Title

Machine Learning Models for Seizure Prediction and Detection from EEG

Author

Ali, Ahmed Siddique

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Machine Learning Models for Seizure Prediction and Detection from EEG

A thesis submitted

in partial satisfaction of the requirements

for the degree Master of Science in Bioengineering

by

Ahmed Siddique Ali

2024

ABSTRACT OF THE THESIS

Machine Learning Models for Seizure Prediction and Detection from EEG

by

Ahmed Siddique Ali

Master of Science in Bioengineering

University of California, Los Angeles, 2024

Professor Wentai Liu, Chair

Epilepsy is a disease characterized by having multiple unprovoked seizures; it can cause serious health complications for the individuals affected. The ability to predict seizures from EEG recordings can improve the standard of care for epilepsy patients by allowing care to be provided in a timely manner. In addition, detecting seizure occurrence in electroencephalogram (EEG) recordings can greatly speed up the time- and labor-intensive process of EEG annotation, which will also improve the level of care for patients. This study sought to both detect seizures and predict them by detecting the preictal stage before onset using machine learning (ML) models. To avoid the need for a manually-annotated preictal phase, a time period known as the "prediction target" was manually chosen and the signal was considered "preictal" if it fell within this period before seizure onset; prediction targets of 30 and 60 seconds were tested. A total of 186 features were extracted from the signal in the time-domain, the frequency-domain, and the time-frequency domain in order to characterize the most information about the signals' shape

and frequency content. The extracted features were used with four different ML models: Support Vector Machine (SVM), random forest, logistic regression, and Multilayer Perceptron (MLP). Of the 4 models, random forest performed the best, with an average accuracy of 65% on classification between ictal, preictal and background, and 85% on detection alone. The model showed strong performance on detecting seizures and an ability to detect the preictal phase. With further improvements, it could become highly effective for both seizure detection and prediction.

The thesis of Ahmed Siddique Ali is approved.

William F. Speier

Corey Wells Arnold

Aaron S. Meyer

Wentai Liu, Committee Chair

University of California, Los Angeles

2024

To my family and friends

Thank you for your unwavering support

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Dr. Wentai Liu for supporting my research project. I would also like to thank Han Xu for his guidance throughout the research process. Lastly, I would like to thank Dr. Bill Speier for his help in formulating this research problem.

1. Introduction

The brain is the primary organ in the Central Nervous System (CNS) and is responsible for controlling all voluntary body functions and thought processes. The brain is made up of cells called neurons. Neurons receive signals from other neurons or environmental signals, and transmit them through dendrites to the cell body or soma; these signals can be either excitatory or inhibitory. Once enough excitatory signals have been reached to cross a certain threshold, the neuron fires an "action potential", a depolarization wave, which travels down the axon and ends by sending a signal to other neurons or cells via synaptics.[1]

1.1 Epilepsy

Epilepsy is a brain disorder characterized by having chronic seizures. A seizure is an episode caused by electrical discharges in the brain, during which a patient will lose voluntary control of their movements. There are two main types of seizures: "partial" or "focal" seizures, in which only a certain part of the brain is affected, and "generalized" seizures in which the seizure spreads and affects the whole brain. While it is possible to have a seizure without having epilepsy, and about 10% of the population will experience one seizure within their lifetime, epilepsy is defined as having at least two unprovoked seizures. Epilepsy affects about 50 million people worldwide, disproportionately affecting people from low-income countries: while about 49 people are diagnosed with epilepsy per 100,000 in developed countries, that rate increases to 139 per 100,000 in low-income countries. Epilepsy can have a severe negative impact on a patient's quality of life, as they have a higher risk of premature death, as well as increased rates of physical and psychological injuries. Epilepsy is primarily treated through antiseizure medication, but this treatment is not universally applicable; in fact, about 30% of epilepsy

patients cannot have their condition controlled with drugs. Patients who cannot use drug treatments are usually treated with surgery instead.[2,3]

1.2. EEG For Epilepsy

The electroencephalogram (EEG) is the most common imaging tool for patients suffering from seizures or epilepsy. In an EEG scan, electrodes around the patient's head to detect changes in voltage, correlated with brain activity. There are two main types of EEG: scalp EEG, in which electrodes are non-invasively placed on the scalp, and intracranial EEG (iEEG), in which electrodes are invasively implanted inside the patient's skull.[4]

EEG recordings can be used to determine the exact timing of seizure occurrence. It is typical for such recordings to be annotated to indicate when the seizure starts and stops. Annotating EEG recordings can be done manually, but this is a time-consuming process. As such, a significant body of research is dedicated to using machine learning (ML) to detect seizures automatically in order to speed this process up.[5]

In addition, EEG can also be used for seizure prediction, using diagnostic tools to predict when a seizure will happen in the minutes before it occurs. Seizure prediction is possible because shortly before a seizure or "ictal" period, there is a "preictal" period, also known as an "aura", in the brain activity which can be detected. The exact length of the preictal stage is highly variable, and can last from a few minutes to a few hours. Predicting seizures is important because it allows patients to make preparations to avoid serious injury, and it allows for timely care and intervention to prevent adverse effects. The earliest studies in seizure prediction started in the 1970s, and the field has grown quite a bit in the decades since. However, creating a seizure

prediction program that stands up to scrutiny is challenging; in a 2007 review, it was determined that many previous studies in seizure prediction used flawed methodology and would not perform as well as they claimed.[2]

One major hurdle has been the lack of comprehensive datasets of epilepsy recordings. Over time, a few databases of epilepsy data have been put together for this purpose, such as the EPILEPSIAE database, the IEEG.org database, the Temple University Seizure Corpus (TUSZ), the CHB-MIT dataset, and more.[6,7,8,9] A 2023 review by Wong et al. compares the properties of these publicly available seizure databases.[10] The databases vary widely in their properties: while some, such as TUSZ and CHB-MIT, use scalp EEG recordings, others use iEEG, such as the IEEG.org.[6,8,9] The CHB-MIT dataset is one of the only ones with long-term continuous recordings lasting over 24 hours, while TUSZ and Siena Scalp have continuous data over shorter time periods, and the University of Bonn database has no continuous data at all.[6,9,11,12] Some databases already have filtering or artifact removal applied, and many have already applied selective segmentation to get the specific ictal, preictal or interictal recordings needed. The University of Bonn and Neurology Sleep Center Hauz Khas databases are the only ones in the review to have class-balanced data.[10,13,14] The length of EEG segments varies a lot as well; while some such as TUSZ and CHB-MIT have segment lengths of around 1 hour, others are much shorter, such as the Kaggle UPenn and Mayo Clinic database which is divided into segments of 1-second.[6,9,14] With such a wide variety, the choice of database can vastly affect an experiment, meaning that choosing the most ideal database is important.[4,10]

1.3. Current Methods of Seizure Prediction

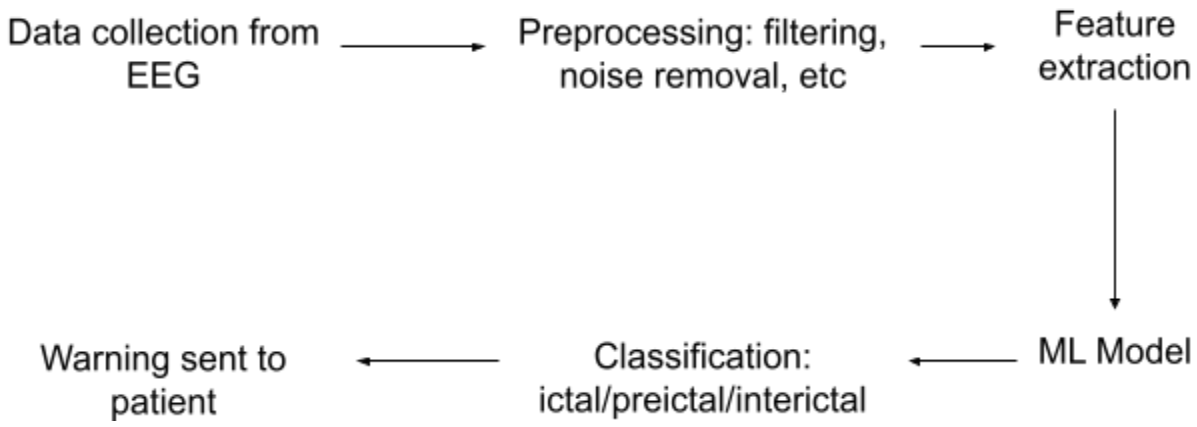

Figure 1: The typical step-by-step process for a seizure prediction algorithm.

A typical seizure prediction algorithm takes data on brain activity, usually from EEG; next, a preprocessing step usually includes filtering out noise, using methods such as the Butterworth and notch filters, as well as using methods to transform the signal such as wavelet transform and Empirical Mode Decomposition (EMD). Past studies have shown Fourier transform, wavelet transform, Butterworth filter, and Common Spatial Filtering to be the best methods for removing noise, as measured by the signal-to-noise ratio.[15] Some studies have also used channel selection during the initial preprocessing phase; this approach is most useful for predicting focal seizures, which are focused on a specific area in the brain rather than generalized.[15] Next, feature extraction is used to pull the relevant information from the signal; extracted features can include statistical features, spectral moments, Hjorth parameters, and more. The algorithm uses the features to determine if the patient is in the preictal stage and a seizure is about to occur, and the patient is alerted if a seizure is predicted. The features are then

fed into an algorithm that decides whether a seizure is about to occur. Finally, an alert is sent to the patient to warn them if a seizure is predicted.[2,15]

One of the key decisions that must be made in a seizure prediction algorithm is determining how far in advance the seizures can be predicted. While earlier seizure prediction is naturally preferable, past research has shown that if the prediction time is too high, the false alarm rate increases as well.[15] In addition, there is a higher chance a seizure will occur somewhere in the prediction window simply because there are more chances for that seizure to occur, so it is possible that random fluctuations in the signal will be incorrectly identified as a predictive marker, and the algorithm will still appear to perform well due to chance. Therefore, a balance has to be struck to predict seizures early while still ensuring that the predictions are providing meaningful information.[2]

There are multiple methods that can be used to analyze the EEG signal. In time-domain methods, the signal is kept in the original form measured across time. Methods in the time-domain include linear prediction and feature-reduction methods such as principal component analysis (PCA) and independent component analysis (ICA). The signal can also be transformed into the frequency-domain using a Fourier transform. It is then possible to calculate the power spectrum, which measures the power of each frequency component of the signal. A third option is to use a transformation into the time-frequency-domain, which keeps both temporal and frequency-based information in the signal; this can be accomplished with a Wavelet transform.[4]

The feature extraction phase is very important, as the algorithm's performance depends heavily on what features it is analyzing. Some of the most common features include statistical

features, such as mean, variance, skewness, median, etc; these features are very simple to calculate and can easily capture the most important information about the data, including its center and shape. Spectral information such as band power is also common, as they provide important information about the data in the frequency domain. The Hjorth parameters were specifically designed for signal processing, and so they are also frequently used in seizure prediction and detection. Other common methods include PCA, a dimensionality reduction method, and entropy, which represents the amount of uncertainty expected in the variable.[4]

ML can be divided into two main methodologies: supervised learning and unsupervised learning. In supervised learning, data is provided with "ground truth" labels that indicate the correct output, and the ML algorithm finds the parameters that most closely match the ground truth. In unsupervised learning, no labels are provided, and the algorithm simply forms clusters of data samples that are the most similar.

Numerous studies into epileptic seizure detection and prediction have been performed. The earliest studies on prediction were performed in the 1970s using linear measures, with the field expanding into non-linear methods in the 80s, as well as specifically trying to detect the preictal stage once it was discovered. Starting in 2002, competitions were held within the international science community to determine which methods performed best on a common dataset. Various ML methods have been used, both supervised and unsupervised. Some of the most common methods include Support Vector Machine (SVM), which has been used by Subasi et al and Fasil & Rajesh, among many others, and random forest, used by Wang et al and Tzimoutra et al.[16,17,18,19] Other methods include K-Nearest Neighbors (KNN), decision tree, logistic regression, and dynamic threshold.[2] In one study, a very simple set of features,

consisting of mean, coefficient of variation, dominant frequency, power spectrum mean, and variance, were used with a linear SVM, which achieved classification rates of over 90%.[20] For detection, some of the most successful studies have used SVM, random forest, and KNN, all achieving accuracies above 90%.[5]

Recently, the use of Deep Learning (DL) methods, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), has begun to gain traction. For detection, Artificial Neural Networks (ANN), Extreme Learning Machine, and Back-Propagation Neural Networks have all also achieved over 90% accuracy, comparable to the best non-DL methods.[5] In 2019, Truong et al developed an algorithm using CNN with STFT for preprocessing which performed very well, reaching 79.8% sensitivity on raw data.[21] Several other studies have used CNN, with some only using it for feature extraction and then feeding the output into another algorithm such as SVM.[15] Other studies have tried using unsupervised learning in order to get around the lack of labeled data. One study used k-means to achieve 91.43% accuracy, and in 2021, Yildiz et al proposed a fully-unsupervised deep learning method for seizure prediction, using variational autoencoder (VAE), with an AUC of 0.83.[22,23]

The goal of this study was to create a ML model that can predict seizures when trained on the TUSZ, the largest freely available dataset of scalp EEG recordings from epilepsy patients. This analysis was done through a feature extraction process that included statistical features, Hjorth parameters, band power, EMD, continuous wavelet transform (CWT), and PCA. The extracted features were analyzed using four ML models: SVM, random forest, logistic regression, and multilayer perceptron (MLP).

Table 1: list of abbreviations used throughout the paper and their meanings.

Abbreviation	**Meaning**
ANN	Artificial Neural Network
AUC	Area Under the Curve
CNN	Convolutional Neural Network
CNS	Central Nervous System
CWT	Continuous Wavelet Transform
DL	Deep Learning
DWT	Discrete Wavelet Transform
EDF	European Data Format
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
ICA	Independent Component Analysis
iEEG	Intracranial EEG
IMF	Intrinsic Mode Function
KNN	K-Nearest Neighbors
ML	Machine Learning
MLP	Multilayer Perceptron
mV	Millivolts
PCA	Principal Component Analysis
RBF	Radial Basis Function
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
s	Seconds
SOP	Seizure Occurrence Period

SVM	Support Vector Machine
TUSZ	Temple University Seizure Corpus
VAE	Variational Autoencoder

2. Methodology

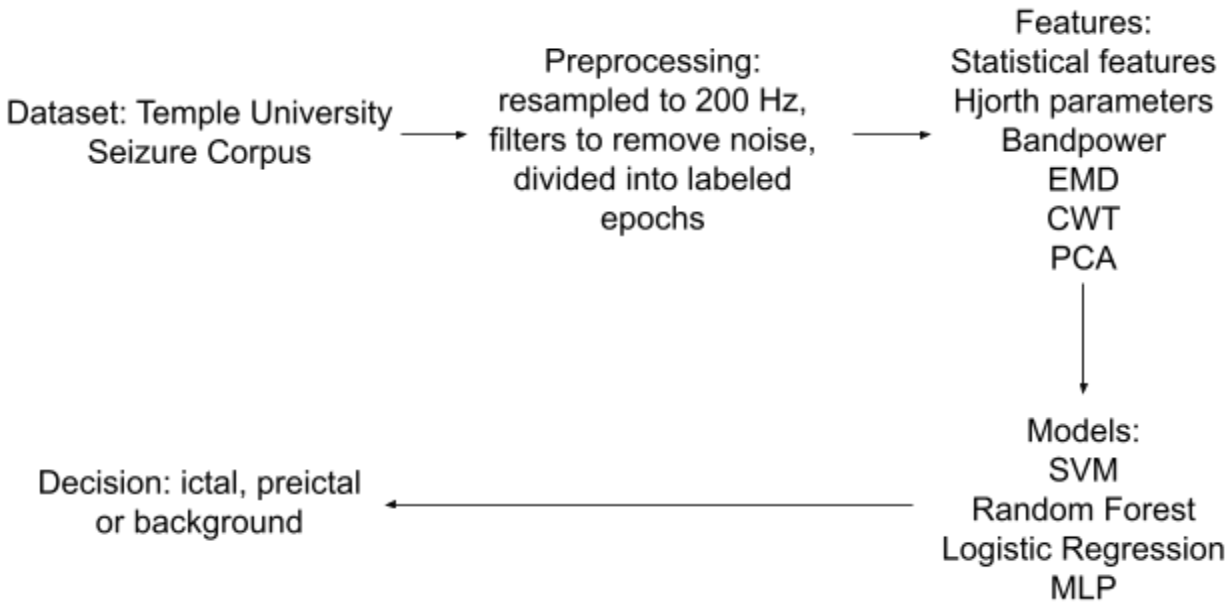

Figure 2: Pipeline of steps for this model, following the same structure as the pipeline in Figure 1.

For this study, EEG data of epilepsy patients was taken from the TUSZ. Data was preprocessed by resampling all recordings to 200 Hz, and then applying filters to remove 60 Hz noise and limit the signal to 0.5-100 Hz, the range relevant for seizure activity. The signal was then divided into epochs, which were labeled as ictal, preictal, or background based on whether they fell during a seizure, the prediction target before onset, or neither. Next, 186 features were extracted, including statistical features, Hjorth parameters, average bandpower, and PCA; in addition, the signal was processed with EMD and CWT and then had features extracted from the resulting waveforms. 4 models were tested on the extracted features: SVM, random forest, logistic regression and MLP. After the models were trained, they were used to decide if a signal epoch was ictal, preictal, or background.

2.1. Dataset

The TUSZ is the largest open-source database of seizure recordings. It is a subset of the Temple University Hospital EEG Corpus, with recordings of patients that experienced seizures, as detected by a Natural Language Processor that scanned the annotations. The most recent release has 822 EEG recordings, with 280 containing seizures, from a total of 315 subjects. EEG recordings are provided in the file format European Data Format (EDF). Many of the subjects were recorded in multiple sessions, and the recordings were additionally divided into multiple EDF files labeled as separate tokens. Each recording also has an annotation file, in Comma-Separated Value format, detailing the start and stop times for events in each channel. Eleven classifications of seizures are included, each with their own labels. Because recordings were taken at many different times under different circumstances, they are not uniform, and the sampling rates vary with some having 250 Hz, 256 Hz, and 400 Hz. The recording sessions are cut into different "tags" with varying lengths. Recordings in this dataset use the tcp_ar and tcp_le montages, each with 22 channels, and the tcp_ar_a and tcp_le_a montages, which only have 20 channels. While the annotations are provided for the bipolar channels, the EDF files provide the raw data for each channel, so each bipolar channel has to be manually added. EEG recordings were read into Python and processed using the MNE library. Each patient has an eight-digit identification code to protect their privacy.[6]

The TUSZ database has a few notable characteristics that make it ideal for this study. First, it has the most patients of any publicly available database, by far, with 315 subjects; in comparison, most of the other publicly available databases have fewer than 100 recordings. The data is also continuous in the short-term and full recordings include the time before and after

seizures, allowing this study to better simulate the data that would be obtained in a real-world scenario. The TUSZ also uses scalp EEG data rather than iEEG; this source is useful because when this is applied in the real world, it is ideal to require the least invasive procedure possible. One weakness of this database is that only the seizures are manually annotated, not the preictal stages; however, since this study uses an automated classification of "preictal" data within a certain range before a seizure, that is not a problem. Another weakness is the general inhomogeneity of the data, given that recordings take place over multiple years with different montages, numbers of channels, and recording lengths. However, this database is still usable because there is enough data per patient to train and test the model on each patient individually, there is less variance among each patient's data.[6,10]

2.2. Preprocessing

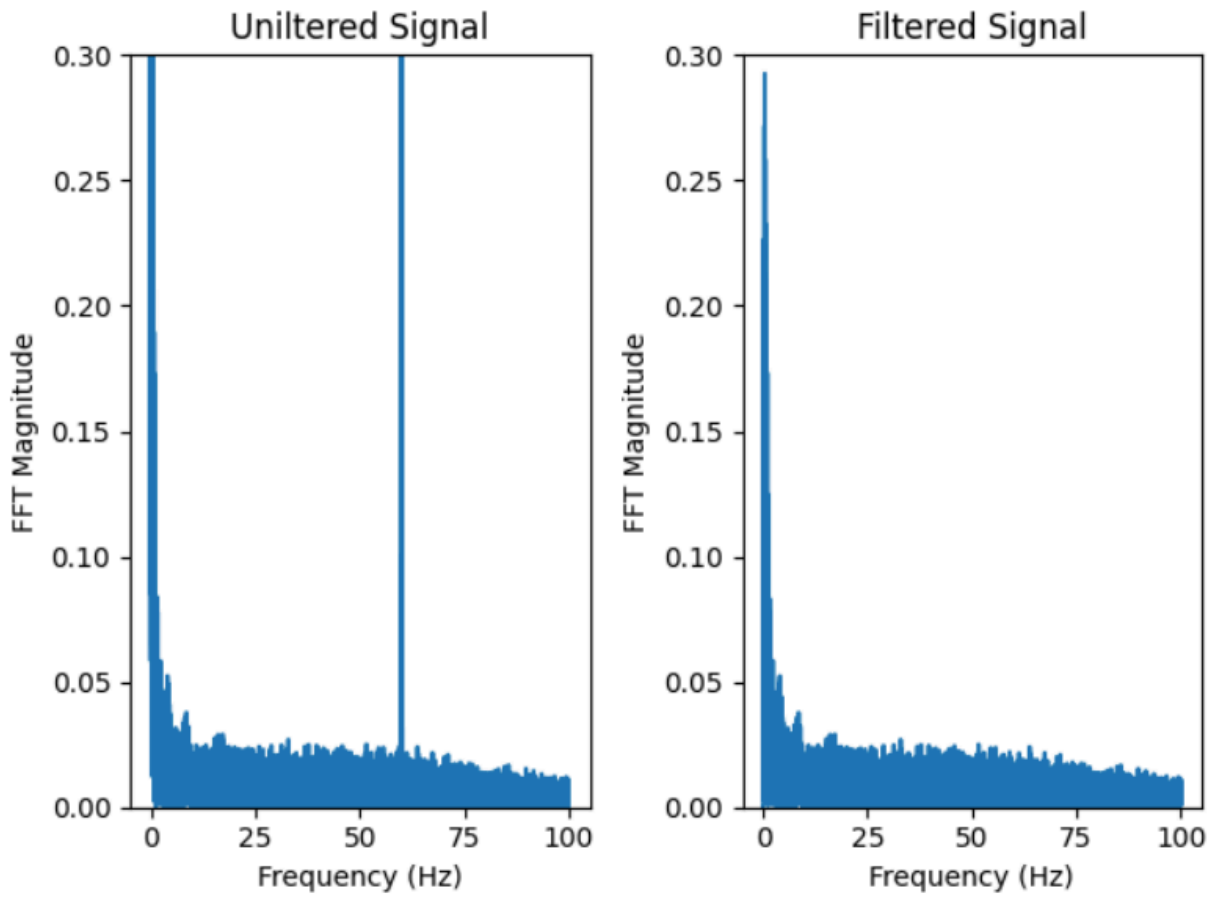

Figure 3: Fourier transforms of a patient's EEG signal before filtering (left) and after filtering (right).

Data from the TUSZ database was resampled at 200 Hz so that all recordings would have a consistent sampling rate. A bandpass filter was applied to eliminate all signals outside the range of 0.5-100 Hz, since frequencies created by epileptic brain activity generally lie within this range and any frequencies outside this range were likely due to noise. Additionally, because the TUSZ contains line noise at 60 Hz, a notch filter was applied to remove the signal at 60 Hz.[24] As the above figure shows, the unfiltered signal includes a massive spike at 60 Hz on the Fourier

transform, which is removed by the filtering, and the signal itself is smaller and amplitude and appears to be less noisy when the filtering is applied.

2.3. Division into Epochs

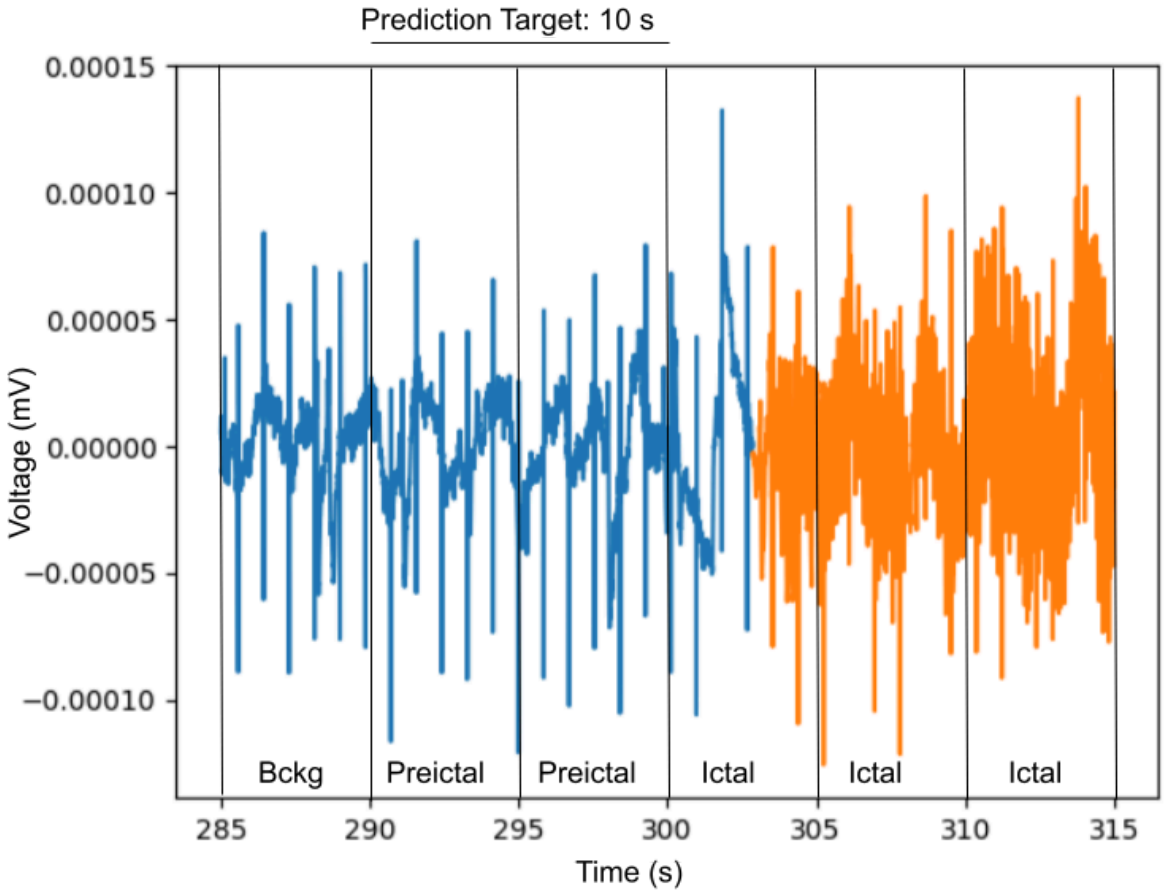

Figure 4: An example showing epoch division on an EEG signal. The signal is colored orange during the seizure and blue for background. For this example, there is no overlap, and the prediction target is 10 seconds.

For this experiment, single-channel analysis was performed. First, an "events" array was created for each channel, with a binary label at each time point of "0" for background and "1" for

seizures of any kind. Next, both the EEG file and the events array were cut into five-second epochs. The epoch division can be applied with an "overlap" to increase the number of epochs using a sliding window. Epochs are classified as "ictal" if at least 50% of the epoch takes place during a seizure, as determined by the events array. All epochs within a certain time range, known as a "prediction target", before a seizure starts are labeled as preictal. A third "background" label is applied for all epochs that are neither ictal or preictal, including the interictal signal. The epoch length, prediction target and overlap are all modifiable and multiple values were tested over the course of this experiment. For the majority of the tests described in this thesis, the settings were an epoch length of 5 seconds, overlap of 3 seconds, and a prediction target of 30 seconds. In order to perform seizure detection without including prediction, the prediction target was set to 0 seconds so that the preictal label would be removed. The above figure demonstrates the process. The figure below shows examples of 5-second epochs for all 3 class labels.

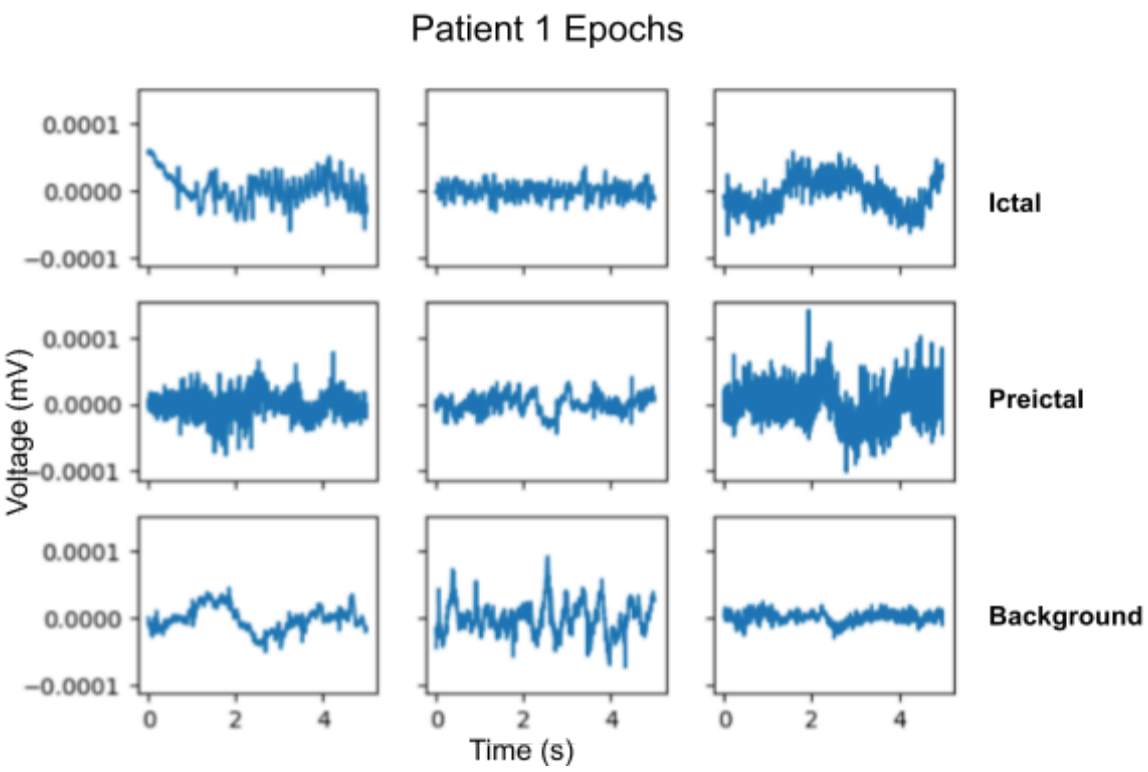

Figure 5: 5-second epochs for each of the three classes.

2.4. Data Balancing

In the original dataset, the number of epochs in each class are very unbalanced. For most files, the ratio of ictal to preictal epochs ranges from about 2:1 to 3:1, while the ratio of background to preictal is much higher, reaching values around 100:1. While it is possible to simply use the entire dataset unbalanced, and simply have the analysis models weighted to match the class imbalance, this approach is impractical for 2 reasons: first, even with weighting, the model is susceptible to overfitting to the training data and appearing to be highly accurate by simply choosing the most common class (in this case "background") for the majority of epochs. Additionally, the more data is analyzed, the longer it takes to extract features and train the model, and it is simply not practical to do this for a large number of patients. In order to balance the

amount of data in each class, for each recording in a given dataset, the number of epochs for each of the 3 classes is counted, and the lowest of these 3 values, usually for preictal epochs, is taken and set as the sample size. Next, random samples of the same sample size are taken from both other classes so that all 3 classes have equal value.

2.5. Patient Selection

A filtering process was applied to select the patients with the most suitable data for analysis. First, a spreadsheet was generated with information for each EDF file, including the patient identifier, the filename, the number of epochs of each label and total, the percentage of epochs of each label, the number of epochs after balancing. Next, certain filters were applied: each recording had to have at least 1% of its epochs come from each class, and they had to have at least 5000 total epochs before balancing. Finally, the total number of files for each patient after filtering were counted up, and the 10 patients with the highest number of recording files were chosen for analysis. For these 10 patients, the number of files ranged from 6 to 33. For simplicity, the 10 patients are renamed Patient 1-Patient 10, with patient 1 having the most files and Patient 10 having the least.

Table 2: patient IDs in the dataset and their labels in this thesis.

Patient	**8-Digit ID**
Patient 1	aaaaardk
Patient 2	aaaaapks
Patient 3	aaaaajqo
Patient 4	aaaaakfo

Patient 5	aaaaates
Patient 6	aaaaanme
Patient 7	aaaaajns
Patient 8	aaaaakkm
Patient 9	aaaaalfs
Patient 10	aaaaammu

2.6. Feature Extraction

A number of different features were extracted from each epoch for use in the SVM: statistical features, Hjorth parameters, band power, EMD features, CWT features, and PCA.

2.6.1. Statistical Features

The features included the first 4 statistical moments of the signal: mean, variance, skewness, and kurtosis, according to equations 1-4. The mean represents the center of the distribution. The variance is a measure of the spread. The skewness measures the asymmetry of the distribution. The kurtosis measures the distribution's "tailedness", or the frequency of outliers.

2.6.2. Hjorth Parameters

The next features extracted are the Hjorth parameters, statistical properties used in signal processing. The first Hjorth parameter is the mobility, which represents the proportion of standard deviation of the power spectrum and is mathematically defined as the square root of the

variance of the time derivative of the signal divided by the variance of the signal. The other Hjorth parameter is the complexity, which computes the similarity of the signal with a sine wave, and is calculated by dividing the mobility of the signal's derivative by the mobility of the signal.[25]

2.6.3. Average band power

The next feature included was the average band power of each "band", or range of signal frequencies, measuring how much of the signal is contained within that band. In order to calculate the band power, first, the Power Spectral Density (PSD), or the proportion of power over frequency, is calculated using the Welch method; next, the PSD is integrated over the frequency range of each band to get the power in that band. The average band power was calculated for 5 bands: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-120 Hz).

2.6.4. Empirical Mode Decomposition

EMD is a process that breaks a signal down into several components known as intrinsic mode functions (IMFs). In each step of EMD, the signal is treated as a fast oscillation superimposed on a slower oscillation. At each step, the fast oscillation is extracted as an IMF, and the remainder of the signal is once again split into a fast and slow oscillation. In this model, the first 5 IMFs were extracted, and the statistical features, Hjorth parameters, and average band powers were all calculated for each of the 5 IMFs.[26]

2.6.5. Continuous Wavelet Transform

In a CWT, the signal is convoluted with a function known as a wavelet function in order to measure their similarity. Unlike a Fourier transform, a wavelet transform is still a function of time, so the transformed signal still includes information about the time-based components of the original signal. The "scale" of the wavelet function stretches or shrinks the signal, and each scale provides information about different frequencies in the original signal. In this model, 10 scales of the "morlet" function are applied, corresponding to frequencies between 1 and 100 Hz, and the statistical features, Hjorth parameters and band powers are extracted from each scale.[27]

2.6.6. Principal Component Analysis

PCA is a dimensionality reduction technique frequently used to simplify the information in datasets with high numbers of dimensions. In PCA, the direction of the highest variance in the original dataset is identified, and a new "axis" is set in this direction, so that the first Principal Component can be calculated as the values of the data on this new axis. A second orthogonal axis is chosen in the direction that will explain the second-highest amount of variance, and a second principal component is calculated in this axis; this process can be repeated for any arbitrary number of principal components up until the number of dimensions in the original dataset. In this model, the values of 10 principal components are extracted and added to the set of features.[28]

In total, 186 features were extracted for each epoch.

2.7. Cross-validation

In cross-validation, a dataset is split into training and testing values, so the model can be fit on the training set and evaluated on the test set. Cross-validation was performed using a Group K-Fold split. In each fold, data was taken from all the EDF files for one patient, and the data was grouped based on which specific EDF file it came from. A 5-fold split was performed with each split being roughly equal in size, meaning that the testing group always contained about 20% of the data.

In order to tune the hyperparameters of the model, nested cross-validation was performed. Nested cross-validation is a process in which after the initial split into training and testing, a "validation" set is withheld from the training set. All possible combinations of the hyperparameters, or manually tuned parameters, of the models are tested on the validation split for each fold in a grid-search process, and the best-performing set of hyperparameters is found. These hyperparameters are used with the testing set and the performance of the model is evaluated. In this model, the "outer" split into training and testing was done with a 5-fold split, while the "inner" split was 4-fold; both were done using a Group K-Fold split.

2.8. Models

The prediction model was done using a "pipeline" with two steps. First, the "Standard Scaler" function in sklearn was used to normalize the features into z-scores. Next, a machine learning model was fit to the training data, and the accuracy metrics were calculated on the testing data. Four models were tested: SVM, Random Forest, Logistic Regression, and MLP.

2.8.1. Support Vector Machine

In an SVM, the model finds hyperplanes in n-dimensional space that separate the features by labels. Since the features are frequently not linearly separable, it is often necessary to transform the features into a higher-dimensional space using some function. In order to reduce the computational complexity of this operation, the "kernel trick" is used, in which dot products are replaced with less computationally expensive "kernel functions".[29]

In this experiment, an SVM was created with a radial basis function (RBF) kernel. The two hyperparameters tuned are C, which is inversely proportional to regularization, with possible values of 10, 100, and 1000, and gamma, the kernel coefficient, with possible values of 0.01, 0.001, and 0.0001.

2.8.2. Random Forest

In a random forest classifier, numerous "decision trees" are generated, which make several decisions at each feature to lead to the final classification. A process known as "bagging" picks a random subset of data for each tree to analyze, and the final decision is made based on the highest consensus of decision trees.[30]

Two hyperparameters were tuned for random forest: the number of estimators, which could be 10, 100, or 1000, and the formula for max features at a single split in a tree, which could be either the square root or log-2 of the total number of features.

2.8.3. Logistic Regression

Logistic regression is a type of regression model that uses a logistic function, which behaves similarly to an exponential function at low values and to a logarithmic function at high values. Logistic regression can be used as a prediction model by using the probability of an outcome, with values between 0 and 1, as the dependent variable. In the data itself, all values will be set to 0 for a negative outcome or 1 for a positive outcome, and the logistic curve is fit to the resulting data; using this curve, probability predictions can be made for any given set of features, and they can be converted to a prediction. For a multiclass case such as this one, a process known as multinomial logistic regression is used.[31]

In this study, the logistic solver is applied with 1000 max iterations. The penalty term is "elastic net", which adds both the L1 and L2 penalties, and the SAGA solver is used. Two hyperparameters were tuned: C, a variable inversely proportional to the regularization strength, which could be 1, 0.1, 0.01 or 0.001; and the L1 ratio, which sets the proportion of the penalty determined by the L1 penalty rather than the L2, which could be 0, 0.25, 0.5, 0.75 or 1.

2.8.4. Multilayer Perceptron

A multilayer perceptron (MLP) is a type of deep learning (DL) algorithm, also known as a neural network. In a DL algorithm, individual nodes called "neurons" take inputs with weights and subject them to some kind of function to get outputs. In an MLP, in-between the input layer which takes the features and the output layer with the classification, there are multiple hidden layers each with a certain number of neurons.[32]

An MLP was used here with 1000 max iterations, with two hyperparameters to tune: the number and size of hidden layers, which could range from 1-2 layers with 50 or 100 neurons, and alpha, which represents the strength of the L2 regularization term and had 4 possible values between -1 and 1 on a logarithmic scale.

2.9. Evaluation Metrics

Several evaluation metrics were calculated for each patient: accuracy, weighted precision, weighted recall, weighted F1 score, and Receiver Operating Characteristic (ROC) Area Under the Curve (AUC). In addition, the F1 scores were calculated for each individual class. The formulas are as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$ [Equation 1]

$$Precision = \frac{TP}{TP+FP}$$ [Equation 2]

$$Recall = \frac{TP}{TP+FN}$$ [Equation 3]

$$F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)}$$ [Equation 4]

The ROC curve is a curve of the True Positive Rate, AKA recall, over the False Positive Rate, which is equal to $\frac{FP}{TN+FP}$. The statistic here is the AUC for this curve. The ROC AUC is calculated using the "one-vs-one" method, in which every pairwise combination of classes is used for "positive" and "negative".

All metrics were calculated using a weighted average across all 3 classes, except for ROC AUC. In addition, the F1 scores for each individual class were calculated as well.

2.10. Prediction Target Determination

In order to determine the optimal prediction target, several prediction target values ranging from 5 seconds to 90 seconds were tested on one patient. The testing and training accuracies, as well as the F1 scores for each class, were recorded and plotted to compare results.

2.11. Accuracy Over Time

For one patient, after the model was trained, it was used to predict the results on all the files in one fold, including the data that had been removed during class-balancing. The data was sorted according to the time relative to seizure onset, and the average accuracy was determined for each 5-second interval. This process was repeated for prediction targets of 15, 30, and 60 seconds.

3. Results

Table 3: the means and standard deviations of all accuracy, precision, recall, F1 score, and ROC AUC across all 10-patients, for each of the 4 models used, with a prediction target of 30 seconds.

Method		Accuracy	Precision	Recall	F1 Score	ROC AUC
SVM	Test	0.61 ±0.10	0.61 ±0.10	0.61 ±0.10	0.60 ±0.10	0.78 ±0.08
	Train	0.73 ±0.09	0.73 ±0.09	0.73 ±0.09	0.73 ±0.09	0.88 ±0.06
Random Forest	Test	0.64 ±0.09	0.65 ±0.09	0.64 ±0.09	0.63 ±0.09	0.82 ±0.07
	Train	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
Logistic Regression	Test	0.58 ±0.09	0.58 ±0.10	0.58 ±0.09	0.56 ±0.10	0.77 ±0.07
	Train	0.64 ±0.08	0.63 ±0.10	0.64 ±0.08	0.63 ±0.09	0.81 ±0.07
MLP	Test	0.60 ±0.10	0.61 ±0.10	0.60 ±0.10	0.59 ±0.10	0.78 ±0.08
	Train	0.72 ±0.09	0.72 ±0.09	0.72 ±0.09	0.72 ±0.10	0.88 ±0.07

Table 4: means and standard deviations of class-specific F1 scores with a prediction target of 30 seconds.

	Train			Test		
Method	**Ictal**	**Preictal**	**Background**	**Ictal**	**Preictal**	**Background**
SVM	0.72 ±0.11	0.49 ±0.13	0.58 ±0.10	0.80 ±0.10	0.67 ±0.10	0.71 ±0.09
Random Forest	0.76 ±0.11	0.51 ±0.12	0.62 ±0.09	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
Logistic Regression	0.70 ±0.09	0.48 ±0.10	0.51 ±0.15	0.74 ±0.08	0.58 ±0.09	0.62 ±0.14
MLP	0.71 ±0.12	0.49 ±0.13	0.57 ±0.10	0.79 ±0.10	0.67 ±0.10	0.70 ±0.09

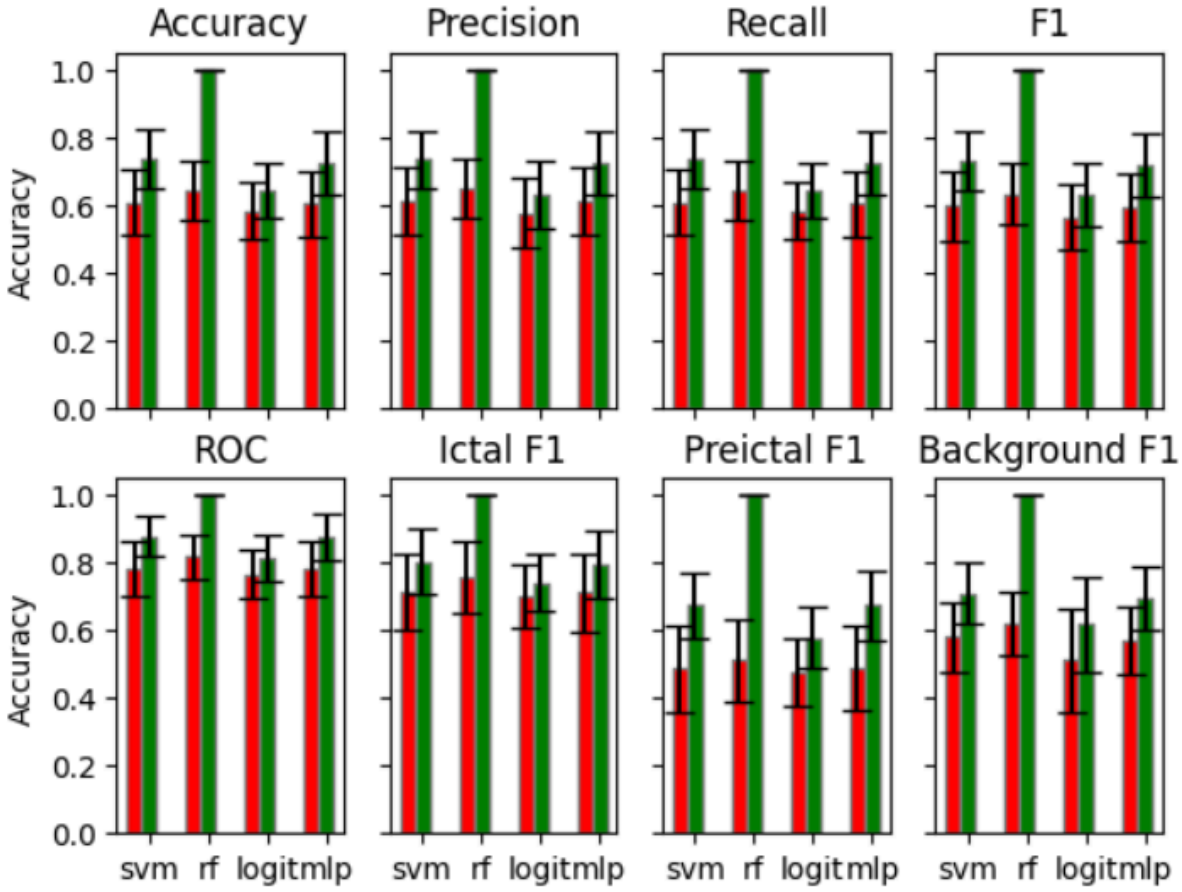

Figure 6: Bar chart of average training and testing metrics across all patients for the four methods, specifically accuracy, precision, recall, f1 score, ROC AUC, and class-specific f1 scores. The length of each error bar is the standard deviation of the 10-patient averages. Red bars are test values, green bars are training values.

The average accuracy for all patients for each method is displayed above for a prediction target of 30 seconds, with error bars equal to one standard deviation. The test accuracies are approximately 0.61, 0.64, 0.58 and 0.61 for SVM, random forest, logistic regression and multilayer perceptron, respectively. Similar values are observed for precision, recall and F1 score. The ROC AUC score is a little higher, with values of 0.78, 0.82, 0.77 and 0.78,

respectively. Random forest has the highest scores on all metrics, although all 4 methods are within one standard deviation of each other. For the class-specific scores, with random forest, they are 0.74 for ictal, 0.53 for preictal, and 0.63 for ictal.

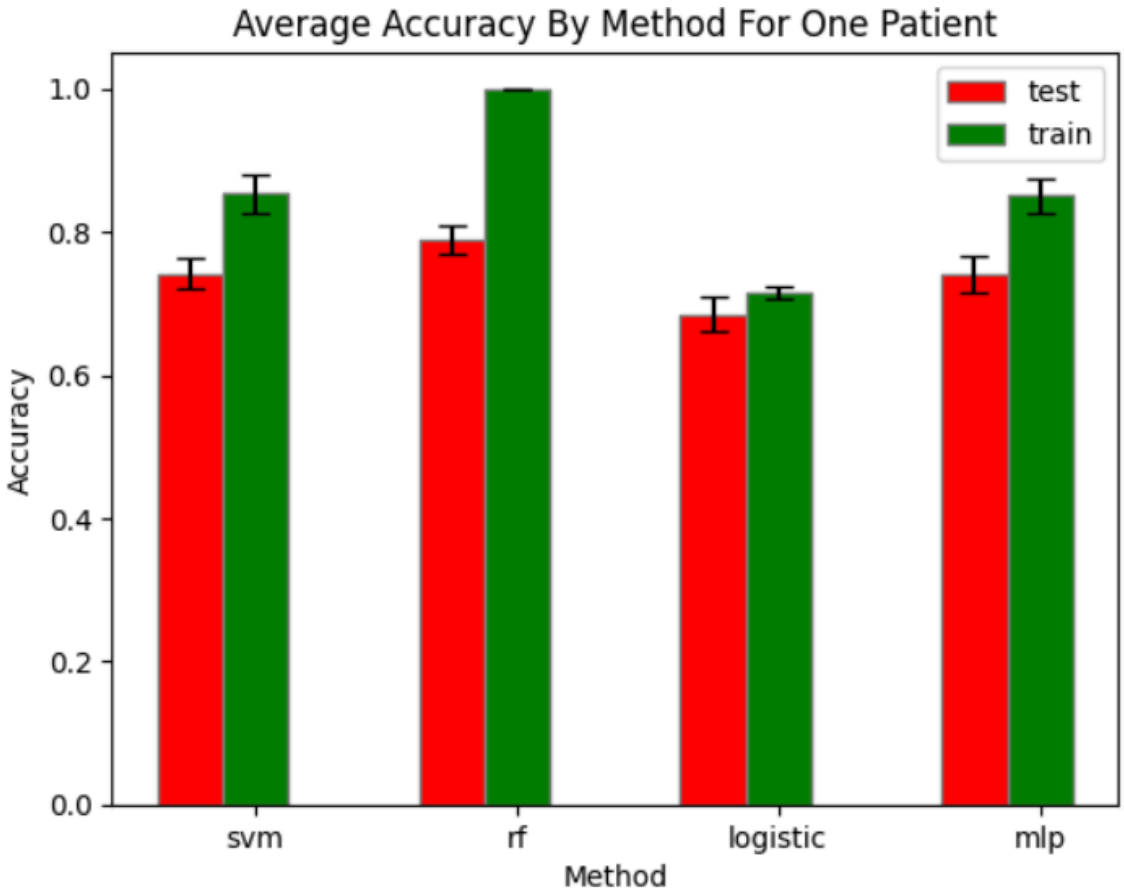

Figure 7: Average training and testing accuracy for Patient 1 for each method. The length of each error bar is the standard deviation of the 10-patient averages.

In Figure 7 above, the means and standard deviations of accuracy are included for Patient 1, on whom the model performed the best. Random forest once again proved to perform the best, and since the standard deviations are smaller, the difference in performance is more significant. A similar pattern was observed on all patients.

Table 5: means and standard deviations of performance metrics for a prediction target of 60 seconds.

Method		Accuracy	Precision	Recall	F1 Score	ROC AUC
SVM	Test	0.61 ±0.09	0.61 ±0.09	0.61 ±0.09	0.60 ±0.10	0.79 ±0.08
	Train	0.73 ±0.09	0.73 ±0.09	0.73 ±0.09	0.73 ±0.09	0.88 ±0.06
Random Forest	Test	0.65 ±0.08	0.66 ±0.08	0.65 ±0.08	0.64 ±0.09	0.83 ±0.06
	Train	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
Logistic Regression	Test	0.59 ±0.08	0.59 ±0.08	0.59 ±0.08	0.58 ±0.09	0.77 ±0.07
	Train	0.65 ±0.08	0.64 ±0.08	0.65 ±0.08	0.64 ±0.08	0.82 ±0.06
MLP	Test	0.61 ±0.10	0.62 ±0.10	0.61 ±0.10	0.60 ±0.10	0.79 ±0.08
	Train	0.72 ±0.09	0.72 ±0.09	0.72 ±0.09	0.71 ±0.09	0.87 ±0.06

Table 6: means and standard deviations of class-specific F1 scores for a prediction target of 60 seconds.

	Train			Test		
Method	**Ictal**	**Preictal**	**Background**	**Ictal**	**Preictal**	**Background**
SVM	0.73 ±0.12	0.51 ±0.11	0.57 ±0.10	0.81 ±0.10	0.68 ±0.10	0.69 ±0.10
Random Forest	0.78 ±0.11	0.54 ±0.11	0.62 ±0.10	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
Logistic Regression	0.71 ±0.09	0.49 ±0.10	0.53 ±0.10	0.75 ±0.09	0.58 ±0.09	0.60 ±0.09
MLP	0.72 ±0.12	0.51 ±0.12	0.57 ±0.10	0.80 ±0.09	0.66 ±0.10	0.68 ±0.09

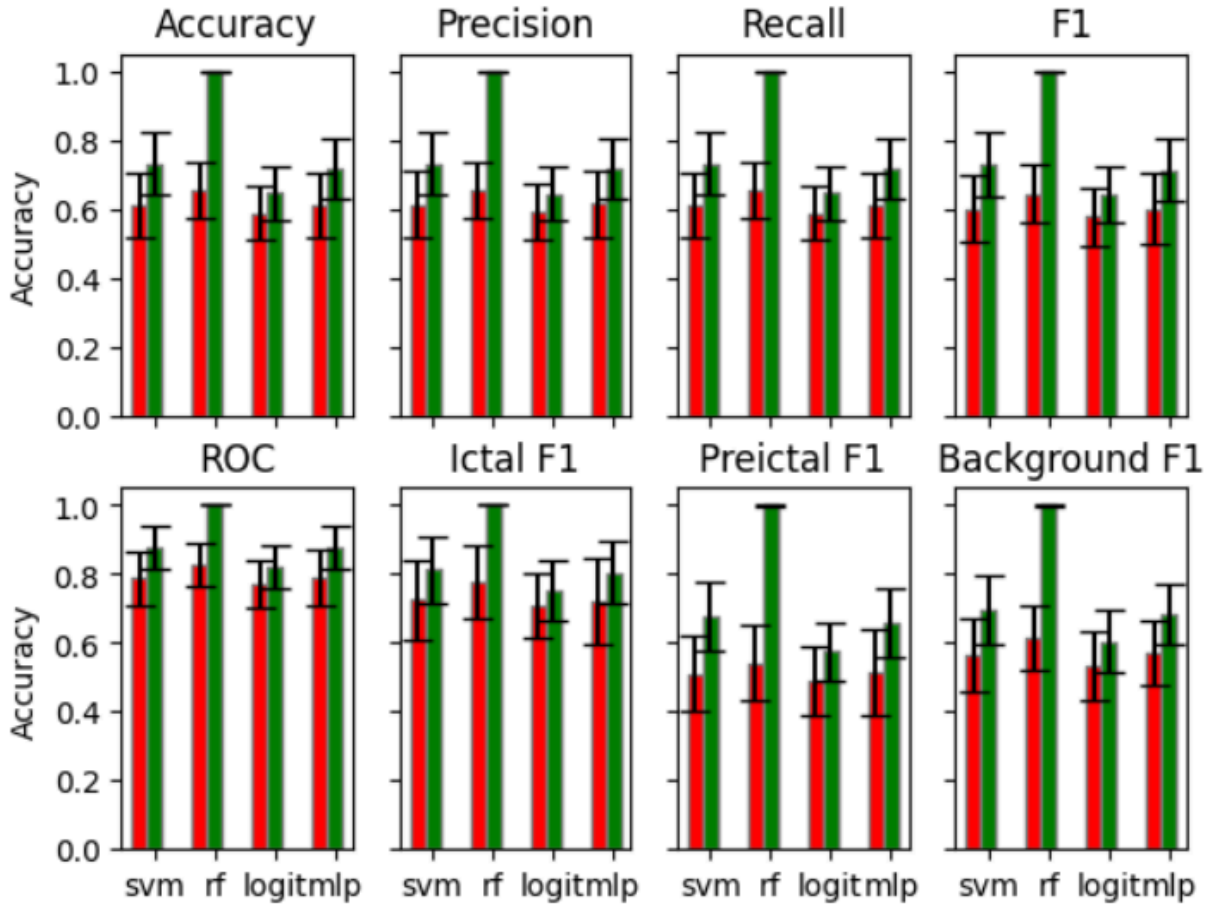

Figure 8: Performance metrics with a 60-second prediction target. The length of each error bar is the standard deviation of the 10-patient averages. Red bars are test values, green bars are training values.

When the prediction target is set to 60 seconds, the results are very similar to the 30-second case, but all the metrics are slightly higher. The same basic patterns are observed as well: random forest performs the best on all metrics, while logistic regression performs the worst.

Table 7: means and standard deviations of performance metrics for 2-class detection.

Method		Accuracy	Precision	Recall	F1 Score	ROC AUC
SVM	Test	0.80 ±0.09	0.81 ±0.08	0.80 ±0.09	0.80 ±0.09	0.87 ±0.08
	Train	0.89 ±0.06	0.89 ±0.06	0.89 ±0.06	0.89 ±0.06	0.94 ±0.04
Random Forest	Test	0.85 ±0.07	0.86 ±0.06	0.85 ±0.07	0.84 ±0.07	0.92 ±0.05
	Train	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00	1.00 ±0.00
Logistic Regression	Test	0.78 ±0.08	0.79 ±0.07	0.78 ±0.08	0.77 ±0.08	0.85 ±0.08
	Train	0.82 ±0.06	0.82 ±0.06	0.82 ±0.06	0.82 ±0.06	0.89 ±0.05
MLP	Test	0.80 ±0.09	0.81 ±0.09	0.80 ±0.09	0.80 ±0.09	0.87 ±0.09
	Train	0.88 ±0.06	0.89 ±0.06	0.88 ±0.06	0.88 ±0.06	0.95 ±0.04

Table 8: means and standard deviations of class-specific F1 scores for 2-class detection.

	Train		Test	
Method	**Ictal**	**Background**	**Ictal**	**Background**
SVM	0.80 ±0.09	0.80 ±0.09	0.89 ±0.06	0.89 ±0.06
Random Forest	0.84 ±0.08	0.85 ±0.07	1.00 ±0.00	1.00 ±0.00
Logistic Regression	0.78 ±0.07	0.77 ±0.08	0.82 ±0.06	0.82 ±0.06
MLP	0.79 ±0.10	0.80 ±0.09	0.89 ±0.06	0.88 ±0.06

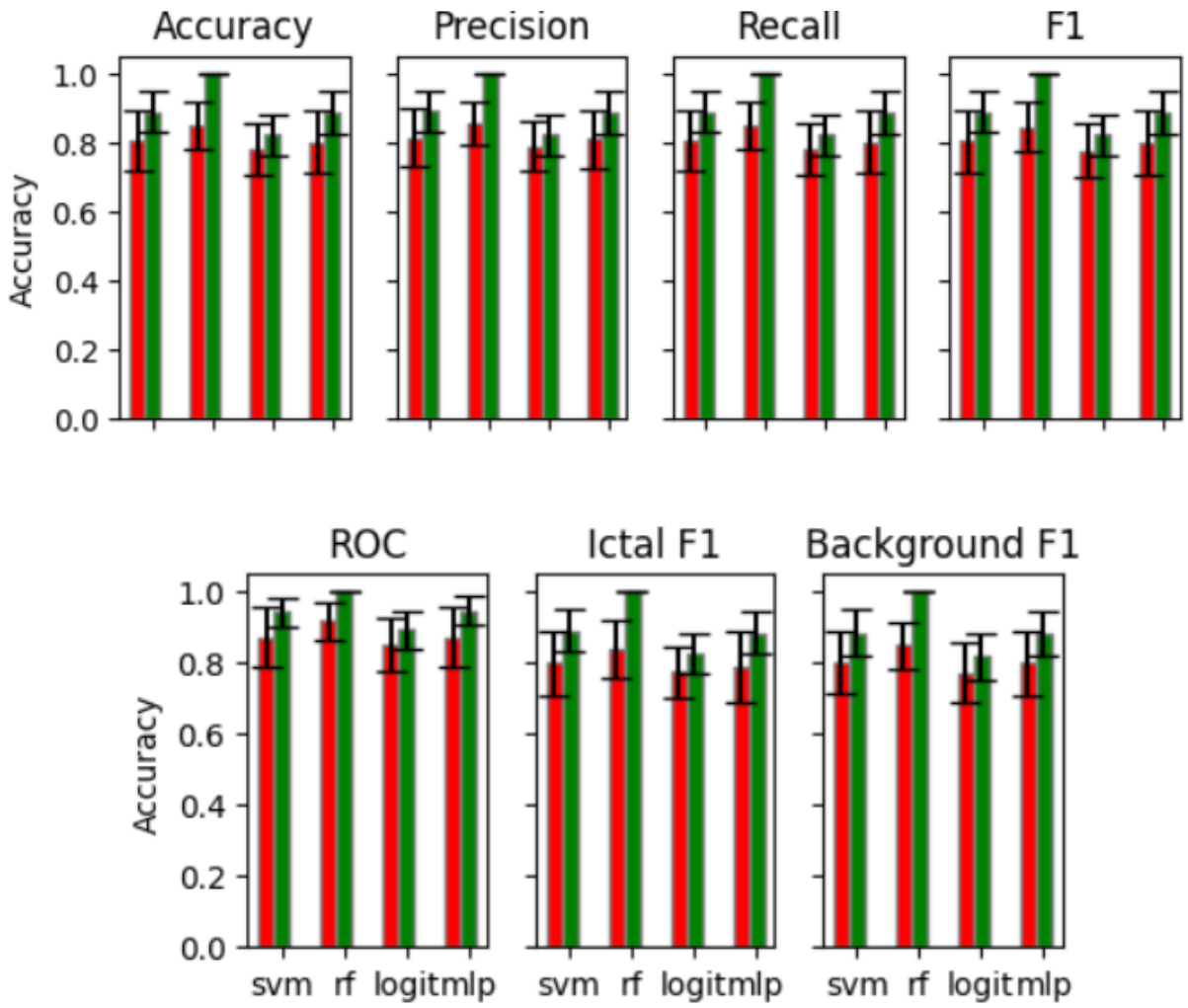

Figure 9: Average performance metrics for 2-class detection. The length of each error bar is the standard deviation of the 10-patient averages. Red bars are test values, green bars are training values.

The above figure shows the results for two-class detection, with no preictal phase included. The model performs much better in this case than for prediction and detection combined. All the accuracy metrics are between 0.75 and 0.85, including the class-specific metrics, except for the ROC AUC, which is in the 0.85-0.95 range.

4. Discussion

4.1. Model Performance

Out of the 4 models tested, random forest had the best performance, with an average accuracy of 0.64, precision of 0.65, recall of 0.65, F1 score of 0.63, ROC AUC of 0.81, and class-specific F1 scores of 0.75 for ictal, 0.53 for preictal, and 0.63 for background. All of these measures were at least a few percentage points higher than the same measures for SVM, logistic, and MLP. Similar patterns were observed on all individual patients. While random forest performed the best and logistic regression performed the worst, the differences among the 4 methods were relatively small, meaning that any of the 4 methods can be used with similar effectiveness.

4.2. Detection vs. Prediction

This feature extraction algorithm performs exceedingly well at detecting the ictal state, making it ideal for detecting seizures. However, while the ictal state is easily distinguishable, it is more difficult to distinguish between the preictal and background states. In Figure 5, this contrast is clearly illustrated: ictal epochs clearly have much more activity, with higher frequencies and amplitude, while the background and preictal epochs are much more similar to each other.

Displayed below is a confusion matrix of the accuracy for Patient 1. Ictal epochs are almost always labeled correctly, with very few ictal epochs being labeled as preictal or background. Meanwhile, while preictal and background epochs are usually labeled correctly, there is a higher number of miscategorizations between the two classes, and very few being categorized as ictal from either class. For preictal epochs, the proportion of correct classifications

is 61%, while only 33% are misclassified as background, meaning that the classifications are not completely random and they are accurate more often than not. For background epochs, the difference is even more pronounced, with 73% of epochs being classified correctly and only 20% being misclassified as preictal. This is most likely because epochs that are farther away from any seizure are more easily classified as background, whereas those closer to the start of the prediction target are harder to differentiate. The precision of each class is 87% for ictal, 72% for preictal, and 65% for background, which indicates that positive results can be trusted for the most part.

Table 9: a confusion matrix of predictions for patient 1.

Ground Truth (rows) Predictions (columns)	Ictal	Preictal	Background
Ictal	4159	185	324
Preictal	288	2847	1533
Background	353	927	3388

4.3. Prediction Target

Since our dataset does not have the preictal phase annotated, one of the biggest challenges was choosing the best "prediction target" to define the preictal phase. This approach is used very widely in the field, but there are many considerations that have to be taken into account to choose effectively. If a prediction target is too small, it may not encompass the full length of the preictal phase, and of course, the less time the patient is informed before the seizure begins, the less useful it is to alert them. However, if a prediction target is too long, it is more likely to be correct due to random chance rather than actually picking up a preictal signal.[2] In

past studies, increased prediction time has sometimes led to increased sensitivity, but also increased false alarm rate, illustrating the pitfalls of increasing the prediction target too far.[15] Prediction time has varied widely in existing studies: Bandarabadi et al only uses an 8-second prediction time, shorter than any we used, while other studies such as Usman et al. use 20-30 minutes, and Yang et al uses prediction time above an hour.[4] Truong et al uses a 30 minute "Seizure Occurrence Period" (SOP), but also factors in a 5-minute "Seizure Prediction Horizon" before the SOP starts, during which time the patient is alerted.[21]

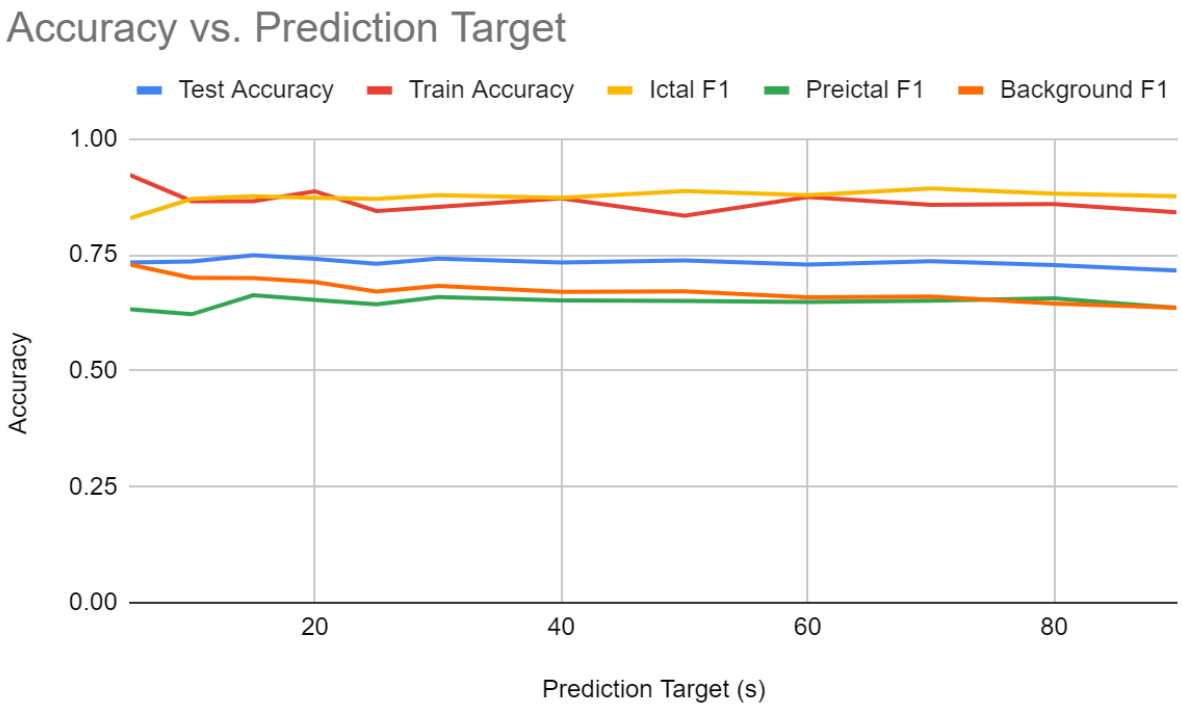

Figure 10: Training and testing accuracies, as well as class-specific F1 scores, at different prediction targets between 5 and 90 seconds.

Figure 10 depicts the accuracy of one patient at different prediction targets. There is very little variation in the accuracy, at any value from 5 seconds to 90 seconds. Based on this test, 30

seconds appeared to be the best option. However, the averages across all 10 patients were slightly higher for the 60-second case than the 30-second case; for random forest, the 30-second case had an accuracy of 0.64, and class scores of 0.75, 0.51, and 0.62 for ictal, preictal and background, respectively; the 60-second case, meanwhile, had an accuracy of 0.65, and class scores of 0.78, 0.54, and 0.62.

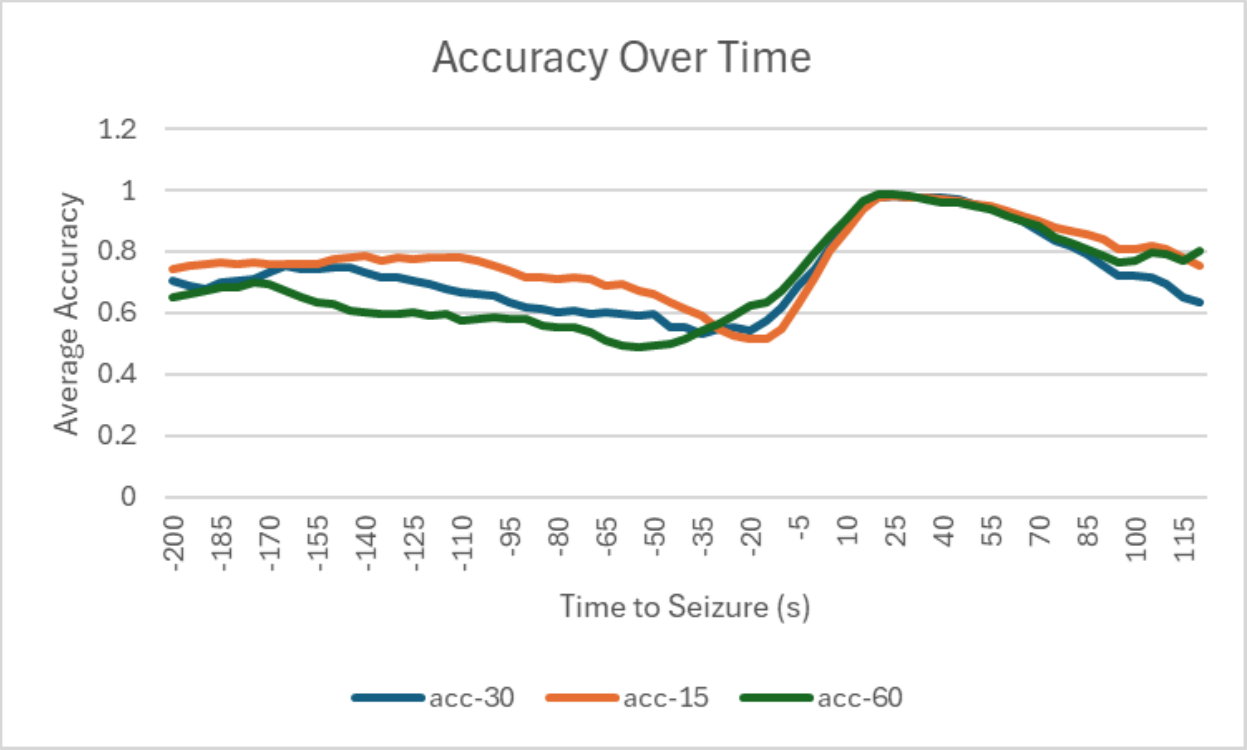

Figure 11: Average accuracy over time, relative to the onset of a seizure. Accuracy was averaged over a 30-second interval, with a 5-second sliding window. The horizontal axis displays the midpoint of each 30-second window.

In order to determine how the accuracy of the model changes over time, the time before a seizure was recorded for each epoch, and the total accuracy was taken at each 10-second interval of time to seizure. The results are displayed in Figure 11 above, with prediction targets of 15

seconds, 30 seconds, and 60 seconds. Interestingly, lower prediction targets seem to perform better at timepoints farther away from the seizure onset, while the time closest to the seizure onset could be predicted most accurately with a higher prediction target.

4.4. Variation By Patient

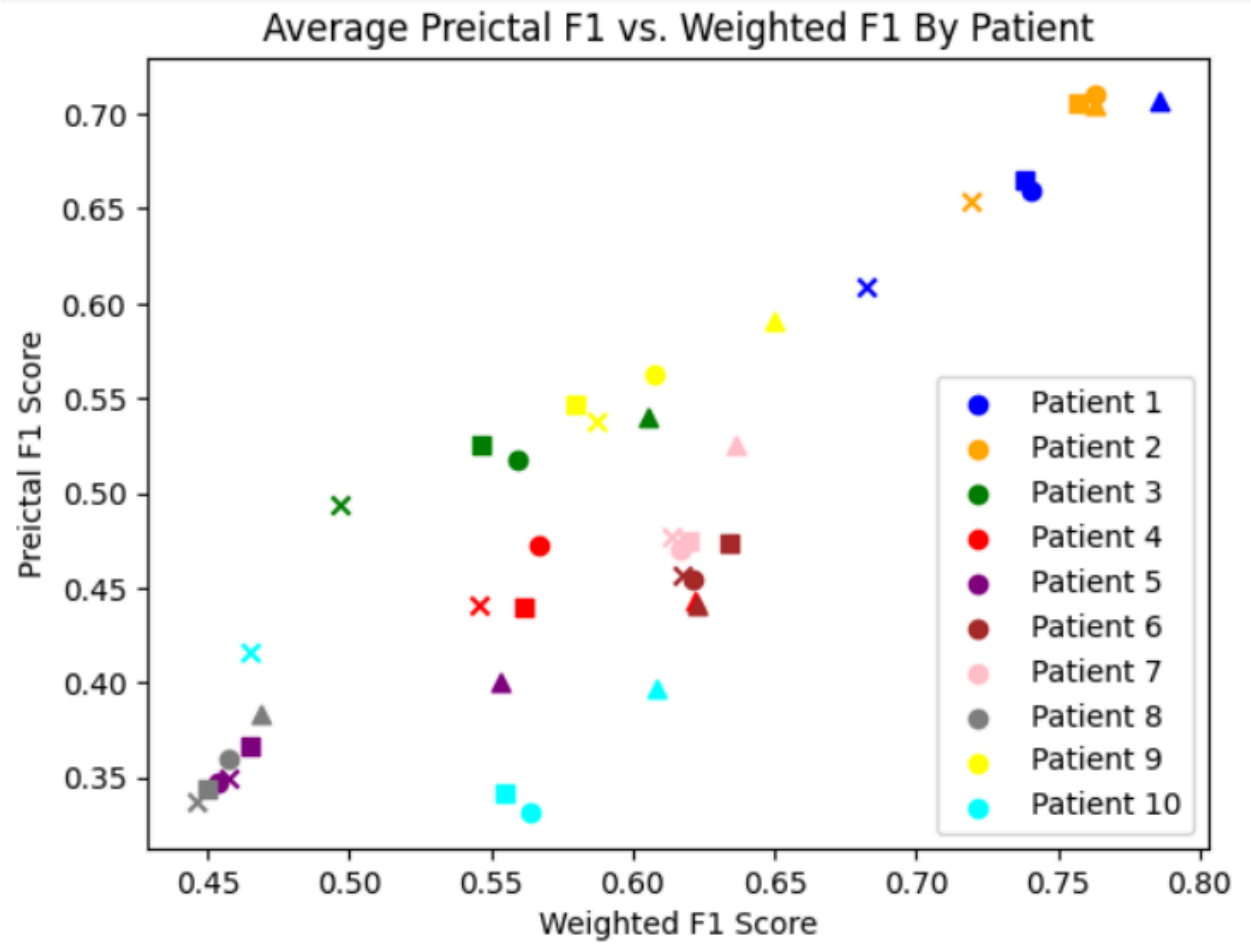

Figure 12: Scatterplot of weighted F1 scores and preictal F1 scores for each patient. Each symbol represents a different model: dots for SVM, triangles for random forest, Xs for logistic regression, and squares for MLP.

All metrics, on all models, showed very high variance based on what patient they were used on. The above scatterplot of accuracy and ictal F1 score demonstrates that both metrics vary widely from patient to patient. While some patients, like Patient 1 and Patient 2, had very strong performance on all metrics, others, like Patient 5 and Patient 8, performed very poorly.

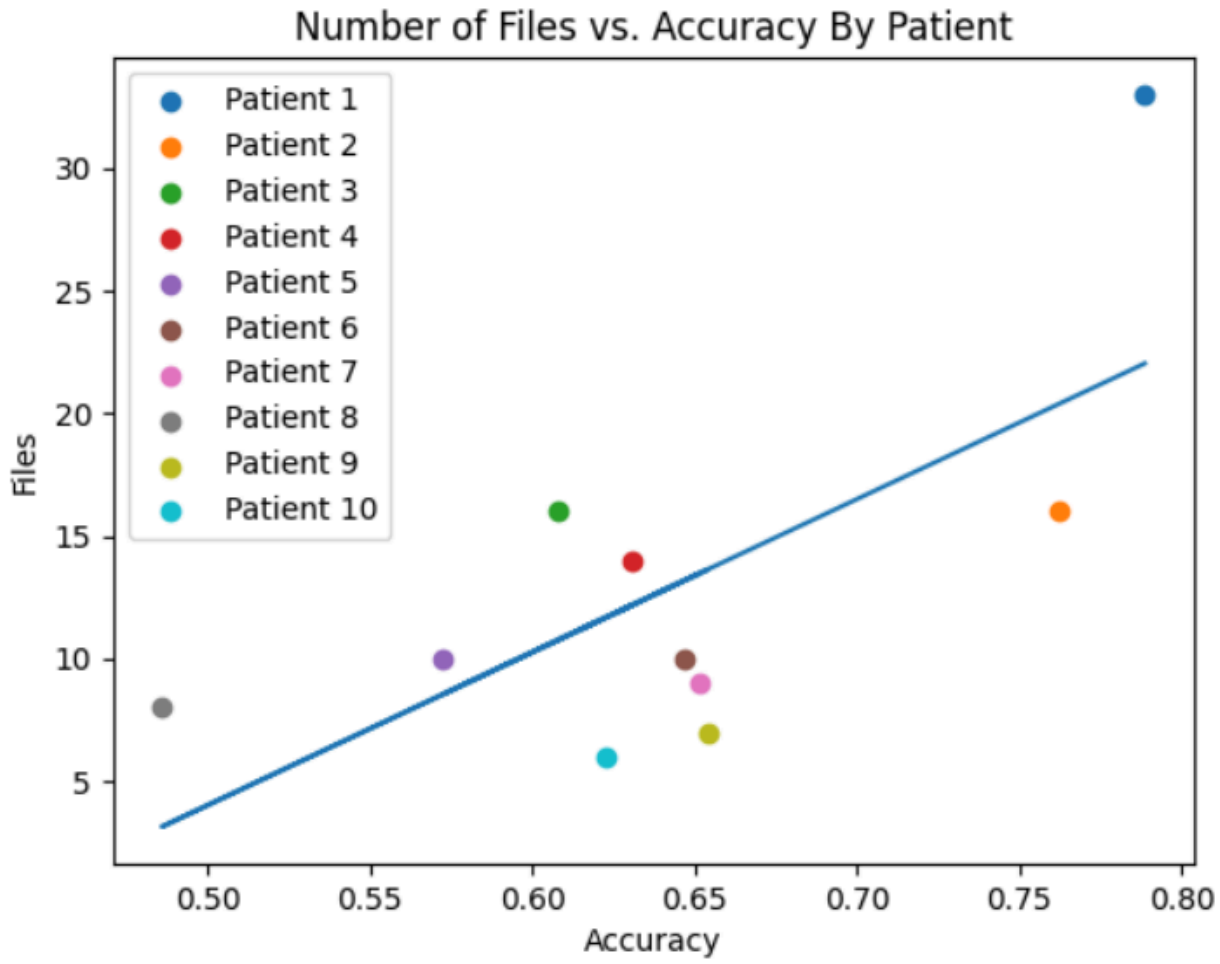

Figure 13: Scatterplot of random forest accuracy and number of EDF files for each patient, with a linear regression trendline.

The above figure plots accuracy against the number of EDF files for each patient. The coefficient of determination is 0.463, and the p-value, relative to a null hypothesis of no

correlation, is 0.03. This indicates that there is a correlation, although it still clearly does not explain everything, since Patient 1 and Patient 3 have the same number of files and yet Patient 2 performs much better. Since the number of files does not directly correspond to the amount of data, it's likely that the imbalance sizes of each file are causing part of the problem. In this case, one possible solution is to split large files into multiple small files preferably splitting around seizure events so each file still includes all 3 classes. Since the files, as they stand right now, include multiple "tokens" taken from a single session, this would not meaningfully change the data collection in any way.

One possible explanation could be that certain patients have more noise or artifacts in the data. An examination of some of the recordings for patient Patient 5, one of the worst-performing patients, shows some extreme spikes that are not likely to occur naturally.

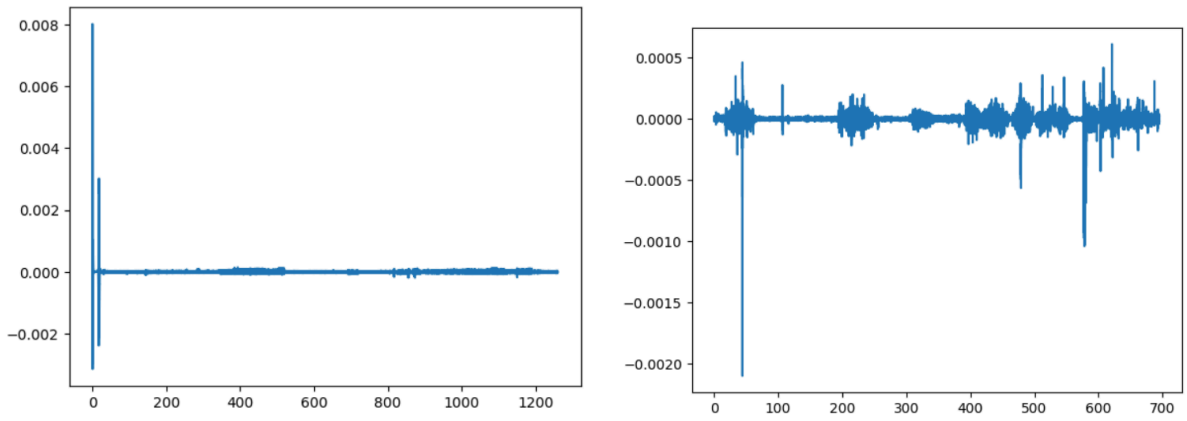

Figure 14: EEG recordings for Patient 5 (left) and Patient 1 (right).

However, it is unclear if this is the cause of the problem, as a recording from Patient 1, the best-performing patient, has a similar spike, albeit with lower magnitude.Though it is not confirmed that it is the cause of the issue, there is a high chance that these spikes are the result of an artifact and not a natural part of the signal. Therefore, further methods to properly filter the

signal should be investigated in order to result in the algorithm performing better and more consistently across all patients.

4.5. Limitations and Future Improvements

One major limitation of this model is the high runtime for each patient. One of the major reasons for the high runtime is the feature extraction phase, for which jobs cannot be run in parallel on separate CPU cores. Future directions could explore reworking the code to allow for running feature extraction jobs in parallel. Additionally, the model could be run with different combinations of features to determine which features are most necessary for the highest accuracy, and features that have little impact could be removed.

Another possible feature that could be implemented is Discrete Wavelet Transform (DWT). DWT is similar to CWT except that the scaling is done at a coarser set of values, compared to the finer scaling of CWT.[33] DWT has been used in feature extraction in the past in studies such as Bekalovna et al, which combines DWT with EMD, and Alikovic et al, which uses DWT in combination with EMD and PCA with multiple ML models and shows very high performance in distinguishing preictal, ictal and interictal signals.[34,35] Liu et al used DWT in a seizure detection model with SVM and achieved 94% specificity and 95% sensitivity.[36] Based on the high performance shown in these studies, it is possible that including DWT in this feature set could improve the model's performance.

Another area for improvement is testing different models with the same features. In particular, deep learning, also known as neural networks, has been shown to have high

performance in other studies, so it is likely that a neural network with greater complexity than the MLP could be constructed to perform well on this set of features.

It is also possible that this model could be improved by using an unsupervised learning model, as this would eliminate the need for labeled data. While this area of study is still in its infancy, a few studies have already been conducted. Several studies have used methods like K-means clustering and Gaussian mixture model on extracted features, and one study uses VAE for a deep learning method with no feature extraction.[33] Unsupervised seizure prediction has been studied as well, with one method using K-means clustering for various distance metrics and achieving accuracy above 90%.[32]

4.5.1. Applications

The algorithm in its current form has limited use in real cases. The most effective use would be for automating annotation of seizure recordings by allowing a physician to easily detect the seizures in an EEG recording. This application has value, as manually annotating large amounts of EEG data takes a very long time and it is useful to speed it up with automation. However, the most useful application of an algorithm like this is to predict a seizure before it starts. In order to do so, the algorithm would need to be applicable in real-time. Furthermore, an alarm system would be needed to alert the patient when the preictal phase has begun, so that they are not caught off-guard by the actual start of the seizure. In order to prevent false reports, the best way to do this would most likely be to only send an alert when multiple preictal epochs are detected in a row. Of course, the tradeoff here is that the more time is required to make a correct prediction, the less useful that prediction is to the patient. In an ideal situation, this would be doable on a portable device so that a patient could go about their life and get a warning before a

seizure occurs; the biggest challenge would be connecting this device to a system that can run the algorithm at a reasonable running time. For future developments, designing the hardware component for this system would be the biggest hurdle to apply the system.

4.5.2. Patient-Independent Model

One of the biggest weaknesses in a system like this is that a significant amount of training data is required from a patient in order to train the algorithm. This means that the algorithm cannot be used unless the patient has already spent significant time undergoing EEG recording, and they would have to experience seizures during the recording period; in addition, the original recordings would have to be manually annotated. While one solution to this problem, as discussed above, could be using unsupervised learning to avoid the need for manual annotation, this would not alleviate the need for recording the patient before use. In order to solve this problem, it would be ideal if an algorithm could be developed that can generalize data across patients, rather than needing to be trained on each patient individually. Some studies have already begun on patient-independent seizure detection: a 2021 study successfully performed patient-independent seizure detection with deep learning with two models 88% and 91% accuracy.[37] This is an emerging innovation in the field of seizure prediction, and further study could be a crucial advancement. While early tests on a patient-generalized application of this model were not successful, it is possible that a patient-generalized model could be developed by building off of this model and feature set, and doing so would open up many new possibilities for applying this model in real life.

5. Conclusion

This study demonstrated a model that can perform both seizure detection and prediction, by detecting the preictal phase. The model performed very well on detecting seizures; using random forest for detection, it could detect seizures at 85% accuracy averaged across all patients, and 90% on the best-performing patient. When used for both prediction and detection together, random forest achieved 64% accuracy average and 79% accuracy for the best patient; for detecting the preictal phase specifically, the performance averaged at 51% and peaked at 70%, based on F1 scores. While these accuracy values clearly could be improved, they show that this feature set includes some characteristics that can be used to distinguish the preictal phase, and with refinement it could be used for seizure prediction more effectively. For both detection and prediction, there are many ways that the model could be improved for future applications, but the current model provides a solid baseline from which to work.

Bibliography

[1] *Action potentials and synapses*. Queensland Brain Institute - University of Queensland. (2023, April 26). https://qbi.uq.edu.au/brain-basics/brain/brain-physiology/action-potentials-and-synapses

[2] Kuhlmann, L., Lehnertz, K., Richardson, M.P., Schelter, B., & Zaveri, H.P. (2018). Seizure prediction — ready for a new era. *Nature Reviews Neurology, 14*, 618–630 (2018). https://doi.org/10.1038/s41582-018-0055-2

[3] World Health Organization. (2024, February 7). *Epilepsy*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/epilepsy

[4] Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., O'Brien, T., & Razi, A. (2021). Machine learning for predicting epileptic seizures using EEG signals: A Review. *IEEE Reviews in Biomedical Engineering*, *14*, 139–155. https://doi.org/10.1109/rbme.2020.3008792

[5] Siddiqui, M. K., Morales-Menendez, R., Huang, X., & Hussain, N. (2020). A review of epileptic seizure detection using machine learning classifiers. *Brain informatics*, *7*(1), 5. https://doi.org/10.1186/s40708-020-00105-1

[6] Shah, V., von Weltin, E., Lopez, S., McHugh, J. R., Veloso, L., Golmohammadi, M., Obeid, I., & Picone, J. (2018). The Temple University Hospital Seizure Detection Corpus. *Frontiers in neuroinformatics, 12,* 83. https://doi.org/10.3389/fninf.2018.00083

[7] Ihle, M., Feldwisch-Drentrup, H., Teixeira, C. A., Witon, A., Schelter, B., Timmer, J., & Schulze-Bonhage, A. (2012). EPILEPSIAE - a European epilepsy database. *Computer methods and programs in biomedicine, 106*(3)*,* 127–138. https://doi.org/10.1016/j.cmpb.2010.08.011

[8] Kini, L. G., Davis, K. A., & Wagenaar, J. B. (2016). Data integration: Combined imaging and electrophysiology data in the cloud. *NeuroImage, 124*(Pt B), 1175–1181. https://doi.org/10.1016/j.neuroimage.2015.05.075

[9] Shoeb, A. & Guttag, J. (2010). Application of Machine Learning To Epileptic Seizure Detection. *ICML 2010 - Proceedings, 27th International Conference on Machine Learning.* 975-982.

[10] Wong, S., Simmons, A., Rivera-Villicana, J., Barnett, S., Sivathamboo, S., Perucca, P., Ge, Z., Kwan, P., Kuhlmann, L., Vasa, R., Mouzakis, K., & O'Brien, T. J. (2023). EEG datasets for seizure detection and prediction- A review. *Epilepsia open*, *8*(2), 252–267. https://doi.org/10.1002/epi4.12704

[11] Detti, P., Vatti, G., Zabalo Manrique de Lara, G. (2020). EEG Synchronization Analysis for Seizure Prediction: A Study on Data of Noninvasive Recordings. *Processes 2020, 8*(7), 846; https://doi.org/10.3390/pr8070846

[12] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Physical review. E, Statistical, nonlinear, and soft matter physics, 64*(6 Pt 1), 061907. https://doi.org/10.1103/PhysRevE.64.061907

[13] Gupta, A., Singh, P., & Karlekar, M. (2018). A Novel Signal Modeling Approach for Classification of Seizure and Seizure-Free EEG Signals. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society, 26*(5), 925–935. https://doi.org/10.1109/TNSRE.2018.2818123

[14] Baldassano, S. N., Brinkmann, B. H., Ung, H., Blevins, T., Conrad, E. C., Leyde, K., Cook, M. J., Khambhati, A. N., Wagenaar, J. B., Worrell, G. A., & Litt, B. (2017). Crowdsourcing seizure detection: algorithm development and validation on human implanted device recordings. *Brain : a journal of neurology, 140*(6), 1680–1691. https://doi.org/10.1093/brain/awx098

[15] Usman, S. M., Khalid, S., Akhtar, R., Bortolotto, Z., Bashir, Z., & Qiu, H. (2019). Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: Review of available methodologies. *Seizure*, *71*, 258–269. https://doi.org/10.1016/j.seizure.2019.08.006

[16] Subasi, A., Kevric, J. & Abdullah Canbaz, M. (2019). Epileptic seizure detection using hybrid machine learning methods. *Neural Computing & Applications 31*, 317–325 https://doi.org/10.1007/s00521-017-3003-y

[17] O.K., F., & R., R. (2019). Time-domain exponential energy for epileptic EEG Signal Classification. *Neuroscience Letters*, *694*, 1–8. https://doi.org/10.1016/j.neulet.2018.10.062

[18] Wang, X., Gong, G., Li, N., & Qiu, S. (2019). Detection Analysis of Epileptic EEG Using a Novel Random Forest Model Combined With Grid Search Optimization. *Frontiers in human neuroscience, 13,* 52. https://doi.org/10.3389/fnhum.2019.00052

[19] Tzimourta, K.D., Tzallas, A.T., Giannakeas, N., Astrakas, L.G., Tsalikakas, D.G., Angelidis, P., & Tsipouras, M.G. (2019) A robust methodology for classification of epileptic seizures in EEG signals. *Health and Technology, 9,* 135–142. https://doi.org/10.1007/s12553-018-0265-z

[20] Seng, C. H., Demirli, R., Khuon, L., & Bolger, D. (2012). Seizure detection in EEG signals using support vector machines. *2012 38th Annual Northeast Bioengineering Conference (NEBEC)*. https://doi.org/10.1109/nebc.2012.6207048

[21] Truong, N.D., Ngueyn, A.D., Kuhlmen, L., Bonyadi, M.R., Yang, J., Ippolito, S., & Kavehei, O. (2018). Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks, 105*, pp. 104-111. https://doi.org/10.1016/j.neunet.2018.04.018

[22] Chakrabarti, S., Swetapadma, A., Pattnaik, P. K., & Samajdar, T. (2017). Pediatric seizure prediction from EEG signals based on unsupervised learning techniques using various distance measures. *2017 1st International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech)*. https://doi.org/10.1109/iementech.2017.8076983

[23] Yıldız, İ., Garner, R., Lai, M., & Duncan, D. (2022). Unsupervised seizure identification on EEG. *Computer methods and programs in biomedicine, 215,* 106604. https://doi.org/10.1016/j.cmpb.2021.106604

[24] McCallan, N., Davidson, S., Ng, K.U., Biglarbeigi, P., Finlay, D., Lan, B.L., & McLaughlin, J. (2023). Epileptic multi-seizure type classification using electroencephalogram signals from the Temple University Hospital Seizure Corpus: A review. *Expert Systems with Applications, 234,* 121040. https://doi.org/10.1016/j.eswa.2023.121040

[25] Hjorth, Bo; Elema-Schönander, AB (1970). EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology. 29 (3)*. 306–310. doi:10.1016/0013-4694(70)90143-4

[26] Wijayanto, I., Hartanto, R., Nugroho, H. A., & Winduratna, B. (2019) Seizure Type Detection in Epileptic EEG Signal using Empirical Mode Decomposition and Support Vector Machine. *2019 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 314-319. doi.org/10.1109/ISITIA.2019.8937205.

[27] *Continuous wavelet transform and scale-based analysis*. MATLAB & simulink. (n.d.). https://www.mathworks.com/help/wavelet/gs/continuous-wavelet-transform-and-scale-based-analysis.html

[28] *What is Principal Component Analysis (PCA)?*. IBM. (2023, December 4). https://www.ibm.com/topics/principal-component-analysis

[29] *What are Support Vector Machines?*. IBM. (2023, December 12). https://www.ibm.com/topics/support-vector-machine

[30] *What is Random Forest?*. IBM. (2021, October 20). https://www.ibm.com/topics/random-forest

[31] *What is logistic regression?*. IBM. (2021, August 16). https://www.ibm.com/topics/logistic-regression

[32] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems, 2(4),* 303–314.

[33] *Continuous and discrete wavelet transforms*. MATLAB & simulink. (n.d.). https://www.mathworks.com/help/wavelet/gs/continuous-and-discrete-wavelet-transforms.html

[34] Bekbalanova, M., Zhunis, A., & Duisebekov, Z. (2019). Epileptic seizure prediction in EEG signals using EMD and DWT. *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)*. https://doi.org/10.1109/icecco48375.2019.9043270

[35] Alickovic, E., Kevric, J., & Subasi, A. (2018). 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. https://doi.org/10.1016/j.bspc.2017.07.022

[36] Liu, Y., Zhou, W., Yuan, Q., & Chen, S. (2012). Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *20*(6), 749–755. https://doi.org/10.1109/tnsre.2012.2206054

[37] Dissanayake, T., Fernando, T., Denman, S., Sridharan, S., & Fookes, C. (2021) Deep Learning for Patient-Independent Epileptic Seizure Prediction Using Scalp EEG Signals. *IEEE Sensors Journal, 21(7),* 9377-9388. https://doi.org/10.1109/JSEN.2021.3057076