University of Toronto

Faculty of Applied Science and Engineering

BME 1473: Acquisition and Processing of Bioelectrical Signals

Project Report

Signal Processing Techniques to Improve Feature Space for EEG-based Epileptic Seizure Detection

Name	**Student Number**
Kalana Gayal Abeywardena	1009505867

This report is submitted in partial fulfillment of the requirements for the module BME 1473: Acquisition and Processing of Bioelectrical Signals

December 2022

Content

List of Tables

List of Figures

1. Introduction

Bio-signal processing is an evolving field of science that is critical in the diagnosis, and prognosis of diseases and timely intervention to reduce the impact of the disease on the patients. In recent years, the research interests have been largely shifted to analyze the functionality of the brain to identify the structural and functional changes and characteristics that correspond to different types of neurological disorders including epileptic seizures, and Alzheimer's. In this project, the broad objective is oriented toward detecting epileptic seizures by analyzing Electroencephalography (EEG) signals recorded in-vivo from the patients diagnosed with the disease.

Manual review of EEG signals is a time-consuming, expensive, tedious activity that is prone to error [1]. In this project, intracranial EEG signals captured at hippocampus formation are preprocessed using signal processing techniques described in Section 2.1 and extracted the features using the techniques described in Section 2.2. The project aims at identifying the effects of the analyzed signal techniques toward accurate seizure detection using EEG signals based on quantitative metrics described in Section 2.3.

1.1. Objective

The EEG signals are generally noisy and non-stationary and hinder the capability of using pattern recognition methodologies to identify the unique differences between Inter-Ictal and Ictal signals to detect epilepsy seizures accurately. The principal objective of this project is to analyze the impact of preprocessing and feature engineering techniques on the final epilepsy detection accuracy.

1.2. Dataset

For this project, a publicly available Epileptogie dataset [2] is used which consists of 5 subsets, each containing 100 EEG segments. The data is sampled at a sampling rate of 173.61 Hz which is recorded using a 128-channel amplifier system with an average common reference. The recorded signals have a spectral bandwidth of 0.5 Hz to 85 Hz. These continuous multichannel EEG recordings from multi-spatial locations are segmented into 23.6s long epochs after visually inspecting the presence of any artifacts [2]. For this project, I selected Intracranial EEG signals that are captured within hippocampus formation. A summary of the dataset is shown in Table 1.

Table 1: Summary of Dataset

Sampling Frequency	173.63 Hz
Inter-Ictal EEG Signal dataset (D)	100 segments
Ictal EEG Signal dataset (E)	100 segments
Recording Site	Hippocampus Formation
Type of Data	Intracranial EEG
Segment Length	23.6s (4097 samples)

2. Methodology

In this section, I discuss the different signal-processing techniques that are used for preprocessing and feature extraction from EEG signals to analyze their impact on seizure detection. The overall project pipeline is shown in Figure 1.

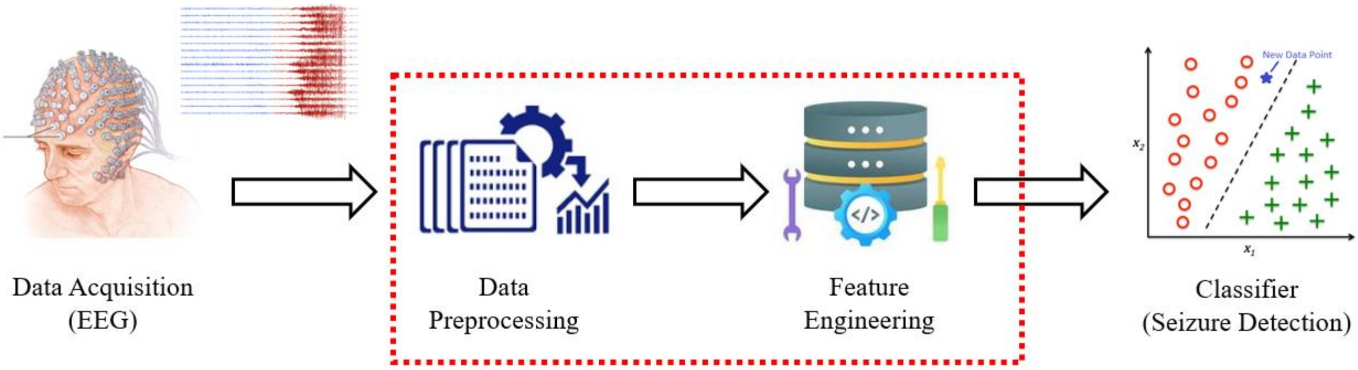


Figure 1: Overall project idea

In the subsequent sections, the methods applied for data preprocessing and feature engineering is discussed in detail.

2.1. Data Preprocessing

I analyze several preprocessing techniques to prepare data for the end classification problem. The preprocessing is done in two steps: (1) sphering the data, and (2) attenuating the noise of the EEG signals.

2.1.1. Data Sphering

As the first step, I perform detrending and whitening of EEG data to achieve weak stationarity such that the subsequent transformations are possible. Generally, EEG signals are not stationary; their statistical characteristics change with time due to the randomness in brain activity. As the quasi-stationarity only lasts for 0.25s in brain activity [3] and since the signals in the dataset are of length 23.6s, data sphering plays an important role in EEG-based seizure detection.

Data sphering is analyzed in two sequential processes: (1) data detrending, and (2) data whitening. The data detrending is simple where the mean of each EEG signal is removed. For data whitening, I adopt *Principal Component Analysis (PCA)* as the primary method and *Zero Component Analysis (ZCA)* as an alternative.

During whitening, we apply a whitening matrix $W \in \mathbb{R}^{N \times N}$ on the original EEG signals to get the uncorrelated dataset $x_w = Wx$. In both PCA and ZCA, the first step is to compute the covariance matrix of the original data $\Sigma_{original} \in \mathbb{R}^{N \times N}$ and use spectral decomposition to obtain Q, the eigenvectors, and Λ the eigenvalues. We obtain $W^{PCA} = \Lambda^{-\frac{1}{2}}Q^T$ and $W^{ZCA} = Q\Lambda^{-\frac{1}{2}}Q^T$ and apply it to the original data to get the whitened data.

2.1.2. Denoising EEG signals

The second part of the data preprocessing is to attenuate the possible artifacts and noise components that result from different sources including the brain rhythms that are caused by other neural activities. In this stage, denoising is performed by attenuating *frequencies of interest* which generally resides in high-frequency bands that capture the effects of some artifacts and line noise. However, this does not suppress any noise that oscillates at similar frequencies as the interested signal portion which is usually difficult [4].

Generally, EEG signals contain several sub-band waveforms that correspond to different rhythmic brain activities and are summarized in Table 2.

Table 2: Rhythmic Brain Activity Sub-bands in EEG Signals

Sub-band	**Frequency Range**
Delta (δ)	2 – 4 Hz
Theta (θ)	4 – 8 Hz
Alpha (α)	8 – 12 Hz
Beta (β)	15 – 30 Hz
Lower Gamma (γ_{low})	30 – 80 Hz
Upper Gamma (γ_{high})	80 – 150 Hz

Based on [2], the spectral components above 40 Hz that corresponds to γ brain rhythms are usually irrelevant for seizure detection and hence it is considered to be noisy. For this purpose, I use *Finite Impulse Response (FIR)*-based low-pass filtering as the primary method and *Discrete Wavelet Transformation (DWT)*-based denoising as the alternative.

I. FIR-based Low-Pass Filtering

FIR filters are of finite length and maintain a linear phase response within the passband. Thus, compared to the *Infinite Impulse Response (IIR)* filters, FIR filters are practically useful in EEG denoising as the artifacts imposed on the filtered signals within the passband are limited. Generally, the following two parameters are critical in defining the FIR filter and based on the options the characteristics of the FIR filter change.

- *Filter Window Type:* This affects the passband ripple, stopband attenuation, and the transition band gap which corresponds to the spectral leakage during the filtering. For this project, I analyzed three window functions: (1) rectangular window, (2) Hamming window, and (3) Blackman-Harris window.
- *Filter Order (or window length):* This determines how sharp the transition band should be. Higher the filter order, the complexity of the FIR filter increases as the number of filter coefficients increases. For this project, I analyzed the filter orders $n = \{16, 32, 64, 128\}$.

Further, the cut-off frequency is another important parameter. However, in this case, since the established work on seizure detection defines the useful rhythmic brain waves that are used for seizure detection [2], I am using the cut-off frequency of 40 Hz.

II. DWT-based Denoising

As the alternative, I used DWT to decompose the signal into multiple bands and attenuate the spectral components in the higher resolutions assuming that the noise operates at the higher resolutions. To attenuate, a hard thresholding-based method is used. For the two highest resolutions, the average fluctuation of the signal (i.e., the variance of the noise) is estimated. As suggested by [4], the threshold was calculated as 7 times the average fluctuation and suppressed the signal fluctuations that exceed the threshold value. For this method, I identified two critical parameters that need further analysis.

- *Wavelet Function:* This is critical to decompose the EEG signal into a sub-band since the shape of the mother wavelet we try to match with the EEG signal changes the wavelet coefficients in each sub-band. For the denoising, I analyzed 4 wavelet functions: (1) Daubechies with 2 taps (db2), (2) Daubechies with 6 taps (db6), (3) Daubechies with 8 taps (db8), and (4) Haar.
- *Filter bank levels:* This defines the number of analysis-synthesis filters in the DWT. This may impact the smoothness of the wavelet function as the level parameter defines the length of the filter. Based on the initial inspection, I keep the number of levels at 6 as the highest resolutions are unchanged in the splits.

2.2. Feature Extraction and Classification

2.2.1. Feature Extraction

As the next step of the project pipeline, useful features are extracted from the preprocessed EEG signals. As the primary method, I used time-frequency features using discrete wavelet coefficients. For this method, I identified three critical parameters that need further analysis.

- *Wavelet Transformation:* There are three main transformation types possible: (1) continuous wavelet transforms (CWT), (2) discrete wavelet transforms (DWT), and (3) undecimated discrete wavelet transform (UDWT). For this project, I further analyzed the latter two transformations as CWT is computationally complex as it allows for continuous scaling and shifting of a wavelet. Further, unlike DWT, UDWT uses both odd and even transformations at each scale. UDWT is a more computationally intensive method than DWT but theoretically can result in better precise frequency localization.
- *Wavelet Function:* For the feature extraction, I analyzed 5 wavelet functions: (1) db2, (2) Daubechies with 4 taps (db4), (3) db6, (4) db8, and (5) Haar.
- *Filter bank levels:* Based on the initial inspection, I keep the number of levels at 6 to limit the search space.

Using the wavelet coefficients extracted from each configuration the features presented in Table 3 are computed. These features are selected based on [3], [5].

As alternatives, we can explore time-domain or frequency-domain features that can be obtained by Short Time Fourier Transform (STFT), Power Spectral Density (PSD), and statistical features for windowed time-domain signals. However, based on the literature, the most useful features are time-frequency features thus I limit my search to wavelet-based time-frequency features.

Table 3: Wavelet Features Extracted

Feature	Description
Log-Sum Energy	Log-transformed wavelet coefficients to achieve robustness to noise
Mean of Absolute Value	Average of absolute wavelet coefficients for each sub-band
Average Power	Average of wavelet coefficients for each sub-band
Standard Deviation	Standard deviation (variance) of wavelet coefficients
Ratio of the Absolute Mean values	Ratios of mean absolute features of adjacent sub-bands

2.2.2. Binary Classification

As the final part of this project, the extracted features are used to detect EEG signal patterns and predict the seizure or non-seizure states. This is not a part of the project that needs further analysis. Hence, I use a Support Vector Machine (SVM) as the classifier with the linear kernel as it is robust for challenging samples that are hard to separate.

2.3. Evaluation Metrics

The following metrics are used for the evaluation of techniques that are used in this project.

Table 4: Metrics used in Project

Metric	Definition
Signal Distortion (dB)	$10\log\left(P^2_{filtered,0-40}/P^2_{unfiltered,0-40}\right)\ dB$
Noise Reduction (dB)	$10\log\left(P^2_{filtered,>40}/P^2_{unfiltered,>40}\right)\ dB$
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$ %
Precision	$TP/(TP+FP)$ %
Sensitivity/Recall	$TP/(TP+FN)$ %
Specificity	$TN/(TN+FP)$ %
F1 Score	$(2\times TP)/(2TP+FP+FN)$ %
Area Under the Curve (AUC)	Area of the Receiver Operating Characteristic (ROC) curve

Where $P^2_{i,0-40}$ is the power of the signal of interest (within 0 – 40 Hz), $P^2_{i,>40}$ is the power of noise of interest (above 40 Hz) for $i\ \in \{filtered, unfiltered\}$, TP is the number true positives, FP is the number false positives, TN is the number true negatives, and FN is the number false negatives of predictions.

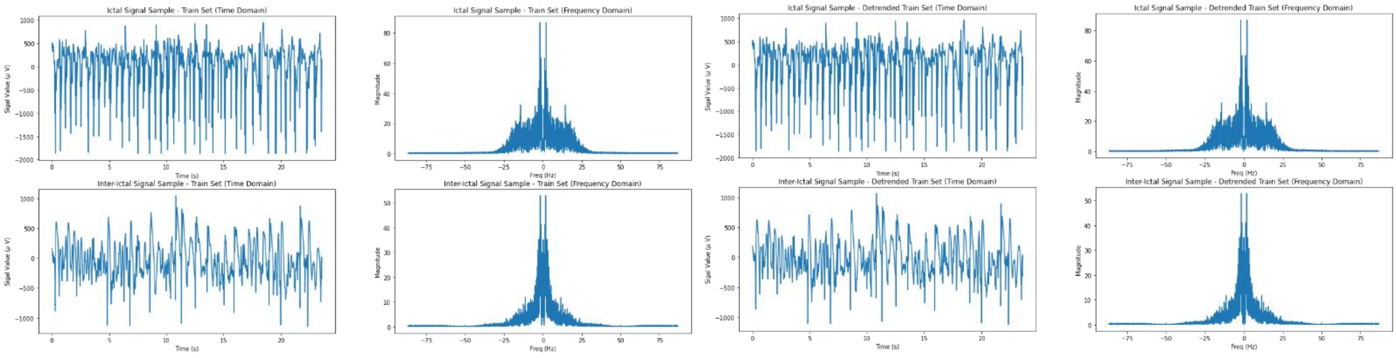
3. Results and Discussion

In this section, the results obtained for the methods discussed in Section 2 are presented concisely and discuss the observations through quantitative and qualitative analysis. For the analysis purpose, the whole dataset was divided into two partitions with balanced data where partition A has 160 EEG samples and partition B has 40 EEG samples. All the results presented here are based on partition A and partition B is used to analyze the generalizability of the chosen techniques from the analysis.

3.1. Data Preprocessing

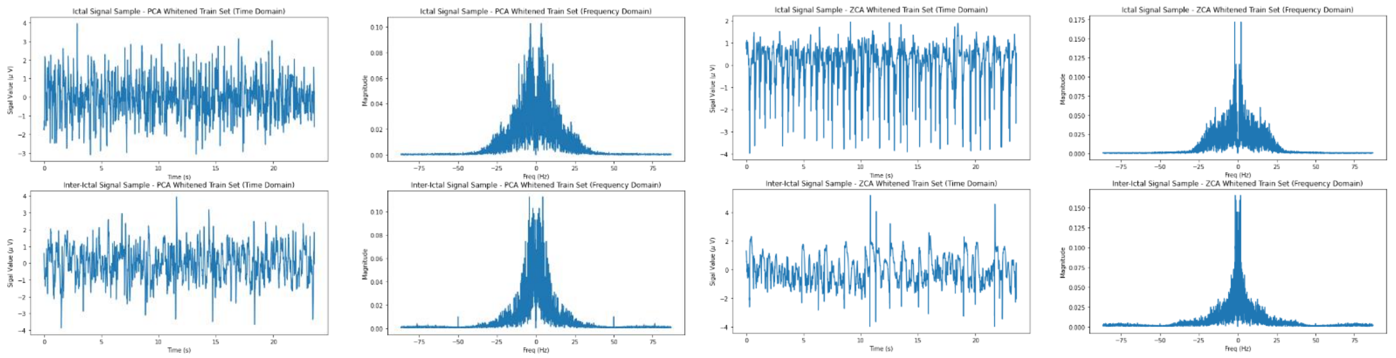
3.1.1. Data detrending and whitening

In Figure 2, two EEG signal samples are visualized for different preprocessing steps applied to observe their effects qualitatively on time and frequency domain representations. As can be seen, the original EEG signals are well-processed in the dataset thus the detrending had a marginal effect on achieving stationarity.



(a) Original EEG signal sample

(b) Detrended original signal sample



(c) PCA whitened original signal sample

(d) ZCA whitened original signal sample

Figure 2: Visualization of EEG sample in Time and Frequency-domain under different data preprocessing techniques applied

Comparing the original signal sample to the PCA-whitened signals, we can see that both the Ictal and Inter-Ictal signals have most of the oscillations within ± 1 from its mean. However, it has changed the spectrum of the original signal. Specifically for Ictal signals, the spectrum now looks more like Gaussian. The output from each technique is evaluated using the classification metrics in Section 2.3 using DWT features obtained by db6. Based on the comparison shown in Figure 3, ZCA has the lowest performance in all the metrics while PCA has marginally improved the performance of binary classification compared to detrended EEG signal classification. Thus, based on this comparison, I move forward with PCA-whitened EEG data for the subsequent analysis.

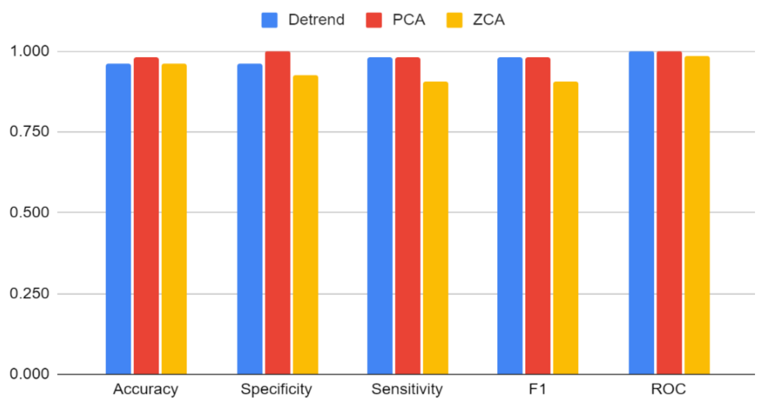
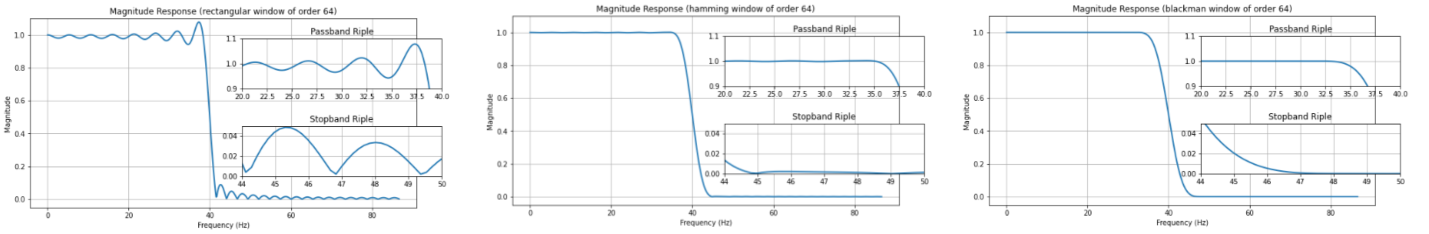


Figure 3: Quantitative performance comparison between different data preprocessing techniques

3.1.2. FIR Filtering for denoising and artifact removal at high-frequency bands

The PCA-whitened signals are then processed to remove frequency components above 40 Hz. With a cut-off frequency fixed at 40 Hz, the three types of window functions are first analyzed using a filter order at $n = 64$ as it provides a better trade-off between filter complexity and performance. The qualitative evaluation is presented in Figure 4.

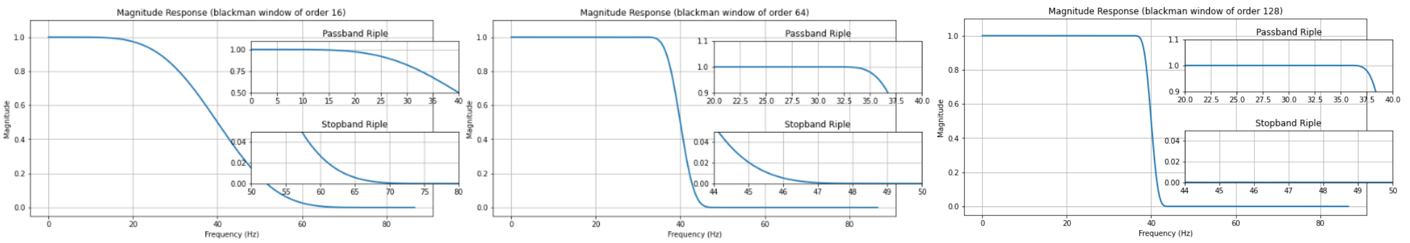


(a) Rectangular window (b) Hamming window (c) Blackman window

Figure 4: Comparison of different window types at n = 64 and cut-off frequency = 40 Hz

By observing the magnitude responses, we can see that a rectangular window with the sharpest transition band suffers from significant passband ripples that distort the magnitude of the filtered EEG signals in the time domain. With Hamming window, the passband ripples are not as significant as with the rectangular window. However, it has a slightly wider transition band compared to a rectangular window which does not have a sharp cutoff. Blackman-Harris window has the best magnitude response with 0 passband or stopband ripples and the widest transition band. However, this doesn't result in significant spectral leakage. Thus, I limit the subsequent search space to the Blackman-Harris window.

Next with the selected Blackman window, I analyze how different filter orders change the magnitude characteristics of the filter which is presented in Figure 5.

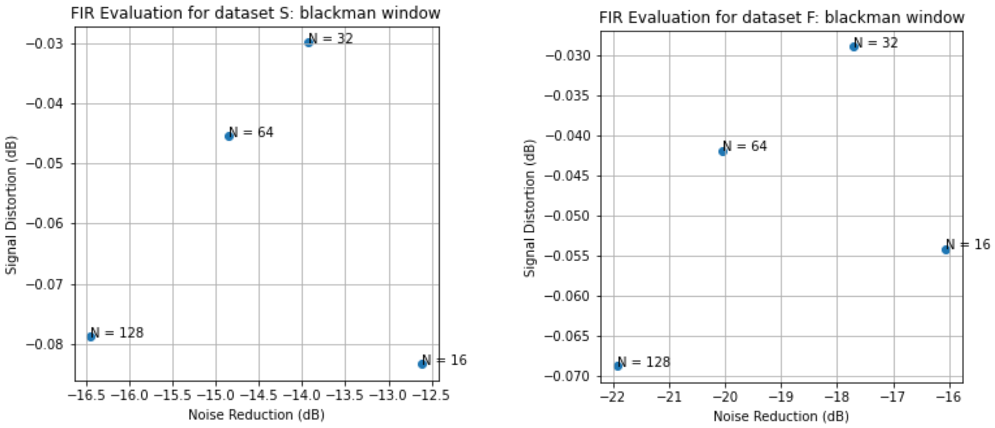


(a) Order = 16 (b) Order = 64 (c) Order = 128

Figure 5: Comparing different filter orders (n) with Blackman window-based FIR filters

We can see that with $n = 16$, the transition band is much wider and causes significant spectral leakage. Further, significant spectral information loss happens in the useful signal band (0 – 40 Hz). Intuitively, this will create more signal distortions within the useful signal region while keeping a significant portion of noise in the higher frequency band. A more precise quantitative analysis using *signal distortion* and *noise reduction* metrics is conducted and presented in Figure 6.

For better performance, the filter should provide a signal distortion $\cong 0\ dB$ as we need the signal band to be identical between the original and filtered signals while noise reduction is significant with a higher negative dB value. Based on the evaluations performed on Ictal and Inter-Ictal signals, $n = 64$ is the ideal filter order that achieves the middle ground between signal distortion and noise reduction. It seems $n = 128$ is better with significant noise reduction, however, the filter complexity is the highest for marginal performance gain. Thus, for the subsequent comparisons for feature extraction methods, I am selecting the Blackman-Harris window with $n = 64$ at a cut-off frequency of 40 Hz.

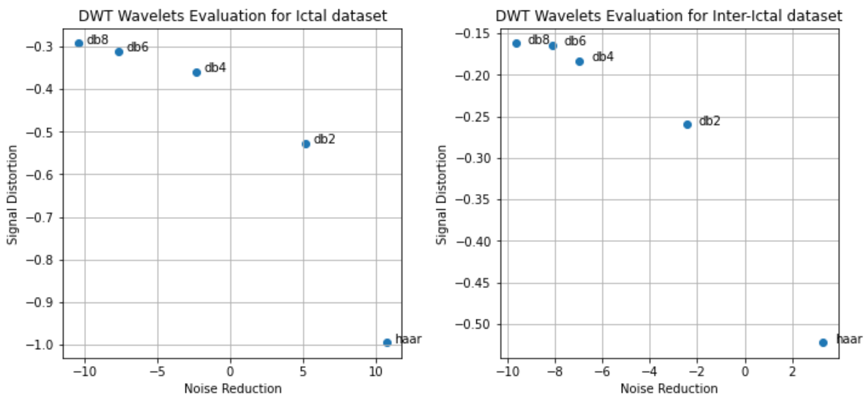


(a) Ictal Signal Filtering (b) Inter-Ictal Signal Filtering

Figure 6: Evaluation of filter order using quantitative metrics (Blackman Window)

3.1.3. Discrete Wavelet Transformation-based Denoising

As an alternative, DWT-based denoising is analyzed. The performance of each wavelet function is quantitatively analyzed by using the same metrics used in Section 3.1.2 which is shown in Figure 7.



(a) Ictal Signal DWT Filtering (b) Inter-Ictal Signal DWT Filtering

Figure 7: Evaluation of different wavelet functions at 6-levels

Based on the evaluation, it is clear that db8 is the best at imposing the lowest signal distortion with the highest noise reduction. Further Figure 8 visualizes the higher resolution bands of an Inter-Ictal EEG signal before and after thresholding for denoising. This provides qualitative evidence that db8 suppresses noisy components sufficiently. However, the signal distortion is still significant compared to FIR-low passing option selected.

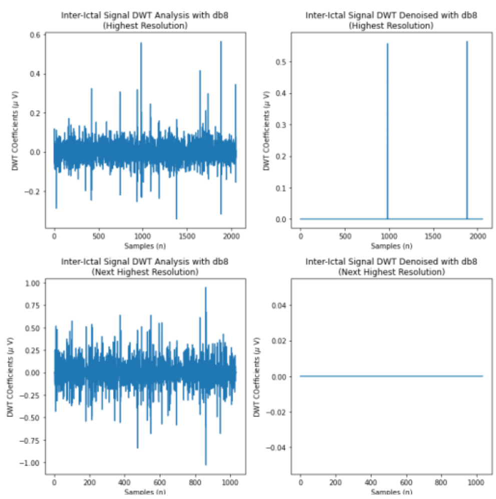


Figure 8: Higher-resolution spectral components before and after filtering

3.2. Feature Extraction and Classification

In the final section of the project, time-frequency features were extracted using wavelet coefficients from the preprocessed EEG data. The quantitative results for the preprocessed EEG signals are present in Table 5 along with the low-passed filtered EEG signals without data sphering. Comparing the results that are obtained for PCA-whitened signals, the wavelet features extracted by the Haar filter using UDWT transform provide the best performance on seizure detection. Further, comparative performance is shown by the features extracted by db6 with DWT transformation where it exceeds Haar UDWT features in terms of specificity. However, the original EEG data without data sphering provides better results with DWT and UDWT features extracted by db6 providing perfect performance. It seems that this dataset is free of artifacts and data sphering does not necessarily improve the seizure detection accuracy.

Table 5: Evaluation of Feature Extraction Configurations

	Feature Extraction Configurations					Evaluation Metrics					
	Transformation		Wavelet Function			Accuracy	Precision	Specificity	Sensitivity	F1	AUC
	DWT	UDWT	db4	db6	Haar						
PCA EEG signals	√		√			0.849	0.868	0.741	0.849	0.847	0.915
		√	√			0.906	0.906	0.926	0.906	0.906	0.974
	√			√		**0.925**	**0.934**	**1.000**	**0.925**	**0.924**	**0.980**
		√		√		0.830	0.831	0.852	0.830	0.830	0.929
	√				√	0.830	0.831	0.852	0.830	0.830	0.879
		√			√	**0.943**	**0.944**	**0.963**	**0.943**	**0.943**	**0.944**
Original EEG signals	√		√			0.943	0.944	0.926	0.943	0.943	0.997
		√	√			0.943	0.944	0.963	0.943	0.943	0.956
	√			√		**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
		√		√		1.000	1.000	1.000	1.000	1.000	1.000
	√				√	0.925	0.925	0.926	0.925	0.925	0.974
		√			√	0.962	0.965	1.000	0.962	0.962	0.986

I analyzed partition B by applying the same techniques selected for partition A to observe whether the techniques are unique or generalizable for any other EEG signals as well. The results are shown in Table 6. The performances for data in partition B data are like the results obtained from partition A. However, choosing the best configuration based on this partition is inconclusive.

Table 6: Evaluation of Techniques on Partition B

	Feature Extraction Configurations					Evaluation Metrics					
	Transformation		Wavelet Function			Accuracy	Precision	Specificity	Sensitivity	F1	AUC
	DWT	UDWT	db4	db6	Haar						
PCA EEG signals	√		√			**0.857**	**0.889**	**0.857**	**0.854**	**1.000**	**1.000**
		√	√			0.714	0.818	0.714	0.689	1.000	0.980
	√			√		0.786	0.792	0.786	0.785	0.857	0.857
		√		√		0.857	0.889	0.857	0.854	1.000	0.898
	√				√	**0.857**	**0.889**	**0.857**	**0.854**	**1.000**	**1.000**
		√			√	**0.857**	**0.889**	**0.857**	**0.854**	**1.000**	**1.000**
Origi	√		√			**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
		√	√			**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

	√			√		**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
		√		√		0.929	0.938	0.929	0.928	0.857	1.000
	√				√	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
		√			√	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

4. Conclusions and Drawbacks

In this project, I analyzed a few signal-processing techniques that are used for data preprocessing and feature extraction in a pipeline to detect epileptic seizures. I used the Epiloptogie dataset which has intracranial EEG signals that are captured at the hippocampus formation and are publicly available. Further, the code implementations for this project can be accessed from here.

Based on the results obtained for data sphering techniques where the EEG signals are decorrelated using PCA and ZCA, it is evident that PCA-based data sphering achieved superior performance for seizure detection. However, this performance improvement was marginal compared to the EEG signals that are only detrended. This may be due to the nature of the dataset I used as the EEG signals in each data directory are segments from continuous multi-channel EEG signals. Thus, certain segments may naturally be correlated, and performing PCA may have caused the correlation information to be lost. Thus, the performance is not as expected from using data sphering as the first step for the EEG-based seizure detection

As the next step, I used denoising which is essential in the seizure detection pipeline. The given data has the spectral energy in higher frequency bands that include γ brain rhythmic patterns which are not useful in detecting epileptic seizures. Using a custom metric defined in Section 2.3, the filtering process was evaluated. Comparing this to the DWT-based denoising method as an alternative, FIR filtering had a significantly better performance. This may again be related to the dataset used as the signals are visually inspected to remove the segments that are subjected to different artifacts. DWT-based denoising is best if artifacts are present from the acquisition or if the frequency bands are overlapping.

After the preprocessing, the features are extracted using wavelet coefficients. Over the other alternatives such as frequency features or temporal features, time-frequency features capture fluctuation details of the domain which is useful in seizure detection. The results show that the original data without any data sphering performs well for the seizure detection task over the data with PCA-whitened EEG signals. As discussed earlier, this may be due to the loss of information from the PCA whitening as certain EEG signals, which are the segments of the same continuous EEG signal acquired, may have a natural correlation that is required to be exploited.

Finally, the comparison was made between partition A and partition B of the EEG data and the results are similar between them using the same analysis techniques. However, this is not the ideal way to conclude that the techniques are performing in general as the data characteristics are the same between the two partitions as they are acquired from the same signals.

5. References

[1] I. Ullah, M. Hussain, E. ul H. Qazi, and H. Aboalsamh, “An automated system for epilepsy detection using EEG brain signals based on deep learning approach,” *Expert Syst Appl*, vol. 107, pp. 61–71, Oct. 2018.

[2] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, vol. 64, no. 6, p. 8, 2001.

[3] T. W. Kjaer, H. B. D. Sorensen, S. Groenborg, C. R. Pedersen, and J. Duun-Henriksen, “Detection of Paroxysms in Long-Term, Single-Channel EEG-Monitoring of Patients with Typical Absence Seizures,” *IEEE J Transl Eng Health Med*, vol. 5, 2017.

[4] J. L. Semmlow and B. Griffel, “BIOSIGNAL and MEDICAL IMAGE PROCESSING Third Edition.”

[5] A. Subasi, J. Kevric, and M. Abdullah Canbaz, “Epileptic seizure detection using hybrid machine learning methods,” *Neural Comput Appl*, vol. 31, no. 1, pp. 317–325, Jan. 2019.