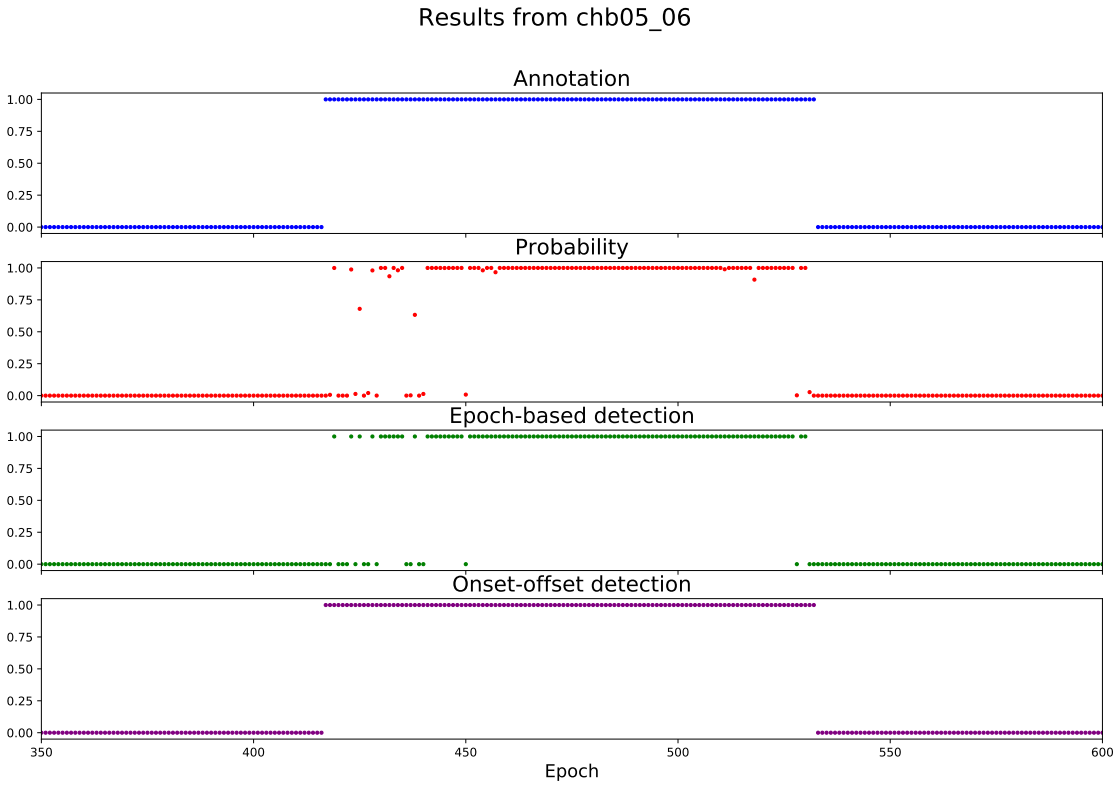



(b) During the seizure event.

Figure 19: Results of the seizure onset-offset detection process tested on `chb01_04`. The first row shows the annotation, the second row displays the seizure probability, the third one presents the decision of each epoch, and the last one demonstrates the result after the onset-offset detection.

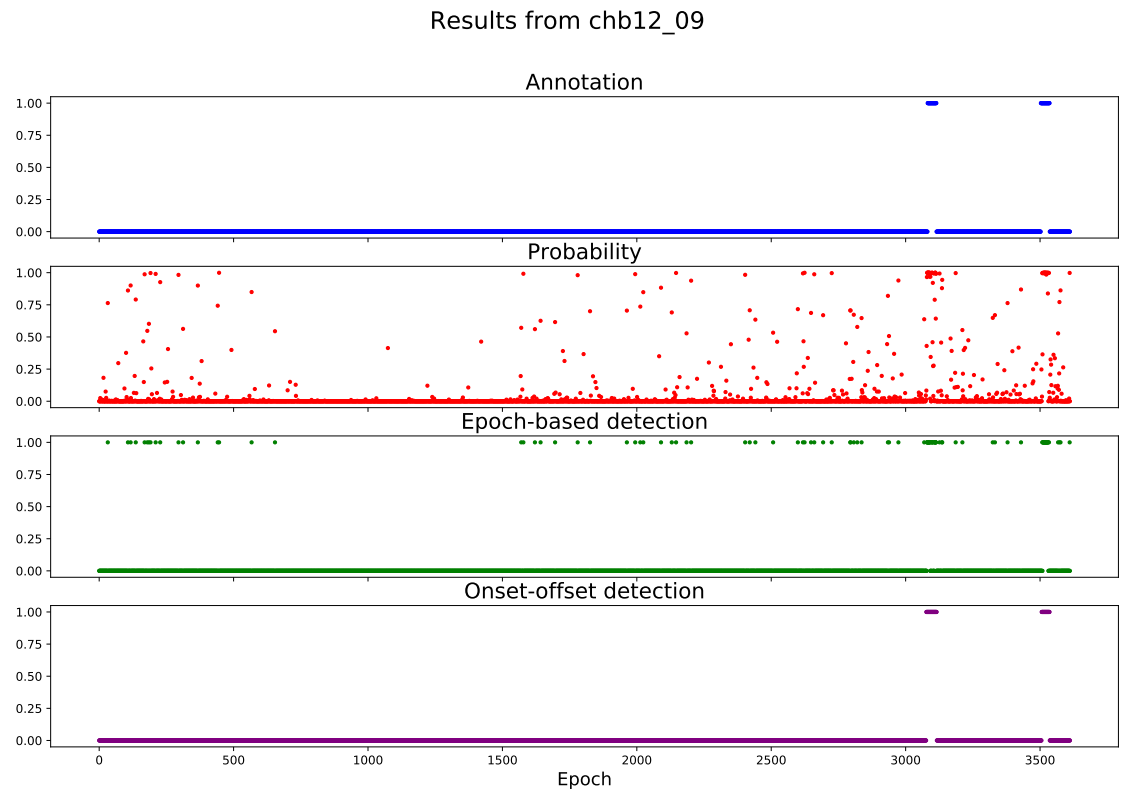



(a) Full record.

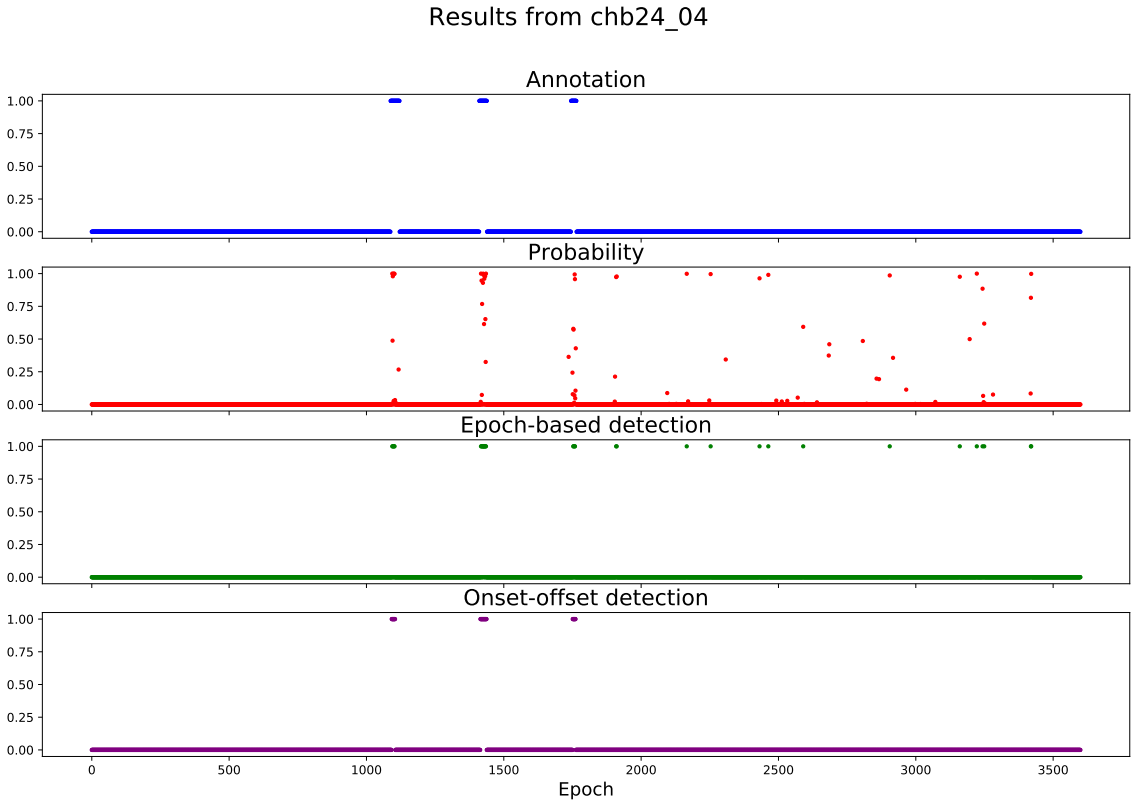



(b) During the seizure event.

Figure 20: Results of the seizure onset-offset detection process tested on `chb05_06`. All normal epochs during the seizure event are converted to epochs containing seizures.

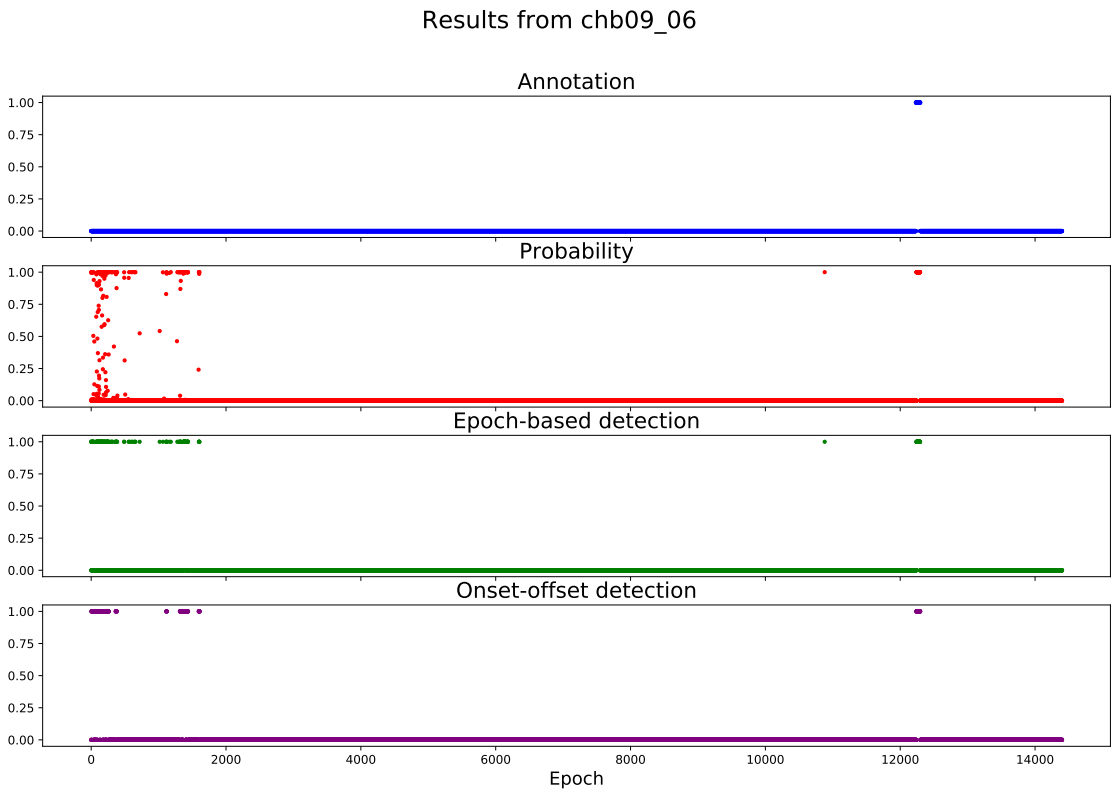



(a) Tested on `chb12_09`.

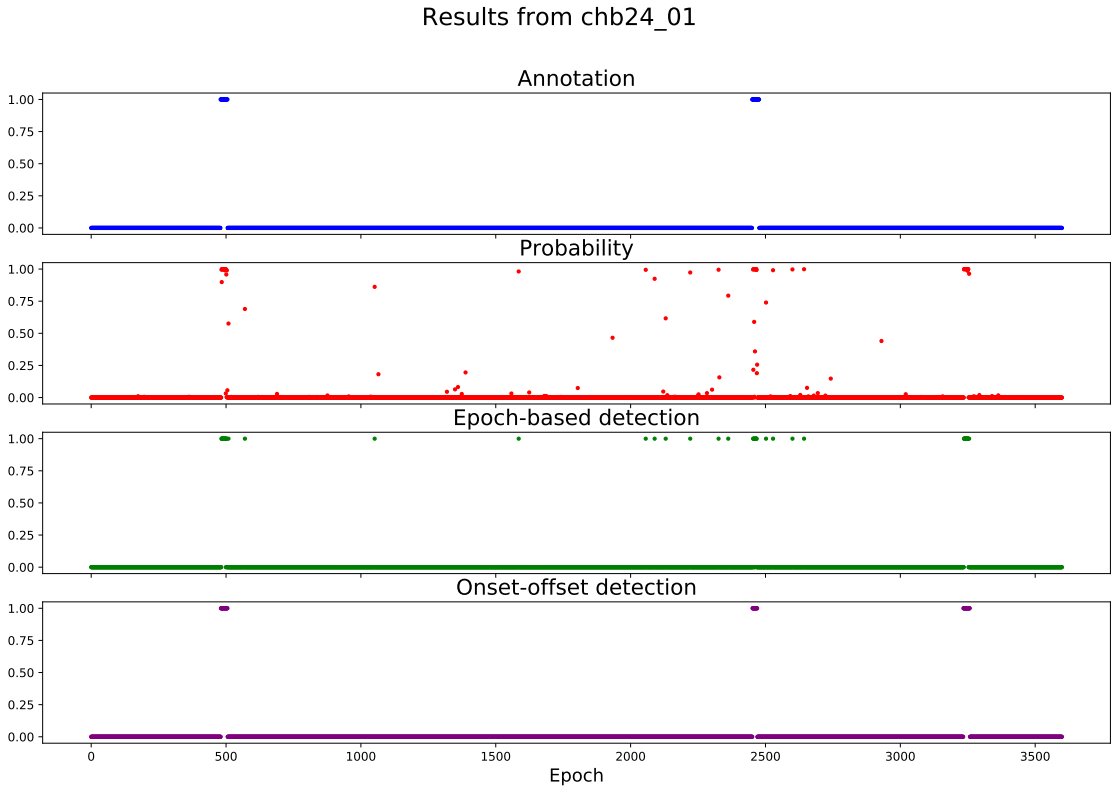



(b) Tested on `chb24_04`.

Figure 21: Examples of using the seizure onset-offset detection process. The output of the onset-offset detector has tremendously less false positives than that of the CNN, extremely reducing FPR/h.

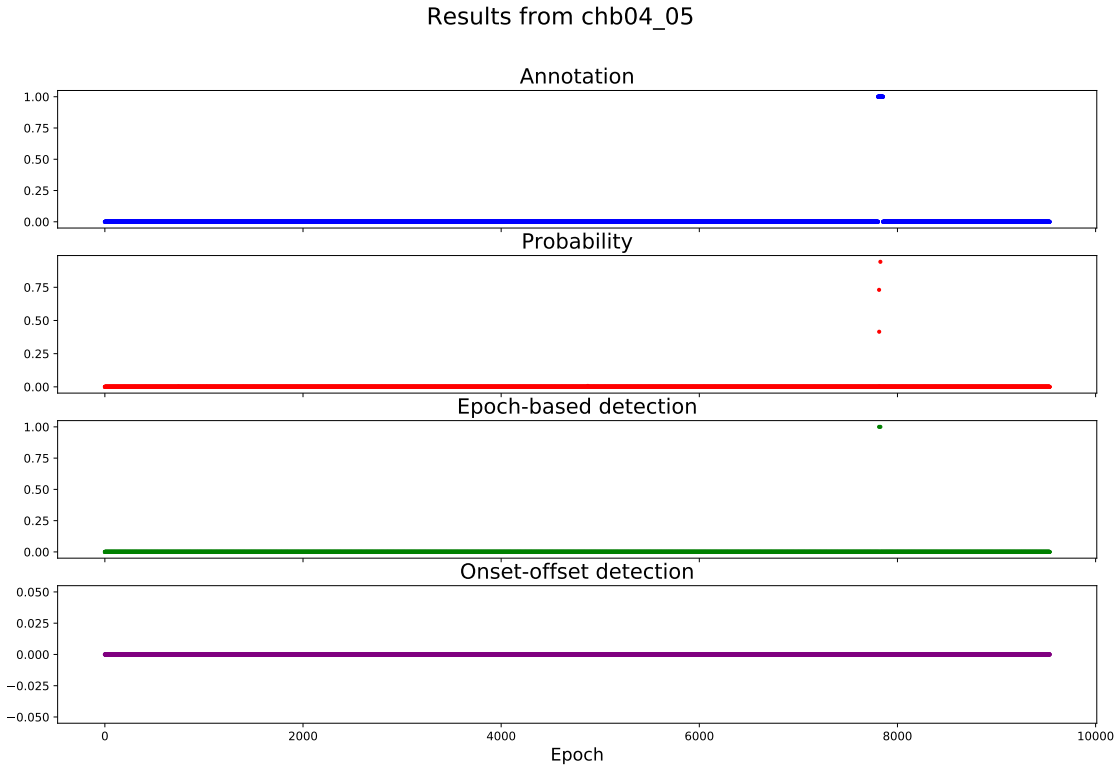



(a) Tested on `chb09_06`.

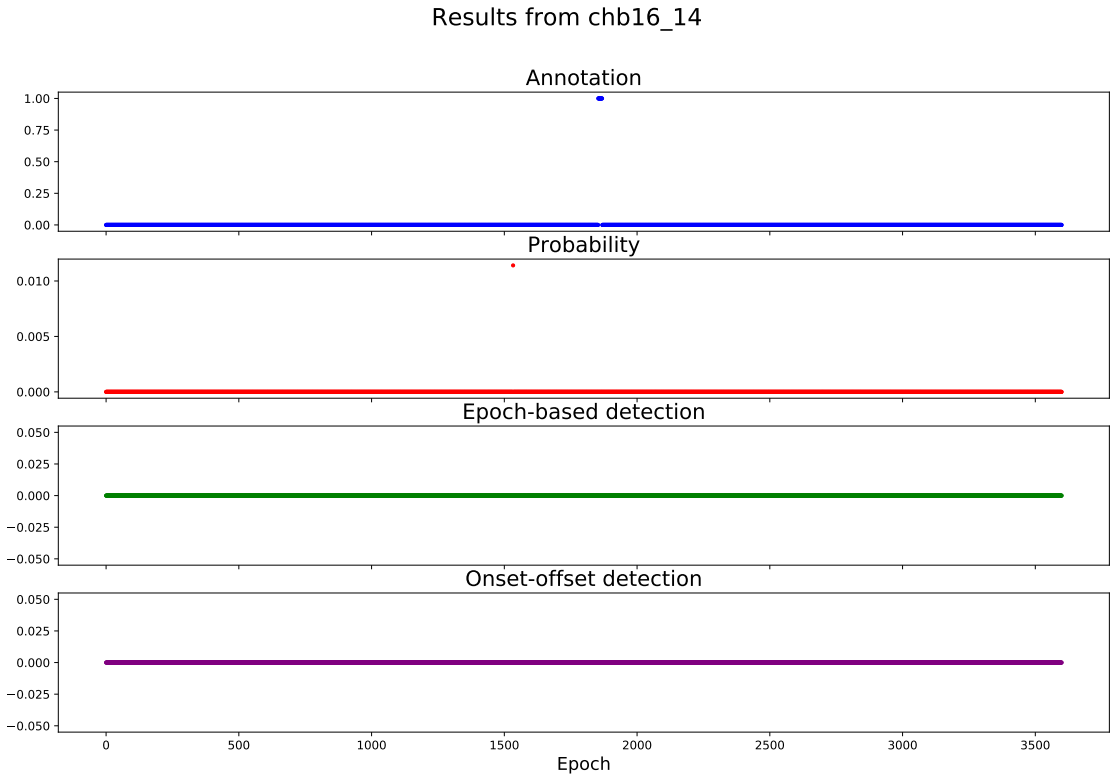



(b) Tested on `chb24_01`.

Figure 22: Examples of using the seizure onset-offset detection process. Even though the final result is improved from applying the onset-offset detection, the final result is still unsatisfactory if the classification output is terrible.

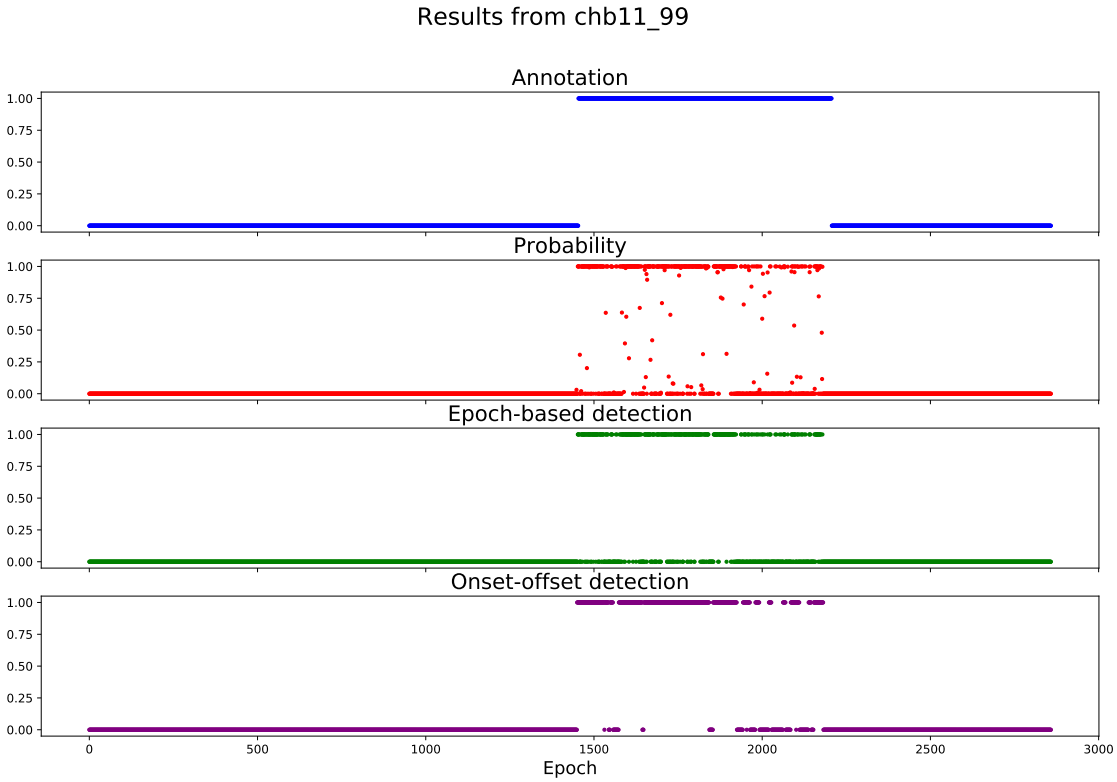



(a) Tested on `chb04_05`.

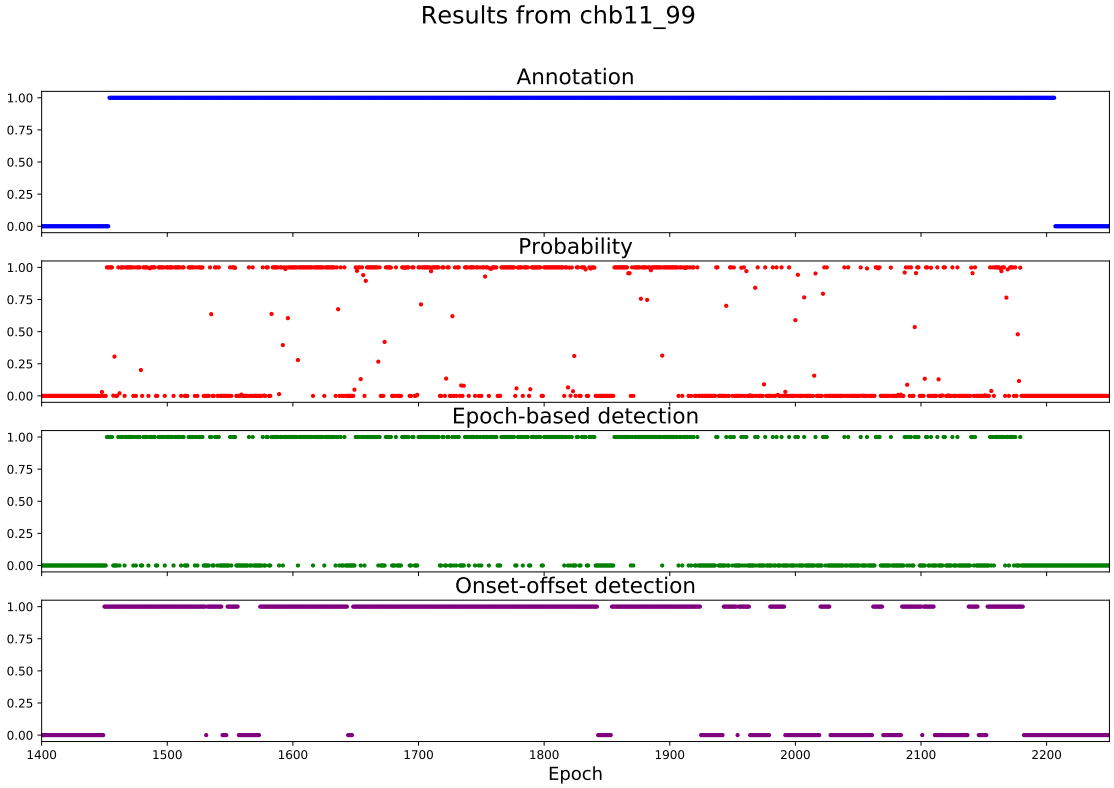



(b) Tested on `chb16_14`.

Figure 23: Examples of using the seizure onset-offset detection process. There are only few TP appearing. So the GDR is reduced in this case because there are few epochs correctly detected as ictal. Those epochs are neglected by the onset-offset detector.




(a) Full record.




(b) During the seizure event.

Figure 24: Results of the seizure onset-offset detection process tested on `chb11_99`. The onset-offset detector gives an unreliable output.

# 10 Conclusion and future work

This thesis aims to provide a detection method of epileptic seizures and the onsets and offsets in multi-channel EEG signals. The whole detection process consisted of two steps: epoch-based classification and onset-offset determination. First, we exploited a deep CNN model as a classifier to detect the seizures of epochs from the EEG signals. The deep CNN was designed to capture temporal and spatial information in multi-channel EEG segments. In a convolutional layer, the concept of filter decomposition was employed to extract features along the temporal dimension and observe feature connections between channels. Second, an onset-offset detector was proposed to indicate the onsets and offsets of the seizure events. The sequence of the epoch-based detection output was simply modified based on knowledge of epilepsy that the seizure clinically appears for some period. The CHB-MIT Scalp EEG database was used to evaluate the classification performances. All records of every subject were modified to have the same montage applied to the model.

As a result, the onset-offset detector could potentially reduce FPR/h and significantly increase sensitivity and $F_1$ while maintaining specificity and accuracy. The results obtained from the output of the onset-offset detection were more reasonable than the results from the epoch-based detection output. Moreover, with more imbalanced data set, our proposed model could determine the seizure onset and offset with ranges of latencies similar to that in other studies.

However, there are some limitations of this work. First, when the CNN model is not well-performed enough, as shown in Section 9.2, the onset-offset detector cannot improve, or sometimes reduce, the performances. From the design of the CNN, we tried to reduce the number of model parameters while maintaining the performance evaluated on `chb24`. It is possible that the model fit to this case may not be suitable in other cases. So, in the future work, we need to explore another deep learning model that is more suitable in this application or apply a technique for imbalanced data set. Second, the onset-offset detector is criteria-based but not adaptive; it requires well-adjusted parameters. This criteria-based scheme is not practical in clinic since conditions are varied depending on each subject. Therefore, we will establish the onset-offset detector based on a machine learning technique that can be adjusted regarding the patient, and mathematically analyze the technique.

# References


[AESAA15] T. Alotaiby, F.E.A. El-Samie, S.A. Alshebeili, and I. Ahmad. A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing*, 2015(1):66–86, 2015.

[AFS+15] U.R. Acharya, H. Fujita, V.K. Sudarshan, S. Bhat, and J.E.W. Koh. Application of entropies for automated diagnosis of epilepsy using EEG signals: A review. *Knowledge-Based Systems*, 88:85–96, 2015.

[AG99] R. Agarwal and J. Gotman. Adaptive segmentation of electroencephalographic data using a nonlinear energy operator. In *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*, volume 4, pages 199–202. IEEE, 1999.

[AKS18] E. Alickovic, J. Kevric, and A. Subasi. Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomedical Signal Processing and Control*, 39:94–102, 2018.

[AMS+12] U.R. Acharya, F. Molinari, S.V. Sree, S. Chattopadhyay, K. Ng, and J.S. Suri. Automated diagnosis of epileptic EEG using entropies. *Biomedical Signal Processing and Control*, 7(4):401–408, 2012.

[AS16] S. Ammar and M. Senouci. Seizure detection with single-channel EEG using extreme learning machine. In *Proceedings of the 17th International Conference on Sciences and*

*Techniques of Automatic Control and Computer Engineering*, pages 776–779. IEEE, 2016.

[ASS+13] U.R. Acharya, S.V. Sree, G. Swapna, R.J. Martis, and J.S. Suri. Automated EEG analysis of epilepsy: A review. *Knowledge-Based Systems*, 45:147–165, 2013.

[ASSK16] M.Z. Ahmad, M. Saeed, S. Saleem, and A.M. Kamboh. Seizure detection using EEG: A survey of different techniques. In *Proceedings of the 2016 International Conference on Emerging Technologies*, pages 1–6. IEEE, 2016.

[ATY+19] Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3):292–358, 2019.

[AWG06] A. Aarabi, F. Wallois, and R. Grebe. Automated neonatal seizure detection: a multistage classification system through feature selection based on relevance and redundancy analysis. *Clinical Neurophysiology*, 117(2):328–340, 2006.

[BLuCS19a] P. Boonyakitanont, A. Lek-uthai, K. Chomtho, and J. Songsiri. A comparison of deep neural networks for seizure detection in EEG signals. *bioRxiv*, page 702654, 2019.

[BLuCS19b] P. Boonyakitanont, A. Lek-uthai, K. Chomtho, and J. Songsiri. A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. *arXiv preprint arXiv:1908.00492*, 2019.

[BR07] A.S. Blum and S.B. Rutkove. *The Clinical Neurophysiology Primer*, volume 388. Springer, 2007.

[BYL84] W.T. Blume, G.B. Young, and J.F. Lemieux. EEG morphology of partial epileptic seizures. *Electroencephalography and Clinical Neurophysiology*, 57(4):295–302, 1984.

[CCS+18] S. Chakraborti, A. Choudhary, A. Singh, R. Kumar, and A. Swetapadma. A machine learning based method to detect epilepsy. *International Journal of Information Technology*, pages 1–7, 2018.

[CODL15] A. G. Correa, L. Orosco, P. Diez, and E. Laciar. Automatic detection of epileptic seizures in long-term EEG records. *Computers in Biology and Medicine*, 57:66–73, 2015.

[CUFK19] G. Chandel, P. Upadhyaya, O. Farooq, and Y.U. Khan. Detection of seizure event and its onset/offset using orthonormal triadic wavelet based features. *IRBM*, 40(2):103–112, 2019.

[FAA+14] R.S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J.H. Cross, C.E. Elger, J. Jr Engel, L. Forsgren, J.A. French, M. Glynn, et al. ILAE official report: a practical clinical definition of epilepsy. *Epilepsia*, 55(4):475–482, 2014.

[FCD+17] R.S. Fisher, J.H. Cross, C. D'souza, J.A. French, S.R. Haut, N. Higurashi, E. Hirsch, F.E. Jansen, L. Lagae, S.L. Moshé, et al. Instruction manual for the ILAE 2017 operational classification of seizure types. *Epilepsia*, 58(4):531–542, 2017.

[FHH+16] P. Fergus, A. Hussain, D. Hignett, D. Al-Jumeily, K. Abdel-Aziz, and H. Hamdan. A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics*, 12(1):70–89, 2016.

[FHT01] J. Friedman, T. Hastie, and R. Tibshirani. *The Elements of Statistical Learning*, volume 1. Springer Series in Statistics New York, 2001.

[GAG+00] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C. Peng, and H.E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23):e215–e220, 2000.

[Gel18] E. B. Geller. Responsive neurostimulation: Review of clinical trials and insights into focal epilepsy. *Epilepsy & Behavior*, 2018.

[GG05] S. Grewal and J. Gotman. An automatic warning system for epileptic seizures recorded on intracerebral EEGs. *Clinical Neurophysiology*, 116(10):2460–2472, 2005.

[Gol10] M.M. Goldenberg. Overview of drugs used for epilepsy and seizures: etiology, diagnosis, and treatment. *Pharmacy and Therapeutics*, 35(7):392–415, 2010.

[Got82] J. Gotman. Automatic recognition of epileptic seizures in the EEG. *Electroencephalography and Clinical Neurophysiology*, 54(5):530–540, 1982.

[GRD+10] L. Guo, D. Rivero, J. Dorado, J.R. Rabunal, and A. Pazos. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *Journal of Neuroscience Methods*, 191(1):101–109, 2010.

[HS97] M.A. Hall and L.A. Smith. Feature subset selection: a correlation based filter approach. In *Proceedings of the 4th International Conference on Neural Information Processing and Intelligent Information Systems*, pages 855–858. Springer, 1997.

[HSW89] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[IS15] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 448–456, 2015.

[Jan17a] S. Janjarasjitt. Epileptic seizure classifications of single-channel scalp EEG data using wavelet-based features and svm. *Medical & Biological Engineering & Computing*, 55(10):1743–1761, 2017.

[Jan17b] S. Janjarasjitt. Performance of epileptic single-channel scalp EEG classifications using single wavelet-based features. *Australasian Physical & Engineering Sciences in Medicine*, 40(1):57–67, 2017.

[JK19] J. M. Johnson and T. M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):27–60, 2019.

[LBH15] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[LYLO14] J. Li, J. Yan, X. Liu, and G. Ouyang. Using permutation entropy to measure the changes in EEG signals during absence seizures. *Entropy*, 16(6):3049–3061, 2014.

[MFF+13] M. Mastrangelo, I. Fiocchi, P. Fontana, G. Gorgone, G. Lista, and V. Belcastro. Acute neonatal encephalopathy and seizures recurrence: a combined aEEG/EEG study. *Seizure*, 22(9):703–707, 2013.

[MP95] J. Malmivuo and R. Plonsey. *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.

[MR18] B. B. Ma and V. R. Rao. Responsive neurostimulation: candidates and considerations. *Epilepsy & Behavior*, 2018.

[NFBA13] R. Nardou, D.C. Ferrari, and Y. Ben-Ari. Mechanisms and effects of seizures in the immature brain. In *Seminars in Fetal and Neonatal Medicine*, volume 18, pages 175–184. Elsevier, 2013.

[NH10] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning*, pages 807–814, 2010.

[OCDL16] L. Orosco, A.G. Correa, P. Diez, and E. Laciar. Patient non-specific algorithm for seizures detection in scalp EEG. *Computers in Biology and Medicine*, 71:128–134, 2016.

[OLC+09] L. Orosco, E. Laciar, A.G Correa, A. Torres, and J.P. Graffigna. An epileptic seizures detection algorithm based on the empirical mode decomposition of EEG. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2651–2654. IEEE, 2009.

[OP16] I. Obeid and J. Picone. The temple university hospital EEG data corpus. *Frontiers in Neuroscience*, 10:196, 2016.

[Pow11] D.M. Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Techonologies*, 2:37–63, 2011.

[PPCE92] F. Pauri, F. Pierelli, G. Chatrian, and W.W. Erdly. Long-term EEG-video-audio monitoring: computer detection of focal EEG seizure patterns. *Electroencephalography and Clinical Neurophysiology*, 82(1):1–9, 1992.

[RT03] A.J. Rowan and E. Tolunsky. *A primer of EEG: with a mini-atlas*. Butterworth-Heinemann Medical, 2003.

[SB10] N.K. So and W.T. Blume. The postictal EEG. *Epilepsy & Behavior*, 19(2):121–126, 2010.

[SEC+04] A.H. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S.T. Treves, and J. Guttag. Patient-specific seizure onset detection. *Epilepsy & Behavior*, 5(4):483–498, 2004.

[SG10a] A.H. Shoeb and J. Guttag. Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on Machine Learning*, pages 975–982, 2010.

[SG10b] A. Subasi and M.I. Gursoy. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications*, 37(12):8659–8666, 2010.

[SHK+14] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[Sho09] A.H. Shoeb. *Application of machine learning to epileptic seizure onset detection and treatment*. PhD thesis, Massachusetts Institute of Technology, 2009.

[SKS+11] A.H. Shoeb, A. Kharbouch, J. Soegaard, S. Schachter, and J. Guttag. An algorithm for detecting seizure termination in scalp EEG. In *Proceedings of 2011 the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1443–1446. IEEE, 2011.

[SLUC15] C. Satirasethawong, A. Lek-Uthai, and K. Chomtho. Amplitude-integrated EEG processing and its performance for automatic seizure detection. In *Proceedings of the 2015 IEEE International Conference on Signal and Image Processing Applications*, pages 551–556. IEEE, 2015.

[SRS+19] V. Sridevi, M.R. Reddy, K. Srinivasan, K. Radhakrishnan, C. Rathore, and D.S. Nayak. Improved patient-independent system for detection of electrical onset of seizures. *Journal of Clinical Neurophysiology*, 36(1):14–24, 2019.

[SS97] S.C. Schachter and D.L. Schomer. *The Comprehensive Evaluation and Treatment of Epilepsy: a practical guide*. Elsevier, 1997.

[SVI+16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[SvWL+18] 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. *Frontiers in Neuroinformatics*, 12:83, 2018.

[SZ15] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[TBB+11] D.J. Thurman, E. Beghi, C.E. Begley, A.T. Berg, J.R. Buchhalter, D. Ding, D.C. Hesdorffer, W.A. Hauser, L. Kazis, R. Kobau, et al. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*, 52(7):2–26, 2011.

[TTM+11a] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan. EEG-based neonatal seizure detection with support vector machines. *Clinical Neurophysiology*, 122(3):464–473, 2011.

[TTM+11b] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G.B. Boylan. Performance assessment for EEG-based neonatal seizure detectors. *Clinical Neurophysiology*, 122(3):474–482, 2011.

[TYK16] N.S. Tawfik, S.M. Youssef, and M. Kholief. A hybrid automated detection of epileptic seizures in EEG records. *Computers & Electrical Engineering*, 53:177–190, 2016.

[VI17] L.S. Vidyaratne and K.M. Iftekharuddin. Real-time epileptic seizure detection using EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11):2146–2156, 2017.

[VRB10] K. Vonck, R. Raedt, and P. Boon. Vagus nerve stimulation and the postictal state. *Epilepsy & Behavior*, 19(2):182–185, 2010.

[XHL16] B. Xu, R. Huang, and M. Li. Revise saturated activation functions. *arXiv preprint arXiv:1602.05980*, 2016.

[YXJZ17] Y. Yuan, G. Xun, K. Jia, and A. Zhang. A multi-view deep learning method for epileptic seizure detection using short-time fourier transform. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 213–222. ACM, 2017.

[Zei12] M. D. Zeiler. ADADELTA: An adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.