EPILEPTIC SEIZURE PREDICTION USING MACHINE LEARNING AND DEEP LEARNING METHODS

by

Burak Gözütok

B.S., in Electrical and Electronics Engineering, Istanbul Sehir University, 2019

B.S., in Computer Engineering, Istanbul Sehir University, 2019

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University

2021

ACKNOWLEDGMENTS

First of all, I would like to thank my thesis advisor, Prof. Dr. Ahmet Ademoğlu, for his guidance and support in my studies. His academic and personal contributions to me are invaluable.

I would also like to thank my colleagues for the time we shared in our Neurosignal Analysis Laboratory. Also, I would like to thank all my other professors at Boğaziçi University for inspiring me.

Finally, I would like to thank my family. Their support has always been with me throughout my education life. I could not have completed this thesis without their sincere support.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Burak Gözütok, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name : Signature:

Date:

ABSTRACT

EPILEPTIC SEIZURE PREDICTION USING MACHINE LEARNING AND DEEP LEARNING METHODS

Epilepsy is one of the most common neurological diseases in the world which negatively affects the daily life of a patient. Predicting epileptic seizures is of great importance for healthcare professionals and patients. The electroencephalography (EEG), which allows for registering brain activity with the help of electrodes placed on the scalp, is generally used to diagnose and monitor epilepsy. In this study, automatic seizure prediction was performed using CHB-MIT dataset which contains EEG data recorded at Boston Children's Hospital. Support Vector Machines (SVM), a common machine learning algorithm chosen as the primary method within this thesis's scope, and three different deep learning methods were compared. The first of these methods was long short term memory (LSTM) classifier with convolutional autoencoder which did not need any feature extraction. The second method used the spectrograms obtained by preprocessing the EEG data which were fed into a convolutional neural network (CNN) based classifier. The last method was based on converting the EEG data into three-dimensional images by applying source localization and performing classification with CNN. Among the methods used, the best result was obtained using source localization based CNN classification with 89.06% specificity, 92.58% sensitivity and 90.41% accuracy. Computational cost of three methods in terms of runtime efficiency were also compared, and it was observed that the SVM, which yielded the lowest classification performance with 74.07% accuracy, worked significantly faster than other methods.

Keywords: EEG, Epilepsy, Seizure Prediction, Deep Learning, Autoencoder, CNN

ÖZET

MAKİNE ÖĞRENİMİ VE DERİN ÖĞRENME YÖNTEMLERİYLE EPİLEPTİK NÖBET TAHMİNİ

Epilepsi, dünya üzerinde en yaygın nörolojik hastalıklardan biridir ve hastaların günlük yaşamlarını yakından etkilemektedir. Epileptik nöbetlerin önceden tahmininin yapılabilmesi sağlık çalışanları ve epilepsi hastaları için ciddi önem taşımaktadır. Epilepsinin teşhisi ve izlenmesinde genellikle kafa derisi üzerine yerleştirilen elektrotlar yardımıyla beyin aktivitesinin izlenmesini sağlayan Elektroensefalografi (EEG) yöntemi kullanılmaktadır. Bu çalışmada, Boston Çocuk Hastanesinde kayıt altına alınan EEG verilerini içeren CHB-MIT verisi kullanılarak otomatik nöbet tahminlemesi yapılmıştır. Çalışma kapsamında birincil yöntem olarak bir makine öğrenmesi türü olan SVM seçilmiş ve bununla beraber üç farklı derin öğrenme yöntemi kıyaslanmıştır. Bu yöntemlerin ilki, herhangi bir öznitelik çıkarmaya ihtiyaç duymayan Evrişimli Otokodlayıcı girdili LSTM sınıflandırıcıdır. İkincisi, EEG verisinin ön işlenmesi ile spektrogramlarının elde edilip daha sonrasında Evrişimsel sinir ağı (CNN) temelli bir sınıflandırıcı ile ele alındığı yöntemdir. Denenen son metod ise, EEG verilerine kaynak yerelleştirme uygulayarak üç boyutlu kayıtlara dönüştürmek ve bunlar üzerinden CNN ile sınıflandırma gerçekleştirmektir. Kullanılan yöntemler arasında en iyi sonuç 89.06% özgüllük, 92.58% duyarlılık ve 90.41% doğruluk ile kaynak yerelleştirme bazlı CNN Sınıflandırma kullanılarak elde edilmiştir. Çalışma kapsamında ayrıca yöntemlerin çalışma süreleri kıyaslanmış ve 74.07% doğruluk ile en düşük sonucu veren SVM'in diğer yöntemlere kıyasla ciddi oranda hızlı çalıştığı görülmüştür.

Anahtar Sözcükler: EEG, Epilepsi, Nöbet Tahminleme, Derin Öğrenme, Otokodlayıcı, CNN

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$tr(\beta)$	Trace of β
β^T	Transpose of β
$x(i)$	(i)th element of vector x
$\|x\|$	Euclidean norm of vector x
σ	Sigmoid function.
y_i	Prediction value for observation
$\hat{y}_i$	True observation
TW	Time Window Length
Δf	Frequency Resolution

LIST OF ABBREVIATIONS

EEG	Electroencephalography
ML	Machine Learning
DL	Deep Learning
1-D	One Dimensional
2-D	Two Dimensional
3-D	Three Dimensional
FIR	Finite Impulse Response
CSP	Common Spatial Pattern
SVM	Support Vector Machine
RNN	Recurrent Neural Network
LSTM	Long Short-term-memory
STFT	Short-Time Fourier Transform
CNN	Convolutional Neural Network
CAE	Convolutional Autoencoder
FFT	Fast Fourier Transform
DPSS	Discrete Prolate Spheroidal Sequences
MAE	Mean Absolute Error
LOOCV	Leave-one-out Cross-validation
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
FP	False Positive
TP	True Positive
FN	False Negative
TN	True Negative
FPR	False Positive Rate
TPR	True Positive Rate

1. INTRODUCTION

Electroencephalogram (EEG) is a method of recording the electrical potential of the brain through electrodes placed on the scalp [4]. EEG is a widely used method for detecting changes in brain activity, and abnormal EEG patterns can arise from different reasons, such as epilepsy, sleep disorders, and brain damage. The term "epileptic seizure" refers to a set of diseases characterized by the repeated discharge from the cerebral cortex, which causes abnormal brain activity [5]. An epileptic seizure comprising this abnormal brain activity is a symptom of epilepsy. Epilepsy is a crucial disease affecting approximately 1% of the world's population [6]. Neurologists usually analyze EEG by visual inspection. However, examining long EEG recordings is time-consuming and imposes a significant burden on neurologists. In this regard, systems that will facilitate the analysis of EEG signals or completely automate this process in detecting epilepsy are of great importance. According to [6], loss of consciousness caused by epileptic seizures, which might be ordinarily harmless, can be fatal when driving, crossing a busy street or swimming. Besides being a significant life-threatening disease, epilepsy can have a crucial psychological and social impact on patients due to unexpected seizures. For these reasons, in addition to detecting epileptic seizures, researchers also emphasize predictive methods. An early warning system, which can be created by predicting seizures, can enable healthcare professionals to take the necessary precautions for the patient. Additionally, detecting the incoming seizure by examining the EEG signals will also lead to neuroprostheses that will suppress the neural brain foci that cause the epileptic seizure with electrical signals.

EEG classification problems can be the subject of many studies for many purposes. They are separated by different titles, according to the method used and the problem solved. Besides seizure prediction, which is the subject of this study, there are several other problems to be solved using EEG such as motor imagery and sleep scoring. The methods used in this regard also differ based on the problem. A recent study [7], uses feature extraction with common spatial pattern (CSP) filters, which is similar to the machine learning approach employed in this thesis. With the widespread use of deep learning algorithms and the easier access to processing power, studies in this area have also increased. Additionally, deep learning algorithms also have many applications. For example, [8, 9, 10, 11] use CNN with deep learning. Of these, [8] prefers to visualize the EEG data with spectral maps and classifies them using those images, and the lowest error rate they obtained was with the CNN model that they have developed. The study in [9] also examined the problem of seizure detection by converting EEG data into 3D visuals and using CNN. The study in [10] performed classification by creating spectrogram images from EEG data. Another study [11] used the short time fourier transform (STFT) process for this visualization. Besides, LSTM method was applied in [2, 12, 13]. While [2] and [13] used raw data as input to LSTM, [12] extracted several EEG features from cross-correlation, time-domain, frequency-domain and graph theory and then input these attributes into LSTM. Studies in [14], and [15] applied the autoencoder method to classify epileptic EEG data. One of the seizure prediction studies [15] used a convolutional autoencoder, while [14] which was a seizure detection study, preferred the stacked autoencoder method. All these developed methods need to be applicable in real-time to achieve their goal. For this reason, [16], which performed classification with CNN, drew attention to this issue and discussed the running time cost of the method that they proposed.

In this study, different preprocessing methods and deep learning methods were employed to predict epileptic seizures from EEG data, and these were compared with a conventional machine learning method called SVM. The first deep learning method is an LSTM classifier based on convolutional autoencoder using raw data. The second method is to classify the data by the CNN method after preprocessing and estimating spectrogram images. Finally, CNN classifier was used to classify 3D source images obtained by source localization of EEG data.

The organization of the thesis is as follows: Chapter 2 gives general background information about seizure prediction, machine learning classification, deep learning classification, and EEG preprocessing. In Chapter 3, data, evaluation method and the models used in this study are explained in detail. Chapter 4 demonstrates the result of each classifier separately. In Chapter 5, the discussion of the results is presented. Finally, in Chapter 6 conclusions and future recommendations are presented.

2. BACKGROUND

2.1 Seizure Prediction

The uncertain timing of epileptic seizures makes the life of epilepsy patients difficult. In this regard, it is crucial to develop seizure prediction systems as a step ahead of seizure detection systems. An early warning system will make patients feel ready for seizures and prevent unpredictable, dangerous consequences. Epileptic EEG recordings are categorized into four different classes as interictal, preictal, ictal, and postictal [17]. Preictal state refers to the period just before the seizure. The ictal state refers to the seizure time, and the postictal state refers to the time following the epileptic seizure. The remaining non-seizure periods are called interictal. Epileptic seizure states are demonstrated in Figure 2.1. Rasekhi et al. define the seizure prediction problem as a binary classification problem between preictal and non-preictal states [18].

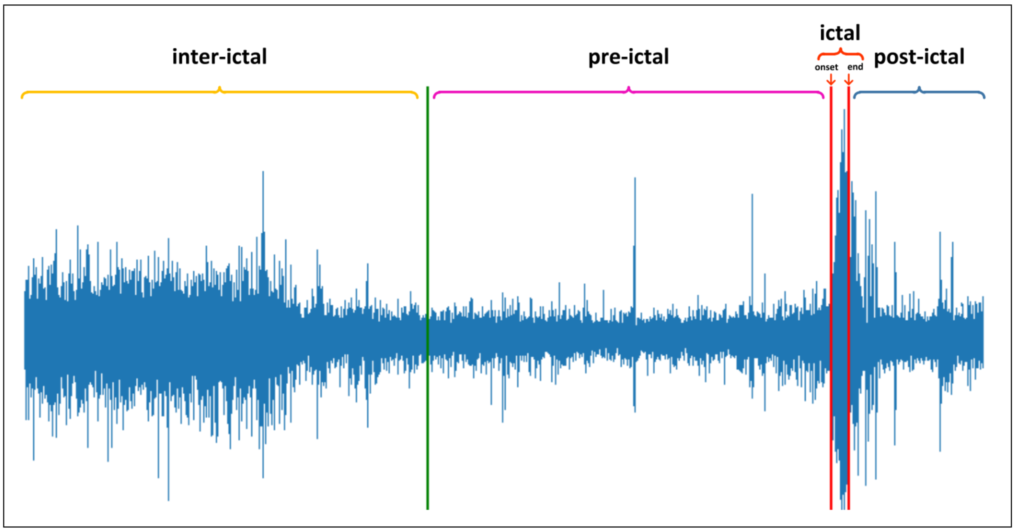

Figure 2.1 Epileptic seizure states [1].

One of the studies conducted in this area [19] states that the main weakness in seizure prediction studies is the lack of comprehensive testing in seizure-free baseline data and the failure to test the performance of the algorithm in different studies and to reduce it to the essential components of multichannel data. In addition, due to the differentiation of EEG activity from person to person, seizure prediction and detection are generally patient-specific [20].

2.2 Machine Learning Classification

Machine Learning is the modelling of systems with computers that make predictions by making inferences from data with mathematical and statistical operations. Machine learning models aim to generalize over the data given for training by employing all available data. Accordingly, it extracts the parameters of the algorithm by understanding the underlying structure of the training data. However, while doing this, it is necessary to confirm that the model does not choose the simple way of memorizing the training data (over-fitting) instead of learning the structure. Therefore, training and validation data should be separate, and the model trained with training data should be evaluated with validation data.

When using machine learning methods, pre-processing and feature extraction are usually applied to the data beforehand. In this way, unnecessary information is removed from the data, allowing for better generalization of the model with less effort. These pre-processing steps vary according to the problem and data type.

There are many machine learning algorithms in the literature, such as nearest neighbor, decision trees, naive bayes, linear regression and support vector machines. In this study, SVM was employed as the machine learning method to be compared with deep learning methods.

2.2.1 Support Vector Machines (SVM)

SVM [21] can be defined as a vector space based machine learning method that finds a decision boundary between the two classes that are far from any point in the training data. The SVM method can be used for both regression and classification problems.

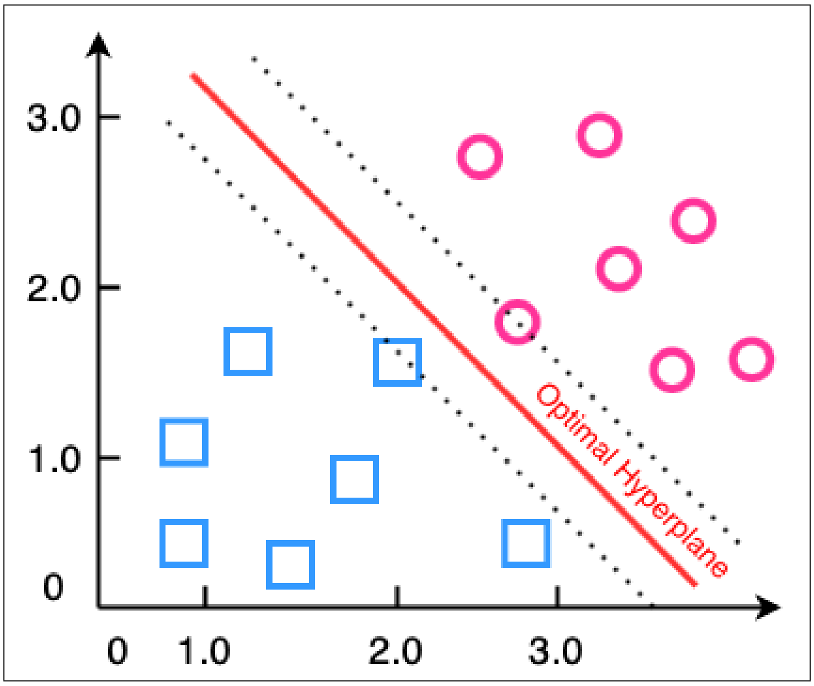

Figure 2.2 Optimal hyperplane using support vector machines.

SVM algorithm which is visualized in Figure 2.2, finds the closest points from both classes to the decision boundary, called support vectors. Distance between these support vectors is called the margin. The purpose of the SVM is to maximize the margin, and the decision boundary where it is maximum is called the optimal hyperplane. This optimal hyperplane created with the SVM algorithm is the separation boundary of the two classes, and new samples can be classified according to which direction they stand relative to the boundary. SVM problem is given by the following optimization problem:

$$Find \max_{w,b} M \tag{2.1}$$

$$s.t. \ \frac{1}{\|w\|} y_i(w.x_i + b) \geq M \geq 0 \ \forall x_i, i = \{1, ..., m\} \tag{2.2}$$

where m is the number of samples, $x \in R^n$ is the input vector, $w \in R^n$ is a vector of parameters (weight vector), $b \in R$ is a constant of hyperplane's equation, M is the orthogonal distance between the hyperplane and each training point and $y_i \in \{-1, 1\}$ indicates the class to which x_i belongs.

2.3 Deep Learning Classification

Deep learning is a field of study that covers artificial neural networks and similar machine learning algorithms with one or more hidden layers. Classical machine learning models are still far from what humans can do in complex problems with their requirements, such as feature extraction. Deep learning models simulate the function of neurons for the human brain and process information with multiple layers. While machine learning models require improvement with manual analysis of the results, deep learning can understand its prediction error and improve itself with the artificial neural network it contains during the training process. Furthermore, thanks to the complex learning structure of deep learning models, they can give successful results by working on the original data without feature extraction. Deep learning has various architectural types with different usage areas such as deep neural networks, recurrent neural networks, convolutional neural networks. Additionally, deep learning models can handle unsupervised and supervised training according to the problem type.

Deep learning is a method frequently used in the field of EEG classification, and it is used in topics such as emotion recognition, motor imagery, mental workload, seizure detection, sleep stage scoring [22].

2.3.1 Recurrent Neural Networks (RNN) and LSTM

RNN is a specialized artificial neural network structure that uses sequence and time data. The architecture of RNN in which inter-node connections are routed backwards allows previous outputs to be used as inputs. In this way, RNN can exhibit time-based behaviour. The structure of RNN is demonstrated in Figure 2.3. However, the vanishing gradient problem, which can be experienced when using RNNs, can

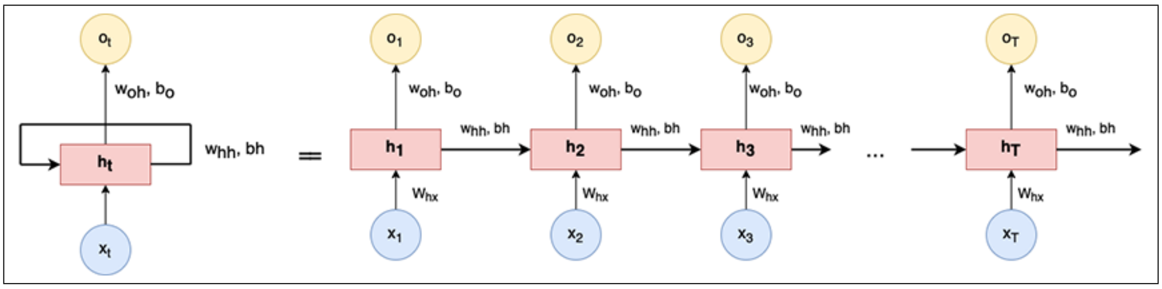

Figure 2.3 Structure of RNN.

weaken the model's success. Loss of weights with back-propagation is an unavoidable problem for RNN. Since chain rules set the weights in each layer, the gradient values will get exponentially smaller and closer to zero as the model steps back. As a solution to this problem, the LSTM model [23], which is a kind of RNN model, has been developed. LSTMs have several gates that control whether new information can be transferred to the next, and with these gates, the gradient flow is better controlled and long-term bonds are conserved. LSTM cell architecture is demonstrated in Figure 2.4. The input gate (i_t) determines what information is entered into the cell and its equations are given as:

$$\begin{aligned} i_t &= \sigma\ (W_i[h_{t-1}, x_t]\ +\ b_i) \\ \tilde{C}_t &= tanh\ (W_C\ .\ [h_{t-1}, x_t]\ +\ b_c) \end{aligned} \tag{2.3}$$

where t is timestamp, σ is sigmoid function, x_t is current input at time step t, h_{t-1} is previous hidden state, W_i is weight matrix of sigmoid operator between input and output gate, b_i and b_c are bias vectors, $\tilde{C}_t$ value generated by tanh and W_c weight matrix of tanh operator between cell state information and network output. The forget gate (f_t) decides what information should be ignored and what information should be kept and its equation is given as:

$$f_t = \sigma\ (W_f[h_{t-1}, x_t]\ +\ b_f) \tag{2.4}$$

where b_f is bias vector and W_f is weight matrix between forget gate and input gate. The output gate (o_t) is used to read the outputs from the cell and its equations are given as:

$$o_t = \sigma \left(W_o \, [h_{t-1}, x_t] + b_o \right)$$
$$h_t = o_t * tanh(C_t) \tag{2.5}$$

where b_o is bias vector and W_o is weight matrix of output gate. With its structure suitable for sequence data, LSTM is used frequently in the classification of EEG data which already contains time information.

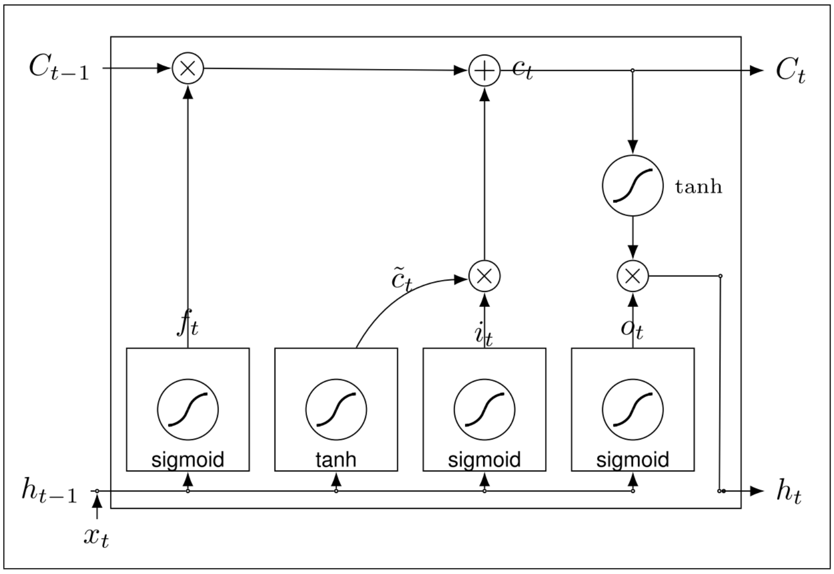

Figure 2.4 LSTM cell architecture [2].

2.3.2 Convolutional Neural Networks

CNN [24] is a deep learning algorithm that is generally used in image processing and takes images as input, usually consisting of convolution and pooling layers. Convolutional neural networks perform an automatic feature extraction with convolution operations on the image data. CNN architecture is demonstrated in Figure 2.5. EEG data when converted into images such as spectrograms, fourier feature maps and topographic maps can be fed into CNN for classification. According to a review by Craik et al. [22] 20% of the EEG classification studies convert data into image format to be as an input to the model. 41% of them classify with various feature calculations. Besides, some algorithms consider two dimensional EEG samples consisting of channels versus time and apply the CNN algorithm directly to the raw data [25].

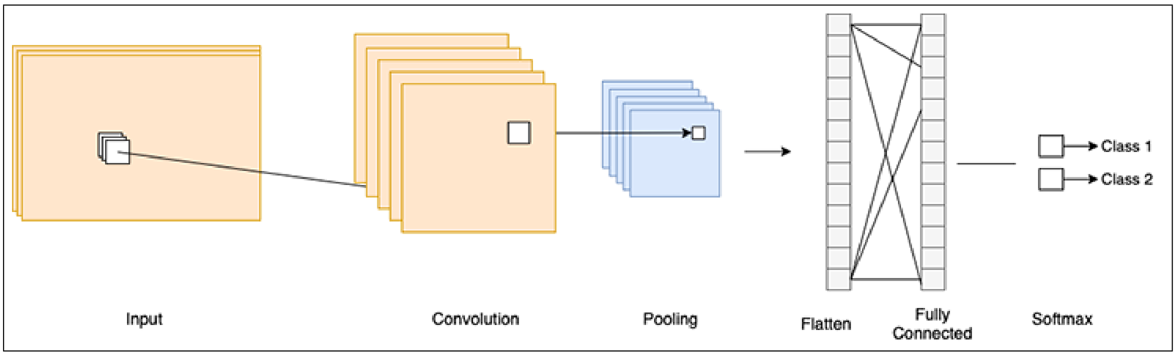

Figure 2.5 CNN Architecture.

2.3.3 Convolutional Autoencoders

Autoencoder is a kind of artificial neural network used for unsupervised learning. It is a structure that learns to reduce the training data to a small size representation and convert it back to the original from this small-sized sample. This two-way structure consists of an encoder and a decoder. An exemplary auto-encoder architecture is shown in Figure 2.6. Since these models aim to recover the original data from the low-dimensional representation, they learn to make the most lossless reduction and preserve important information. In this respect, autoencoders can be used as unsupervised feature extractors. It is also often used in size reduction problems, as it learns to obtain a small-sized representation of the data. Besides, CA is an autoencoder type in which CNN is used. Convolution and pooling layers play a role in the encoder part of CA, while deconvolution and unpooling layers are used in the decoder part. CA have many uses, from noise reduction [26] to feature extraction [27] on EEG.

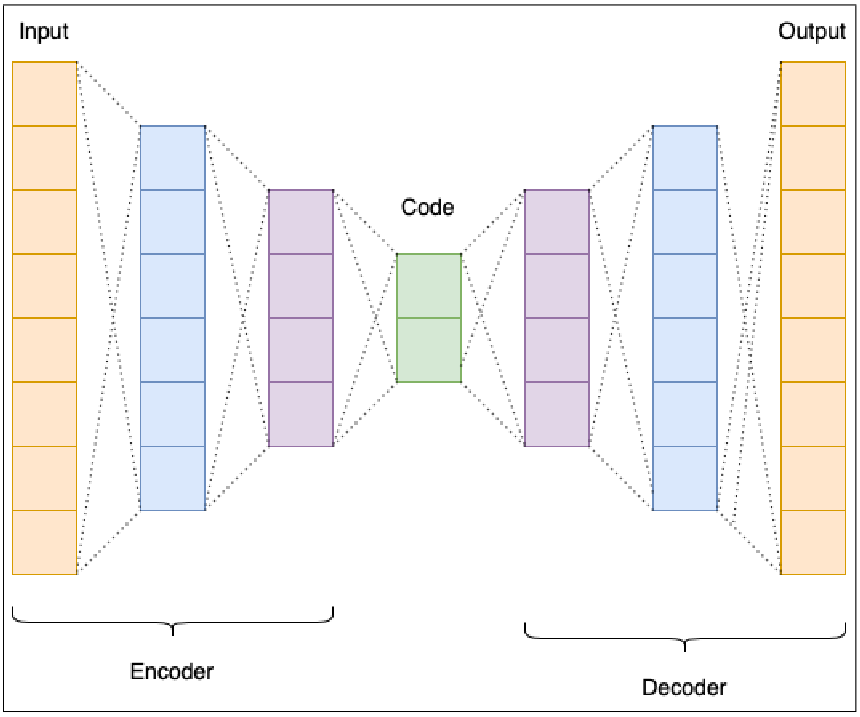

Figure 2.6 Auto-encoder example.

2.4 EEG Preprocessing and Feature Extraction

Although it is known that the main difference between deep learning algorithms and classical machine learning algorithms is that deep learning does not need feature extraction, feature extraction still can be used in deep learning to refine the model further. In addition, the application of pre-processing may also affect classification success. These pre-processing steps can be general or domain-specific, and there are several methods for EEG pre-processing. These can be ordinary methods such as bandpass filtering or they can be more complex as power spectrum, time-frequency analysis or parametric modeling.

2.4.1 Sliding Window Approach for EEG Inputs

EEG data containing seizure information can consist of different classes in a single recording and is hard to process in a single pass because it consists of long recordings (measured in hours) for each epileptic condition. For this reason, the sliding window approach is used, which divides the data into different windows in the time axis. These windows consist of two distinct variables as the length of the window (L) and percentage of overlap or displacement (D) [3]. Appropriate L and D may vary depending on the problem and data. For example, [15] used 5-second non-overlapping windows, the study [28] used 50% overlapping 10-second windows for preictal, 75% overlapping 10-second windows for interictal, and [29] used 30-second non-overlapping windows. The sliding window method is demonstrated in Figure 2.7.

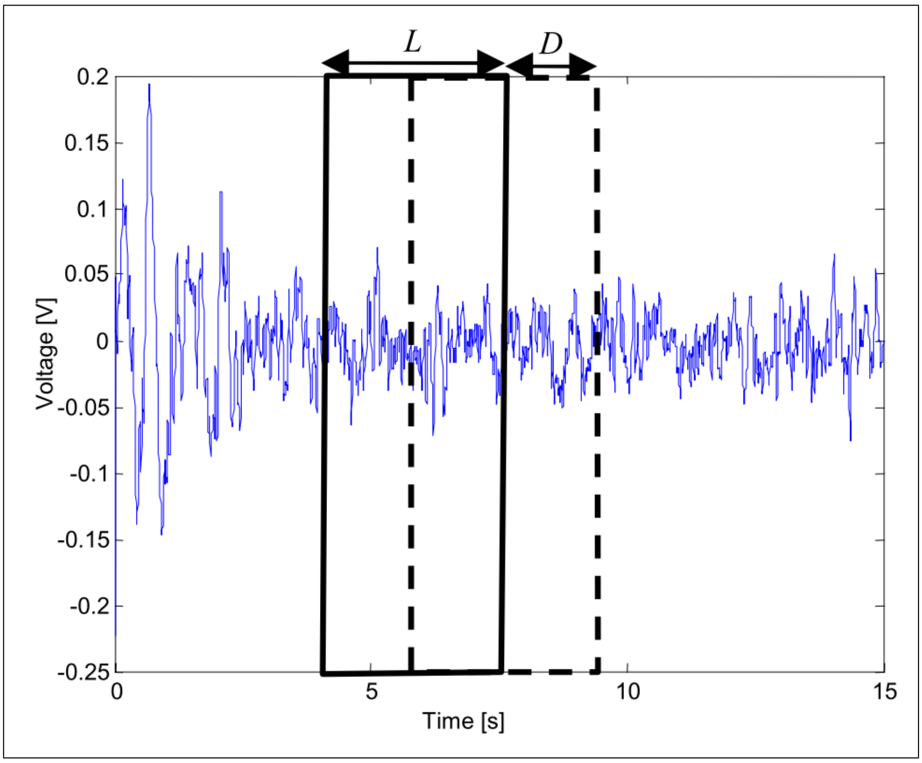

Figure 2.7 Sliding windows [3].

2.4.2 Bandpass Filtering

Bandpass filtering is an application of signal filtering that transmits frequencies within a specific range and attenuates frequencies outside that range. It is a widely used method to remove physiological artifacts and noise from EEG signals and to extract the frequency in a specific range of interest. The ideal frequency range may vary according to the problem to be solved. EEG frequency bands, which are categorized according to their functional characteristics, are generally used when choosing the frequency range.

These can be listed as Delta band (1-4 Hz), Theta Band (4-7 Hz), Alpha band (8-12 Hz), Beta band (12-30 Hz) and Gamma Band (30 - 100 Hz) [30].

2.4.3 Common Spatial Patterns

CSP [31] is a feature extraction method widely used in signal processing. It tries to maximize the distinguishability between two classes using spatial filters [32]. The CSP method splits the signal into additive subcomponents, and these components can be used as features. Let $X \in R^{NxT}$ denote the EEG signal, where N is the number of channels and T is the number of samples. Then, the spatial covariance matrix of trial can be represented as:

$$R = \frac{XX^T}{tr(XX^T)} \tag{2.6}$$

Then, composite spatial covariance can be calculated as:

$$R = \overline{R_1} + \overline{R_2} = U\,\Sigma\,U^T \tag{2.7}$$

where $\overline{R_1}$ and $\overline{R_2}$ denotes averaged normalized covariance over all trials of each group and U denotes the matrix of eigenvectors and Σ denotes the diagonal matrix of corresponding eigenvalues. The full projection matrix is formulated as:

$$W = B^T\,\Sigma^{-\frac{1}{2}}\,U^T \tag{2.8}$$

where B denotes the matrix of eigenvectors for whitened spatial covariance matrix. For features, the corresponding n eigenvector is used from the beginning and end when the eigenvalues are sorted. Then feature vector can be calculated as:

$$Z = W^T X \tag{2.9}$$

$$f_q = log\,(\frac{var(z_q)}{\sum_{i=1}^{2n} var(z_i)}) \tag{2.10}$$

where z_q indicates the q-th row vector of Z and f_q represents the $2n$ dimensional feature vector [33].

2.4.4 Spectrograms

The spectrogram is a visualization method that shows the time-dependent variation of power spectra at different frequencies. Spectrograms usually consist of two-dimensional (time x frequency) visual patterns that show the power variation over frequency and time with colour dimension. Spectrum, on the other hand, shows the power distribution of the frequency at a particular time. The continous and discrete power spectrum formula is:

$$PS(f) = \frac{1}{T}\int_0^T r_{xx}(t)e^{-j2\pi m f_1 t}dt \qquad m = \{0, 1, 2, 3...\} \tag{2.11}$$

$$PS[m] = \sum_{n=0}^{N-1} r_{xx}[n]e^{-\frac{j2\pi mn}{N}} \qquad m = \{0, 1, 2, 3...N\} \tag{2.12}$$

where $r_{xx}(t)$ and $r_{xx}[n]$ are autocorrelation functions [34]. For the spectrogram calculation, the signal is divided into windows of equal lengths and the spectrum is calculated for each window. The spectrogram can be obtained by displaying the magnitudes for each window with a colourmap on an image. The spectral analysis of actual data differs from the ideal spectrum due to the data's discrete, finite and aperiodic nature. A method known as multitaper spectrum [35] is used for the spectral estimation to be more accurate and close to the true spectrum. This method reduces estimation bias by taking an average of many independent estimations from the same sample. Also, all of the samples use "taper" functions, which reduces bias in single samples. The detailed application of this method is explained in the study [35].

2.4.5 EEG Source Localization

Source localization is the process of identifying from which part of the brain the activities in the brain originates. By using source localization, EEG signals can be converted to 3D source images. The location and intensities of each active source are determined by inverse problem solving using regularization or Bayesian methods [36]. If available as a head model, a patient-based MRI model is preferred, but if not, general models such as MNI brain can also be used [37]. Minimum norm estimation is used here as an inverse method for projecting the sensor data measured from the scalp electrodes to the source space locations defined on the tessellated MNI cortex model.

In the brain maps produced by this method, the brain electrical activity can be seen, and even the differences between preictal and interictal stages can be observed for some cases. The images of preictal and interictal windows of Subject 1 in different time windows estimated by this method are demonstrated in Figure 2.8.

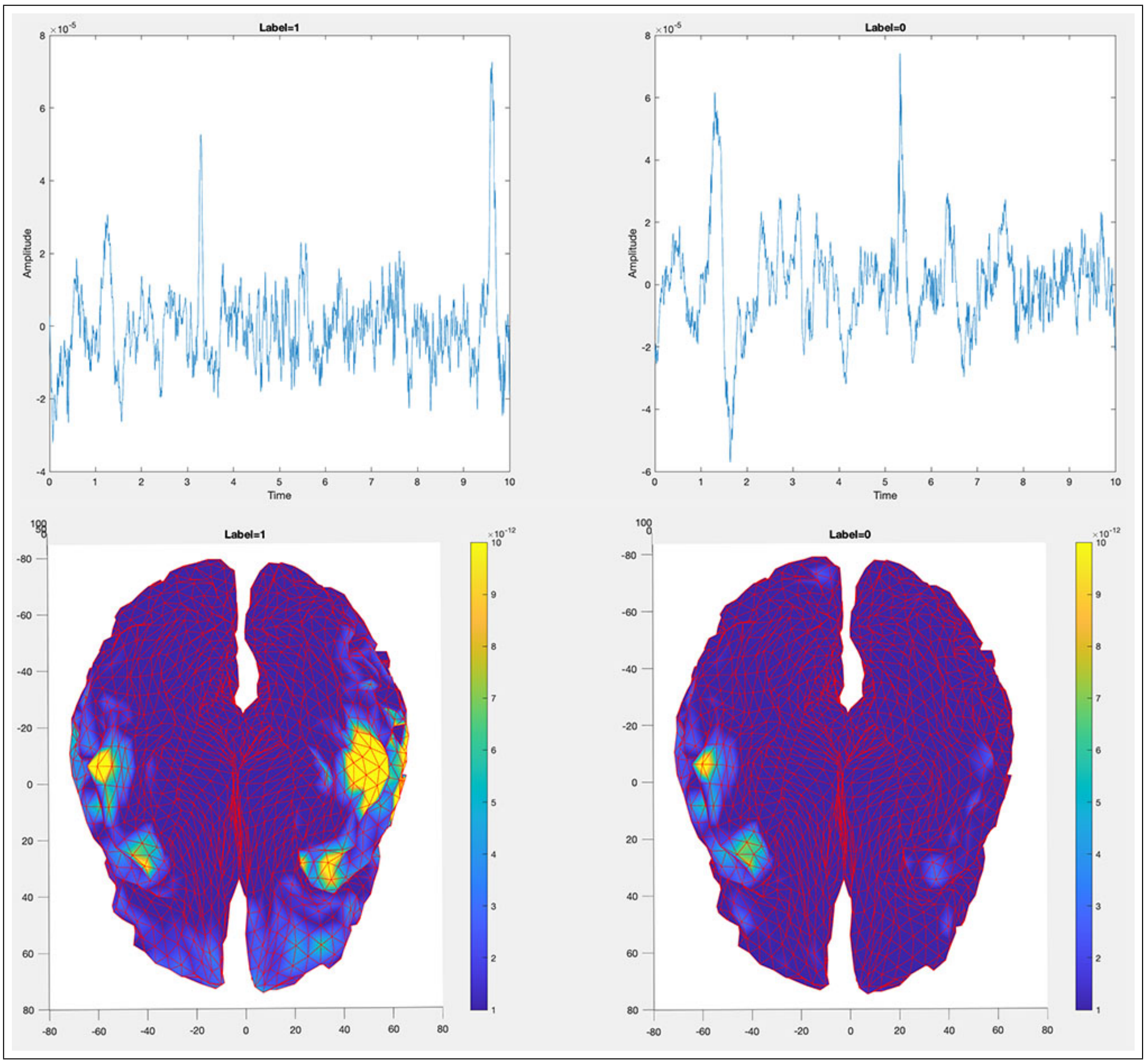

Figure 2.8 10 sec single channel EEG data segments for preictal (above left) and interictal stages (above right). Distribution of average power over 10 sec window for the multichannel EEG data of preictal (below left) and interictal (below right) stages projected to source space.

3. METHODOLOGY

3.1 Data

In this study, Physionet CHB-MIT Scalp EEG Database [38, 39] was used as data. This data includes EEG recordings of epilepsy patients recorded at Children's Hospital Boston. The data pool contains 23-channel EEG recordings of 22 subjects and 198 seizures in total. Electrode positions in these recordings are selected according to the International 10-20 system, and the sampling rate of the recordings is 256 Hz. The dataset contains many different record files for each patient. While some of these recordings are regular recordings, others are moments of epileptic seizures. Seizure start and end times are labelled on the data.

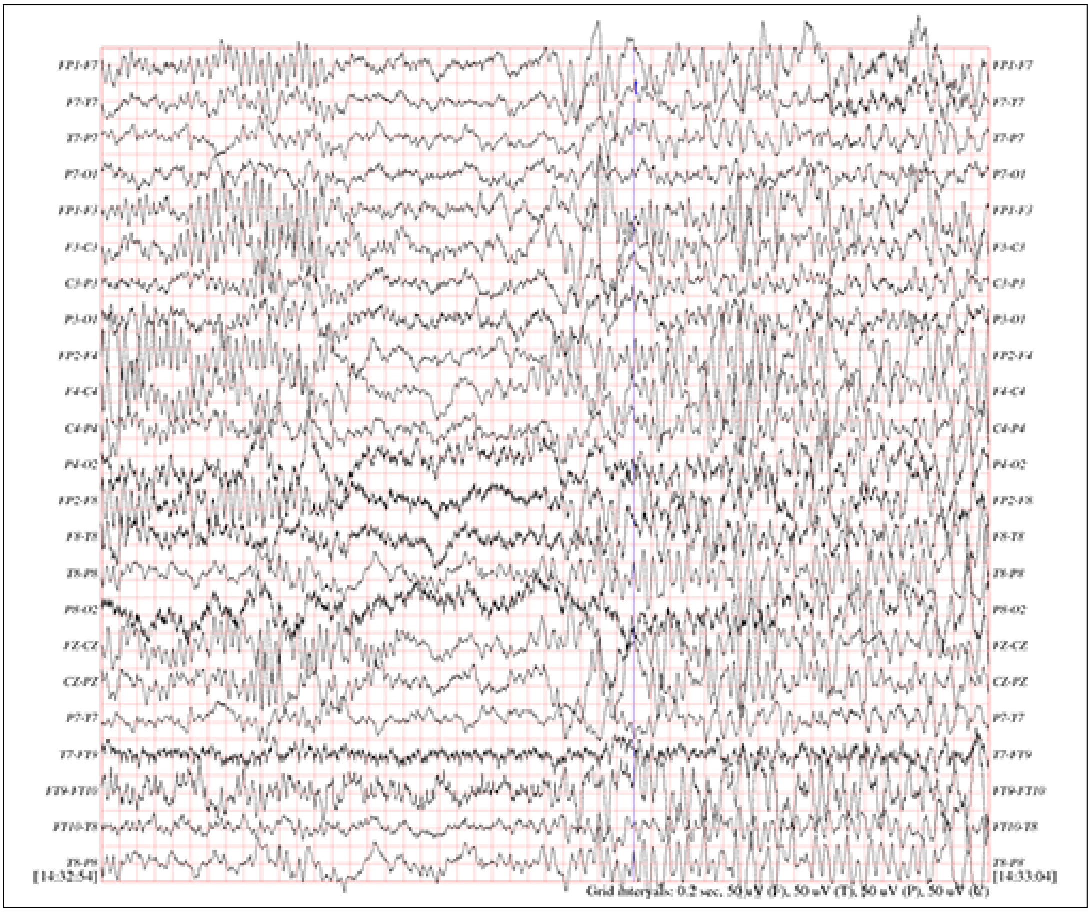

Figure 3.1 Sample signal from CHB-MIT scalp EEG database.

To predict the seizure in the EEG, the preictal state, which is the moment just before the seizure, must be determined. For this reason, classification will be between interictal states and preictal states, as in [18]. Preictal and interictal states should be selected manually, as the available data only includes seizure moments. For this, the interval before a specific time of each seizure should be selected as the preictal state. In this study, different preictal lengths were employed in the baseline model and how they were determined will be explained in the next section. Among the 15, 30 and 60-minute intervals tried, it was seen that the worse result was the 60-minute preictal selection. Since there was no significant difference in the results between 15 and 30, 30 minutes was preferred to predict the seizure earlier. These results can be seen in Table 3.1.

Table 3.1
Ideal preictal length comparision.

	Specificity	Sensitivity	FPR	AUC	Accuracy
15 min	78.64%	80.03%	0.21	79.33%	79.48%
30 min	**78.67%**	**78.32%**	**0.21**	**78.49%**	**79.04%**
60 min	77.47%	72.23%	0.22	74.85%	75.54%

The number of seizures in the data of each subject is listed in Table 3.2. Some conditions have been imposed to correctly distinguish between interictal and preictal times from data labelled seizure times. Firstly, the 60-minute time after the seizure was determined as postictal and not included in the interictal state. Secondly, despite the preictal length chosen as 30 minutes, the minimum preictal time length was limited to 15 minutes if there was not enough preictal recording due to the lack of pre-seizure recording or because it was too close to the previous seizure. The selected samples and subjects that did not meet these conditions were excluded from the data.

While processing the data, a balanced distribution of data was formed by selecting an equal number of interictal and preictal samples for each subject. Afterwards, many different samples were created by dividing the data into 10-second non-intersecting time windows which corresponds to 2560 time points in a 256 Hz frequency signal.

Table 3.2
Number of seizures per subject.

Subject	**Number of seizure**
1	7
2	3
4	4
5	5
8	5
9	4
16	10
18	6
23	7

3.2 Machine Learning Model (Baseline)

In the first method chosen for the classification problem, the SVM, a machine learning model, was preferred. In order to denoise the signals and obtain the signal information of interest, the EEG data were firstly filtered between 0.5 and 45 Hz with a Finite Impulse Response (FIR) band-pass filter. Afterwards, feature extraction was performed on 10-second signal segments with the CSP method using the Python MNE Library [40], and by using these segments, data were classified as binary (preictal or interictal) with the SVM classifier. The number of components in which the EEG signal was decomposed in the CSP method was determined as eight due to different trials. The scheme of the system implemented by SVM is shown in Figure 3.2.

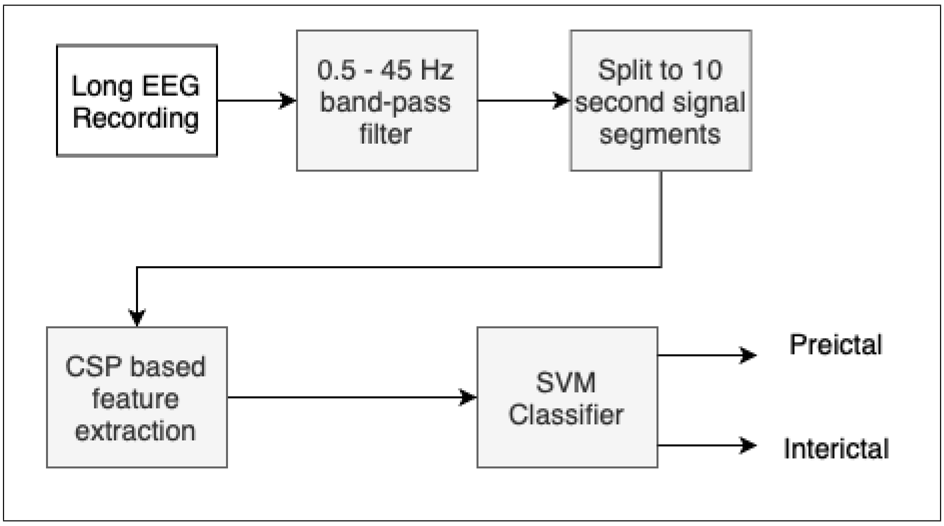

Figure 3.2 Flow of SVM classifier.

3.3 Deep Learning Models

In this study, different deep learning models were also applied, and the results were compared. The deep learning methods used are CA based classifier over raw signals, CNN based classifier over spectrograms and CNN based classifier over EEG source images.

3.3.1 Convolutional Autoencoder Classification

In the model proposed in this section, a convolutional autoencoder-based classifier has been developed. In the encoder part of the model, features are extracted through the convolution and pooling layers. In this way, the data size is gradually decreased, and is reduced from three dimensions to two dimensions in the last layer. On the other hand, the decoder enlarges the data with upsampling and deconvolution layers to restore it to its original state. The structure developed in this study is demonstrated in Figure 3.3. Since this model aims to generate the original signal again after projecting the data into the latent space, it learns to find the optimum reduced representation with the encoder without losing crucial information in the data. This structure, which reduces large-sized data to a smaller size, works as a feature extractor and preserves the most descriptive information. During the training of the autoencoder model, mean absolute error (MAE) is used as the loss function. MAE is determined as:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | y_i - \hat{y}_i| \tag{3.1}$$

where $|y_i - \hat{y}_i|$ represents the absolute error. Since the problem is signal reconstruction and is desired to obtain continuous values, this loss function is preferred.

In the next step, the encoder part of the trained auto-encoder model is separated and combined with the LSTM-based classifier. This classifier also includes a dropout layer with a reduction ratio of 0.2 at the input and 0.5 at the output of the LSTM layer. In this way, it is aimed to prevent the model from overfitting the training data and under-performing on the test data. During the training, "Binary Cross Entropy" is preferred as the loss function, and "Adam" algorithm [41] is used as the optimizer. This classifier, shown in Figure 3.4, learns to make predictions with the data in the low-dimensional space by using the feature extraction ability learned by the pre-trained encoder. Thus, the model can both learn faster and capture the essential features. Since the seizure pattern and brain structures of the patients differ, this classifier is trained for each patient separately.

3.3.2 Spectrogram Based CNN Classification

Another deep learning method used in this study is the CNN classifier working with spectrograms. This method includes a preprocessing step that converts EEG signal segments to spectrogram images. In this method, two dimensional spectrogram of each channel is estimated separately by using multi-taper spectrum for each signal segment, and combined channel-wise to make the data three dimensional. Creating a

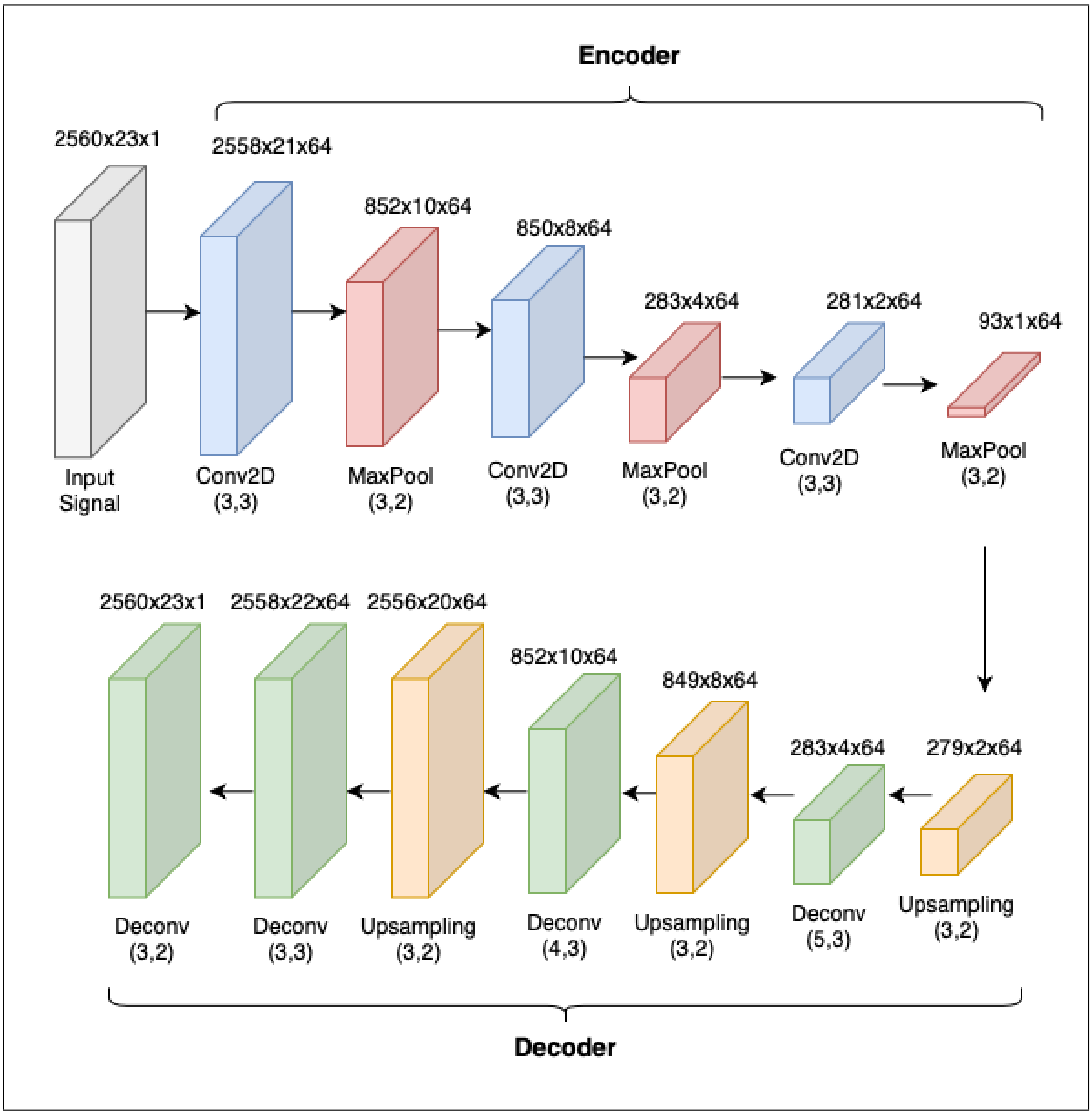

Figure 3.3 Auto-encoder model structure.

multi-taper spectral estimation consists of three steps. The first stage is the generation of discrete prolate spheroidal sequence (DPSS) tapers. In this regard, the parameters required for DPSS are number of samples (L) and half bandwidth (L1/2). These parameters can be calculated as,

$$L = \lfloor TW * fs \rfloor \tag{3.2}$$

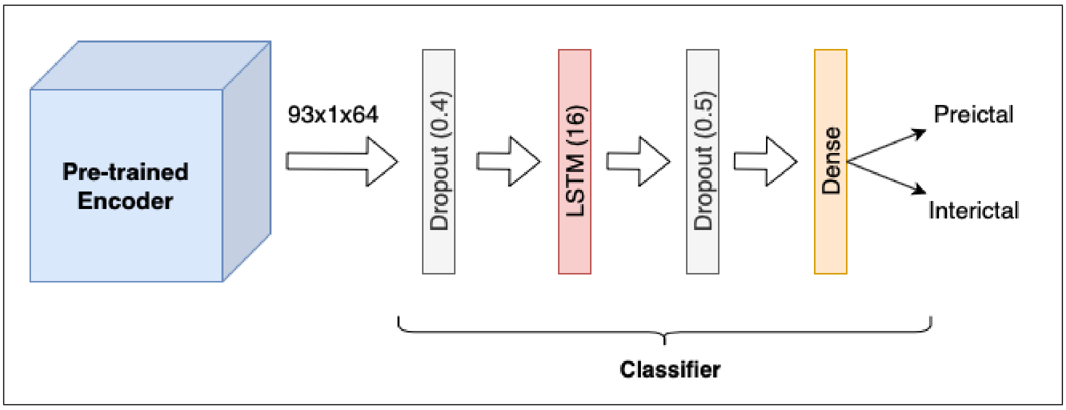

Figure 3.4 Classifier with pre-trained encoder.

$$\frac{L1}{2} = \lfloor 2 * TW * \Delta f \rfloor - 1 \tag{3.3}$$

where TW is the time window length, f_s is the sampling rate, and Δf is frequency resolution. After the DPSS tapers have been generated, the next step is to calculate the single-taper spectrum estimate for each DPSS taper. To do this calculation, we apply Fast Fourier Transform (FFT). Finally, a multi-taper spectrum can be estimation by taking the mean of the single-taper spectrums. The parameters used for estimating the spectrograms are as follows: time window length TW = 1.0 sec, spectral resolution $\Delta f = 45Hz$. Time windows with 50% overlaps are used to show spectral changes in EEG in the course of time.

During training, "Binary Cross Entropy" is chosen as the loss function, "Adam" is used as the optimizer in the auto-encoder model. This system is illustrated in Figure 3.5, and example spectrograms are demonstrated in Figure 3.7. As can be seen in the random samples selected for both preictal and interictal, the preictal sample is more active than the interictal sample, and there is a difference that can be used in decision making for this classifier as well.

The 3D spectrograms are input to the CNN and classified as preictal or interictal. This classifier consists of 3D convolution, max pooling, batch normalization, dense and dropout layers. Additionally, rectified linear unit (ReLU) is preferred as the activation

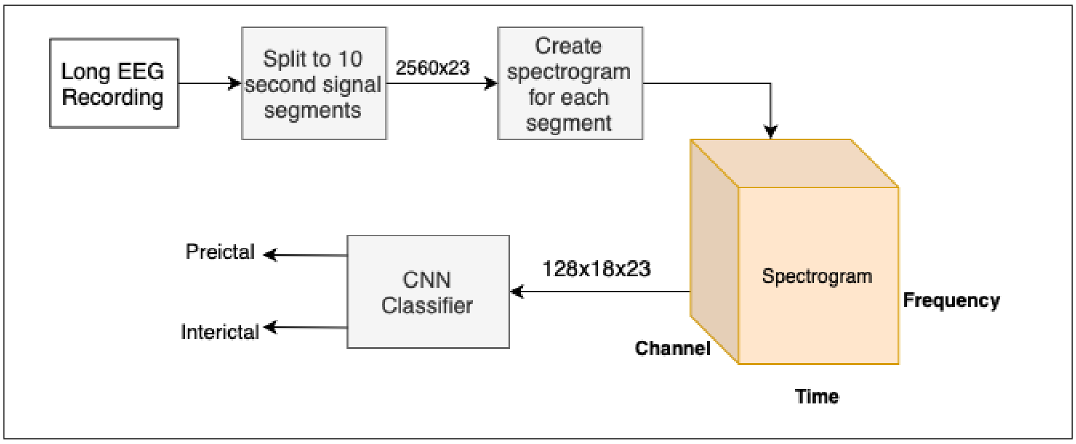

Figure 3.5 Flow of spectrogram based classifier.

function in the outputs of the convolution layers. In order to prevent over-fitting of the model and ensure that it can generalize, L2 regularization with 0.1 ratio and L1 regularization with 0.01 ratio is applied on the weights of all convolution layers. The 3D classifier developed for this classification problem is demonstrated in Figure 3.6.

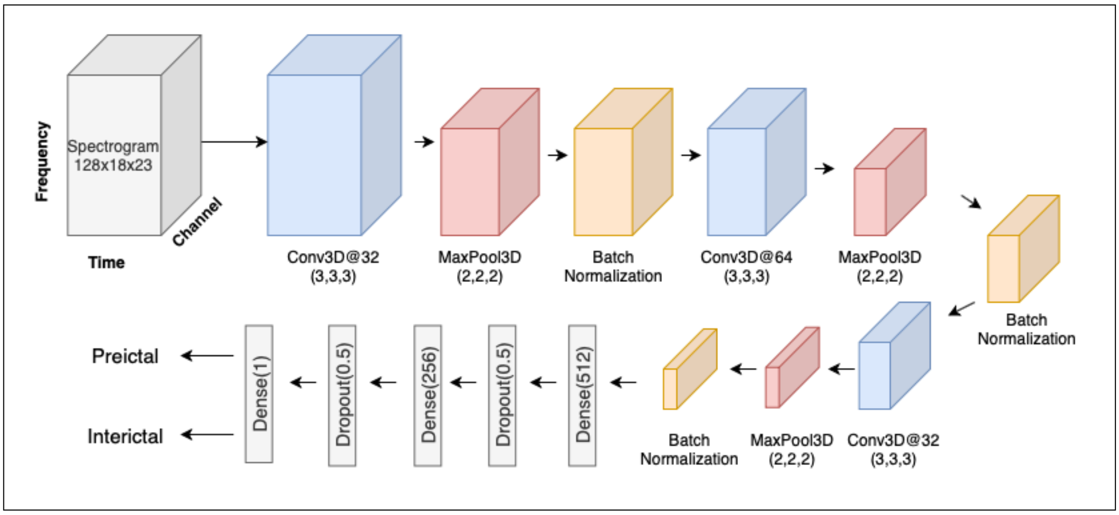

Figure 3.6 Spectrogram classifier.

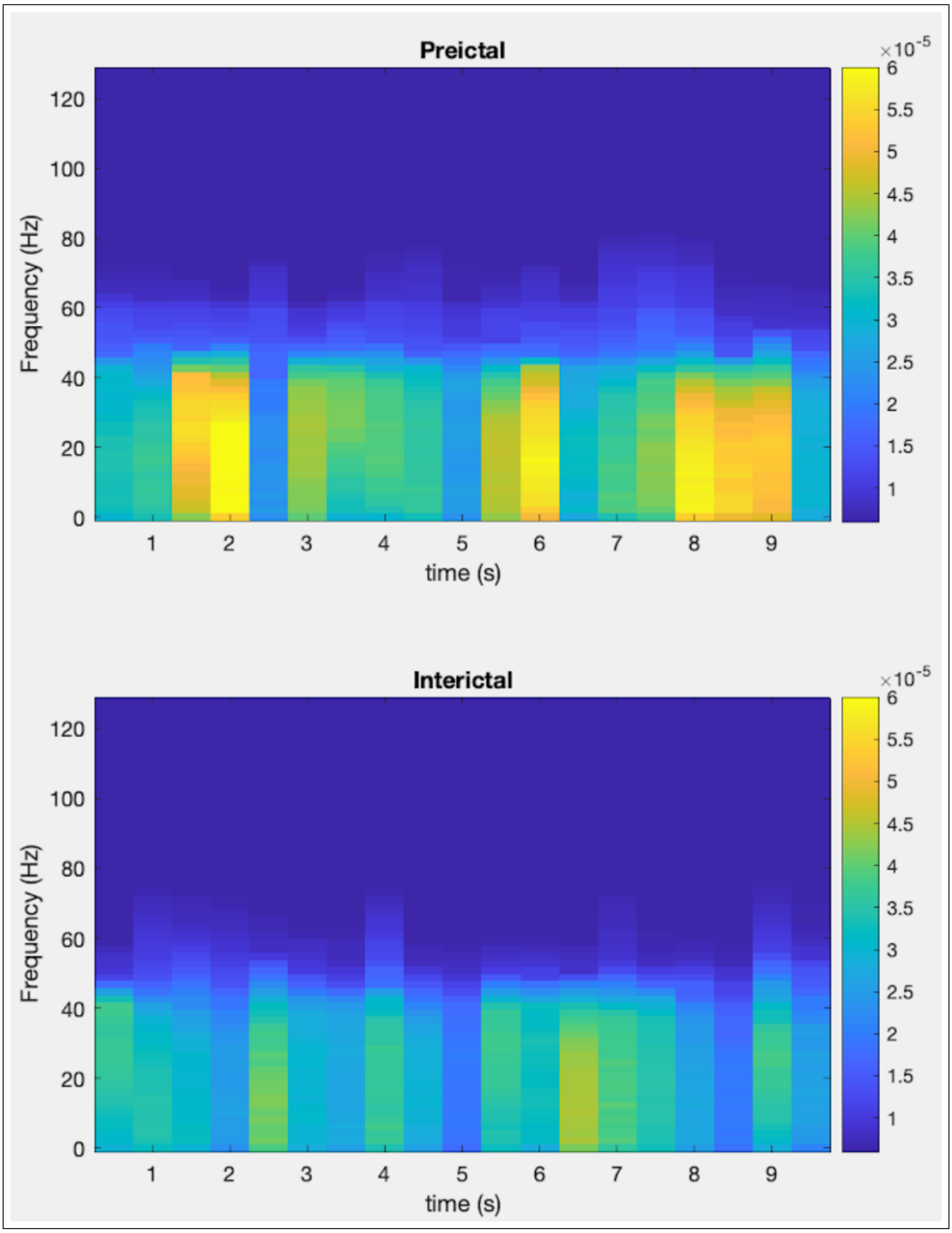

Figure 3.7 Example spectrograms for both preictal and interictal states.

3.3.3 Source Localization Based CNN Classification

Another deep learning method used in this study is the classifier, which works on 3D brain images estimated by source localization and includes 3-D convolution layers. In this method, each time window of EEG signals is converted into 3D source maps. The source localization aims to project the sensor space EEG data to source space. The forward problem is given as:

$$v = Lj + e \tag{3.4}$$

where v is sensor space EEG data, j is a vector of current dipoles at different sources, e is the noise and L is the lead field matrix that links source amplitudes j to electric potential v. L is obtained from the MNI atlas brain using tesellation based realistic head modeling and forward modeling algorithms available in Statistical Parameter Modeling (SPM). To achieve this, each vertex is mapped to a 3D location in the MNI atlas brain and to obtain a 3D EEG volume and smoothed with a spatial filter to achieve a continuous image.

The inverse problem is solved using the minimum norm algorithm. The minimum norm solution to Eq. 3.4 is given as $\hat{j} = Tv$ where $T = L^T(LL^T + \lambda C_e)^{-1}$, C_e is the covariance matrix of the error e which is assumed to be normally distributed as $\mathcal{N}(e|0, C_e)$ and λ is the weight coefficient. Our optimally chosen λ value is $\lambda = 1.25$.

Each time window consisting of 10 seconds are mapped on the source space and its average power spanning in the 13-45 Hz band is computed. The signal frequency is limited to provide meaningful information in source maps. The selected frequency range is chosen as the 13-45 Hz range, that is, the Beta and Gamma bands. 3D EEG power image is fed into CNN for classification. The block diagram of this system is represented in Figure 3.8, and the details of the classifier are shown in Figure 3.9. As in the other models, the loss function is "Binary Cross Entropy", and the optimizer is "Adam".

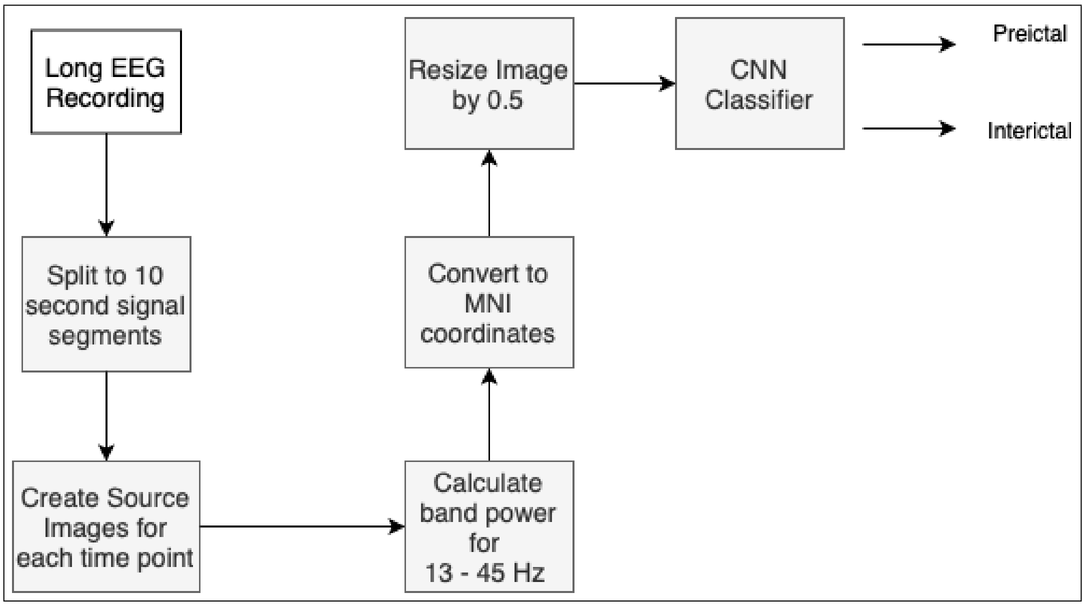

Figure 3.8 Source localization classifier schema.

3.4 Evaluation Method

The unbalanced distribution of the data set can make learning difficult in deep learning models. Since the dataset contains many more interictal periods than preictal periods, the number of interictal samples was taken to be same as the number of preictal samples. In the testing phase of the model, the leave-one-out cross-validation (LOOCV) method was preferred. In this method, the training is repeated for each patient for the number of seizures recorded at different times. In each training step, the data of all states for selected seizure were excluded, then training was conducted with the remaining data, and a test was conducted with the data of the selected seizure. In this way, it was ensured that each sample of data that the model did not use in learning could be tested for the generalization of the model. For performance measurement, the average of tests performed with all seizures of a patient was computed. The way this method works is demonstrated in Figure 3.10.

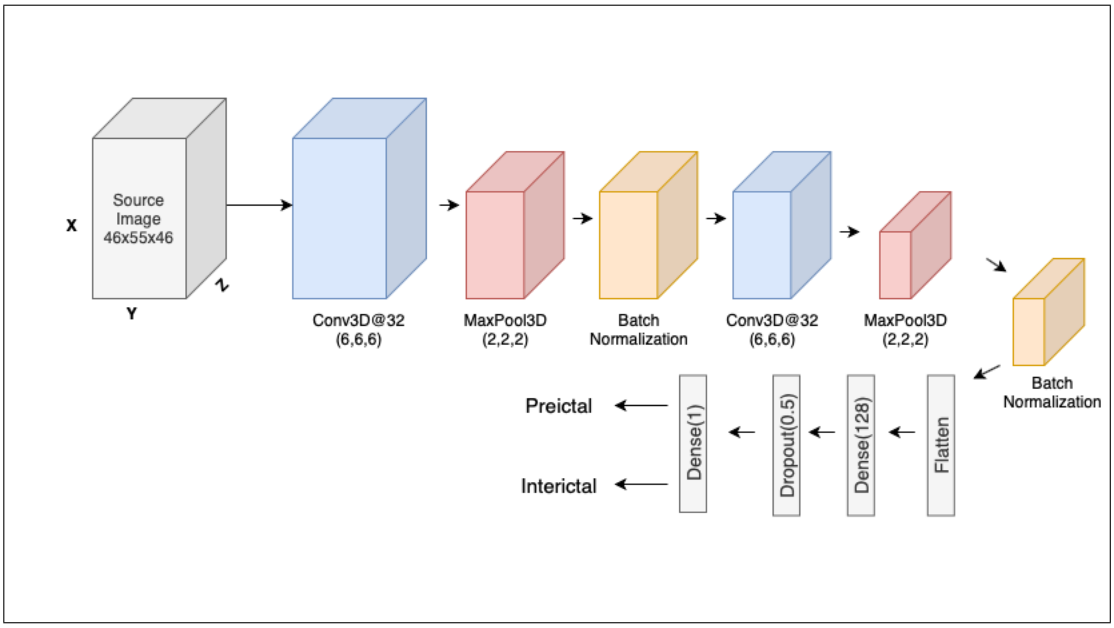

Figure 3.9 Source localization CNN classifier.

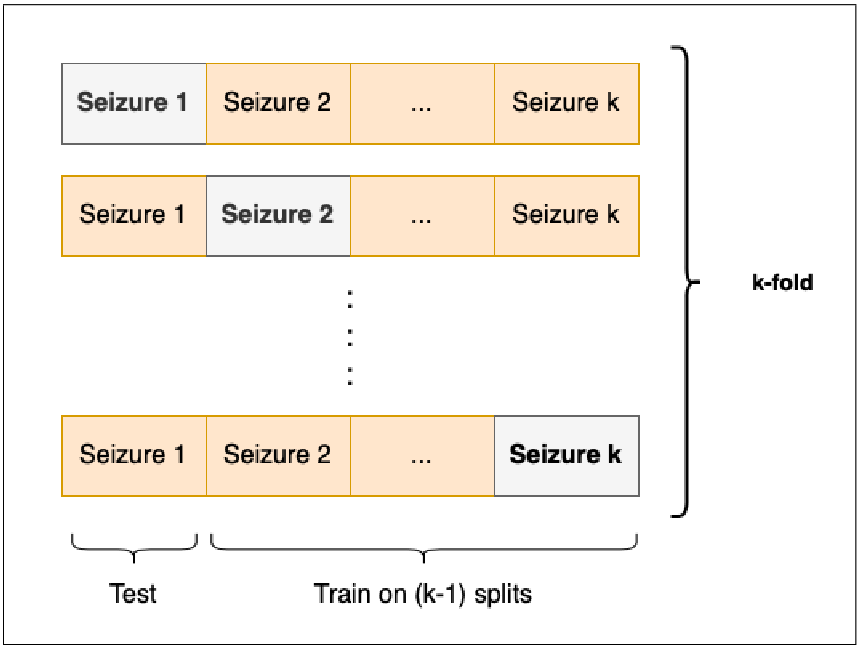

Figure 3.10 Leave-one-out cross-validation method.

4. RESULTS

In this study; the specificity, sensitivity, false positive rate (FPR), accuracy and area under the curve (AUC) metrics were used to measure system performance. AUC corresponds to the area under the ROC curve. A high AUC rate means low false positive and high true positive rate and represents a better model. Performance parameters are defined as:

$$Specificity = \frac{TN}{TN + FP} \tag{4.1}$$

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

$$FPR = \frac{FP}{FP + TN} \tag{4.3}$$

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \tag{4.4}$$

where, TP represents number of correctly predicted preictal windows, TN represents number of correctly predicted interictal windows, FP represents number of interictal windows that are predicted as preictal and FN represents number of preictal windows that are predicted as interictal.

4.1 SVM Based Classifier

An average of 70.38% specificity and 86.23% sensitivity scores were obtained among subjects in the study conducted with the SVM-based classifier, which was run on CSP-based features. In addition, the average FPR was 0.296. There were many false positives for Subject 16 and Subject 23 (FPR=0.601 and 0.577). The best performance of the model was obtained with 93.11% specificity and 97.1% sensitivity scores for Subject 1. All results of SVM classifier are listed in Table 4.1.

Table 4.1
SVM based classifier results.

	Specificity	**Sensitivity**	**FPR**	**AUC**	**Accuracy**
Subject 1	93.11%	97.15%	0.069	95.13%	94.49%
Subject 2	97.92%	78.61%	0.021	88.26%	91.48%
Subject 4	66.52%	81.52%	0.033	74.02%	68.20%
Subject 5	62.87%	75.98%	0.371	69.42%	67.24%
Subject 8	69.57%	90.11%	0.304	79.84%	76.32%
Subject 9	87.68%	94.86%	0.123	91.27%	88.58%
Subject 16	39.81%	87.49%	0.601	63.65%	55.70%
Subject 18	73.70%	89.44%	0.263	81.57%	78.95%
Subject 23	42.23%	80.88%	0.577	61.55%	45.69%
MEAN	**70.38%**	**86.23%**	**0.296**	**78.30%**	**74.07%**

4.2 Autoencoder Based Classifier

An average of 85.61% specificity and 83.07% sensitivity was achieved among subjects in the autoencoder-based classifier. Besides, FPR was 0.144 and accuracy was 84.59%. The lowest performance of the autoencoder model, which gave higher results than SVM, was in Subjects 16 and 23, similar to the SVM method (FPR= 0.283 and 0.337). Likewise, the best result was also obtained with Subject 1. Results of the auto-encoder based classifier are given in Table 4.2.

4.3 Spectrogram Classifier

An average of 89.96% specificity and 88.83% sensitivity scores were obtained with the classifier running with spectrograms, while the FPR was 0.101. The mean values of the other scores, AUC and accuracy, were 89.39% and 89.18%, respectively.

Table 4.2
Autoencoder based classifier results.

	Specificity	**Sensitivity**	**FPR**	**AUC**	**Accuracy**
Subject 1	96.05%	96.11%	0.039	96.08%	96.09%
Subject 2	98.19%	69.44%	0.018	83.82%	88.61%
Subject 4	87.59%	72.36%	0.124	79.98%	85.90%
Subject 5	90.93%	85.49%	0.091	88.21%	89.12%
Subject 8	79.51%	85.89%	0.205	82.70%	82.23%
Subject 9	92.26%	95.28%	0.077	93.77%	92.62%
Subject 16	71.67%	67.33%	0.283	69.50%	70.21%
Subject 18	87.96%	88.15%	0.120	88.06%	88.02%
Subject 23	66.28%	87.61%	0.337	76.95%	68.52%
MEAN	**85.61%**	**83.07%**	**0.144**	**84.34%**	**84.59%**

The model gave the lowest performance for Subject 8 (FPR=0.184), unlike other models. The best score was obtained with Subject 1 as in other models (Specificity = 97.39%, Sensitivity=96.80%). Results of this classifier are listed in Table 4.3.

4.4 Source Localization Based Classifier

An average of 89.06% specificity and 92.58% sensitivity scores were obtained with the classifier running with source localized images, while the FPR was 0.109. The mean values of the other scores, AUC and accuracy, were 90.82% and 90.41%, respectively. The model gave the lowest performance for Subject 16 (FPR=0.223 AUC=84.00%). The best score was obtained with Subject 1 as in all other models (Specificity = 94.17%, Sensitivity=97.45%). Results of this classifier are listed in Table 4.4.

Table 4.3
Spectrogram based classifier results.

	Specificity	**Sensitivity**	**FPR**	**AUC**	**Accuracy**
Subject 1	97.39%	96.80%	0.026	97.09%	97.09%
Subject 2	86.39%	93.98%	0.136	90.18%	89.85%
Subject 4	84.79%	83.27%	0.152	84.03%	84.40%
Subject 5	90.00%	86.41%	0.100	88.20%	88.21%
Subject 8	81.61%	76.93%	0.184	79.27%	77.96%
Subject 9	99.46%	98.87%	0.005	99.16%	99.38%
Subject 16	88.70%	79.17%	0.113	83.94%	83.95%
Subject 18	89.07%	89.10%	0.109	89.09%	89.14%
Subject 23	92.18%	94.93%	0.078	93.55%	92.66%
MEAN	**89.96%**	**88.83%**	**0.101**	**89.39%**	**89.18%**

Table 4.4
Source localization based classifier results.

	Specificity	**Sensitivity**	**FPR**	**AUC**	**Accuracy**
Subject 1	94.17%	97.45%	0.058	95.81%	95.81%
Subject 2	97.22%	92.08%	0.028	94.65%	94.68%
Subject 4	82.55%	91.03%	0.174	86.79%	84.08%
Subject 5	85.00%	86.97%	0.150	85.98%	85.97%
Subject 8	91.55%	96.48%	0.084	94.02%	93.78%
Subject 9	95.89%	93.45%	0.042	94.67%	95.46%
Subject 16	77.68%	90.32%	0.223	84.00%	83.99%
Subject 18	83.33%	89.56%	0.167	86.45%	85.51%
Subject 23	94.14%	95.91%	0.059	95.03%	94.42%
MEAN	**89.06%**	**92.58%**	**0.109**	**90.82%**	**90.41%**

5. DISCUSSION

5.1 Performance

It has been observed that deep learning methods are more successful than the machine learning-based SVM method within the scope of this study. While the average AUC was 78.30% in the classifier developed with SVM, this rate was above 80% in all deep learning-based classifiers. Among the deep learning models, the lowest score was obtained with auto-encoder with 84.34% and the highest score was reached using source localization with 90.82%. Among the models tried in this, study it has been seen that the complex structure of deep learning models can give high results even without pre-processing. Obtaining the best results with source localization shows that the performance of deep learning models can be increased when combined with the correct preprocessing methods. In the classifier developed with source localization, 89.06% specificity, 92.58% sensitivity, 0.109 FPR, 90.82% AUC and 90.41% accuracy scores were obtained. The spectrogram based model comes in second place. 89.96% specificity, 88.83% sensitivity were obtained. Although the auto-encoder-based model lags behind the other deep learning models, it differs from other methods in that it eliminates the need for preprocessing, and in this respect, it can be preferred with its rapid integration to different data.

Table 5.1
Average scores of each model.

	Specificity	**Sensitivity**	**FPR**	**AUC**	**Accuracy**
SVM	70.38%	86.23%	0.296	78.30%	74.07%
Autoencoder	85.61%	83.07%	0.144	84.34%	84.59%
Spectrogram	89.96%	88.83%	0.101	89.39%	89.18%
Source Localization	89.06%	92.58%	0.109	90.82%	90.41%

5.2 Run-time Considerations

For the epileptic seizure predictor to be successful, its adaptability to a real-time system is also crucial. The algorithm to be used should be able to work on real-time EEG recordings. Within the scope of this study, the estimation time of each tried method was calculated. The results were calculated and averaged over 1000 different samples. Time calculations were made on the Google Colab environment. The running times of all the developed algorithms are shown in Table 5.2. In the developed systems, it has been observed that the machine learning-based SVM algorithm works the fastest with a total of 0.76 ms, including the feature extraction process with CSP. On the other hand, deep learning algorithms are relatively slow compared to the machine learning-based system, although they yield for similar time scores among themselves while making predictions. Since the prediction times alone are around 30-40 ms, the difference among deep learning algorithms cannot affect the real-time applicability of the system. However, the pre-processing times required for the spectrogram and source localization algorithms considerably extend the running times of the system. It takes 372 ms to generate spectrograms from EEG data and 1.45 s to create source localization images. The Autoencoder-based method does not require pre-processing, so it only uses the prediction time and can estimate within 41 ms on the GPU and 61 ms on the CPU. In addition to the long pre-processing times of source localization and spectrogram algorithms, they give high performance, which might be acceptable according to the system to be developed. However, in cases where speed is much more critical, an auto-encoder based algorithm may be preferred.

Table 5.2
Runtime of each algorithm for single sample (Calculated with the average of 1000 trials).

	Preprocess Time	**Prediction Time**
Spectrogram	372 ms	32 ms on GPU - 92ms on CPU
Autoencoder	-	41 ms on GPU - 61 ms on CPU
CSP + SVM	0.76 ms total	
Source Localization	1.45 s	34ms on GPU - 45ms on CPU

6. CONCLUSION

This study discussed epileptic seizure prediction on EEG data, and different classification methods were tried and compared. Within the scope of the study, an SVM classifier working with CSP-based feature extraction, a deep learning-based CNN classifier working on the generated spectrogram images, a CNN classifier working on source localization images, and a classifier working with a deep learning-based convolutional autoencoder structure were developed and analyzed. Among these different methods developed, it was seen that the system working on source images worked best, while the machine learning-based classifier gave the lowest performance. Although deep learning-based methods are much more successful than SVM, they lag in terms of runtime.

In the epilepsy prediction system to be developed, the balance between the speed and performance requirement and the available system resources will determine the model type to be selected. Performing epileptic seizure prediction with a high accuracy enables developing an early warning system for healthcare professionals. In future studies, this method can be improved by experiments on different datasets. The developed approaches can be adapted to embedded software and hardware systems in the future and integrated with epilepsy monitoring systems in clinical settings.

7. List of publications produced from the thesis

1. B. Gozutok and A. Ademoglu, "Epileptic Seizure Prediction Using Convolutional Autoencoder Based Deep Learning," 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021, pp. 1-4

APPENDIX A. EPILEPSY DATABASES

A.1 CHB-MIT Scalp EEG Database

This dataset [38], consists of training data of 22 patients recorded at Children's Hospital Boston. The total recording time is 969 hours, and the data sampling rate is 256 Hz. The International 10-20 System is followed for the electrode placement.

A.2 Siena Scalp EEG Database

This dataset [42], contains EEG data of nine male and five female patients, recorded at the Unit of Neurology and Neurophysiology of the University of Siena. The data sampling rate is 512 Hz, and the International 10-20 System is followed for the electrode placement. The total recording time is 128 hours.

A.3 TUH EEG Seizure Corpus (TUSZ)

This data [43] consists of training data of 265 patients and 504 hours of EEG recordings. However, this data contains tens of different EEG channel configurations and recordings at different sampling rates. This inconsistency aspect, since the records come from multiple sources, differentiates this dataset from the others.

REFERENCES

1. Ma, D., J. Zheng, and L. Peng, "Performance evaluation of epileptic seizure prediction using time, frequency, and time–frequency domain measures," *Processes*, Vol. 9, p. 682, Apr. 2021.

2. Alhagry, S., A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *International Journal of Advanced Computer Science and Applications*, Vol. 8, no. 10, 2017.

3. Firpi, H., E. Goodman, and J. Echauz, "Epileptic seizure detection by means of genetically programmed artificial features," in *Proceedings of the 2005 conference on Genetic and evolutionary computation - GECCO '05*, ACM Press, 2005.

4. Binnie, C. D., and P. F. Prior, "Electroencephalography.," *Journal of Neurology, Neurosurgery & Psychiatry*, Vol. 57, pp. 1308–1319, Nov. 1994.

5. Bajaj, V., and R. B. Pachori, "Classification of seizure and nonseizure EEG signals using empirical mode decomposition," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 16, pp. 1135–1142, Nov. 2012.

6. Schulze-Bonhage, A., "Prediction of epileptic seizures," *Nervenheilkunde*, Vol. 27, no. 05, pp. 421–424, 2008.

7. Xygonakis, I., A. Athanasiou, N. Pandria, D. Kugiumtzis, and P. D. Bamidis, "Decoding motor imagery through common spatial pattern filters at the EEG source space," *Computational Intelligence and Neuroscience*, Vol. 2018, pp. 1–10, Aug. 2018.

8. Jiao, Z., X. Gao, Y. Wang, J. Li, and H. Xu, "Deep convolutional neural networks for mental load classification based on EEG data," *Pattern Recognition*, Vol. 76, pp. 582–595, Apr. 2018.

9. Wei, X., L. Zhou, Z. Chen, L. Zhang, and Y. Zhou, "Automatic seizure detection using three-dimensional CNN based on multi-channel EEG," *BMC Medical Informatics and Decision Making*, Vol. 18, Dec. 2018.

10. Vrbancic, G., and V. Podgorelec, "Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks," *Elektronika ir Elektrotechnika*, Vol. 24, Aug. 2018.

11. Wang, Z., L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time fourier transformation and deep neural networks for motor imagery brain computer interface recognition," *Concurrency and Computation: Practice and Experience*, Vol. 30, p. e4413, Jan. 2018.

12. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in Biology and Medicine*, Vol. 99, pp. 24–37, Aug. 2018.

13. Hussein, R., H. Palangi, R. K. Ward, and Z. J. Wang, "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals," *Clinical Neurophysiology*, Vol. 130, pp. 25–37, Jan. 2019.

14. Huang, D. S., K. Han, and A. Hussain, "Classification of Epileptic EEG Signals with Stacked Sparse Autoencoder Based on Deep Learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9773, p. V, 2016.

15. Daoud, H., and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Transactions on Biomedical Circuits and Systems*, Vol. 13, pp. 804–813, Oct. 2019.

16. Mirowski, P., D. Madhavan, Y. LeCun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clinical Neurophysiology*, Vol. 120, pp. 1927–1940, Nov. 2009.

17. Hosseini, M.-P., D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Optimized deep learning for EEG big data and seizure prediction BCI via internet of things," *IEEE Transactions on Big Data*, Vol. 3, pp. 392–404, Dec. 2017.

18. Rasekhi, J., M. K. Mollaei, M. Bandarabadi, C. Teixeira, and A. Dourado, "Epileptic seizure prediction based on ratio and differential linear univariate features," *Journal of Medical Signals & Sensors*, Vol. 5, no. 1, p. 1, 2015.

19. D'Alessandro, M., R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients," *IEEE Transactions on Biomedical Engineering*, Vol. 50, pp. 603–615, May 2003.

20. Alotaiby, T. N., S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, "EEG seizure detection and prediction algorithms: a survey," *EURASIP Journal on Advances in Signal Processing*, Vol. 2014, Dec. 2014.

21. Cortes, C., and V. Vapnik, "Support-vector networks," *Machine Learning*, Vol. 20, pp. 273–297, Sept. 1995.

22. Craik, A., Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, Vol. 16, p. 031001, Apr. 2019.

23. Hochreiter, S., and J. Schmidhuber, "Long short-term memory," *Neural Computation*, Vol. 9, pp. 1735–1780, Nov. 1997.

24. LeCun, Y., P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in *Shape, Contour and Grouping in Computer Vision*, pp. 319–345, Springer Berlin Heidelberg, 1999.

25. Yanagimoto, M., and C. Sugimoto, "Recognition of persisting emotional valence from EEG using convolutional neural networks," in *2016 IEEE 9th International Workshop on Computational Intelligence and Applications (IWCIA)*, IEEE, Nov. 2016.

26. Leite, N. M. N., E. T. Pereira, E. C. Gurjao, and L. R. Veloso, "Deep convolutional autoencoder for EEG noise filtering," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Dec. 2018.

27. Wen, T., and Z. Zhang, "Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals," *IEEE Access*, Vol. 6, pp. 25399–25410, 2018.

28. Shahbazi, M., and H. Aghajan, "A generalizable model for seizure prediction based on deep learning using CNN-LSTM architecture," *2018 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2018 - Proceedings*, pp. 469–473, 2019.

29. Muhammad Usman, S., S. Khalid, and M. H. Aslam, "Epileptic Seizures Prediction Using Deep Learning Techniques," *IEEE Access*, Vol. 8, pp. 39998–40007, 2020.

30. Timplalexis, C., K. Diamantaras, and I. Chouvarda, "Classification of sleep stages for healthy subjects and patients with minor sleep disorders," in *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, IEEE, Oct. 2019.

31. Koles, Z. J., M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topography*, Vol. 2, no. 4, pp. 275–284, 1990.

32. Ai, Q., Q. Liu, W. Meng, and S. Q. Xie, "Chapter 6 - EEG-Based Brain Intention Recognition," in *Advanced Rehabilitative Technology* (Ai, Q., Q. Liu, W. Meng, and S. Q. Xie, eds.), pp. 135–166, Academic Press, 2018.

33. Wang, J., Z. Feng, and N. Lu, "Feature extraction by common spatial pattern in frequency domain for motor imagery tasks classification," in *2017 29th Chinese Control And Decision Conference (CCDC)*, IEEE, May 2017.

34. Semmlow, J., "The fourier transform and power spectrum," in *Signals and Systems for Bioengineers*, pp. 131–165, Elsevier, 2012.

35. Prerau, M. J., R. E. Brown, M. T. Bianchi, J. M. Ellenbogen, and P. L. Purdon, "Sleep neurophysiological dynamics through the lens of multitaper spectral analysis," *Physiology*, Vol. 32, pp. 60–92, Jan. 2017.

36. Muñoz-Gutiérrez, P. A., E. Giraldo, M. Bueno-López, and M. Molinas, "Localization of active brain sources from EEG signals using empirical mode decomposition: A comparative study," *Frontiers in Integrative Neuroscience*, Vol. 12, Nov. 2018.

37. Michel, C. M., and D. Brunet, "EEG source imaging: A practical review of the analysis steps," *Frontiers in Neurology*, Vol. 10, Apr. 2019.

38. Shoeb, A., "Application of machine learning to epileptic seizure onset detection and treatment," *Diss. Massachusetts Institute of Technology*, pp. 157–162, 2009.

39. Goldberger, A. L., L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, Vol. 101, no. 23, 2000.

40. Gramfort, A., M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, Vol. 7, no. 267, pp. 1–13, 2013.

41. Kingma, D. P., and J. Ba, "Adam: A method for stochastic optimization.," *CoRR*, Vol. abs/1412.6980, 2014.

42. Detti, P., G. Vatti, and G. Z. M. de Lara, "EEG synchronization analysis for seizure prediction: A study on data of noninvasive recordings," *Processes*, Vol. 8, p. 846, July 2020.

43. Shah, V., 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*, Vol. 12, Nov. 2018.