"SeizeSure: An Epilepsy Classifier.

By Ahmed ElAlfy and Thisal Abeysekera

Table of contents

What is Epilepsy?

Epilepsy is a neurological disorder that causes sudden bursts of electrical activity in the brain, also known as seizures, Which can lead to changes in behavior, movements, or consciousness and affect the quality of a person 's life.

- Approximately 2.4 million people are diagnosed with epilepsy each year.
- Epilepsy can occur at any age, but it's most commonly diagnosed in young children and older adults.
- While not curable,It can be managed exceptionally well with the right treatments.
- Early detection and treatment of epilepsy can significantly improve the quality of life of those individuals affected by the condition.

How to identify Epilepsy?

To identify epilepsy, healthcare professionals use various techniques, including:

1. **Medical History:** Gathering information about the individual's symptoms, any witnessed seizure activity, medical history, and family history of epilepsy or related conditions.

2. **Physical Examination:** A thorough examination to check for physical signs or neurological abnormalities associated with epilepsy.

3. **Electroencephalogram (EEG):** This test measures the electrical activity in the brain and can detect abnormal brain waves that may indicate epilepsy.

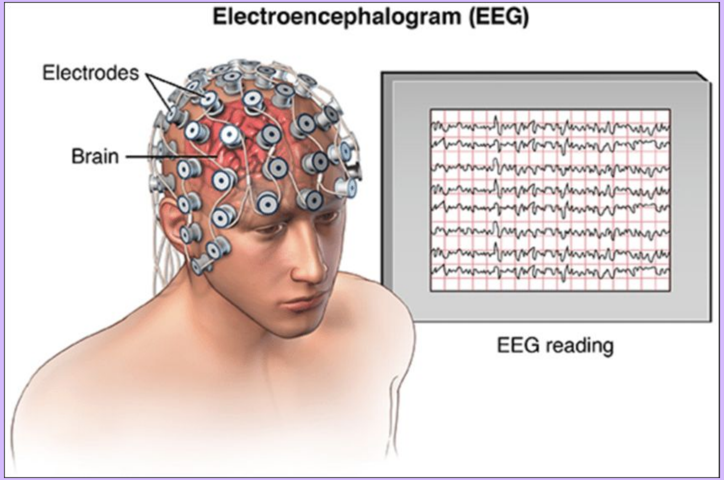
4. **Imaging Studies:** Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scans can help identify structural abnormalities or lesions in the brain that may be causing seizures.

These techniques are often used in combination to provide a comprehensive evaluation and diagnosis of epilepsy.

What is an EEG signal?

An electroencephalogram (EEG) is a test that measures electrical activity in the brain using small, metal discs (electrodes) attached to the scalp. Brain cells communicate via electrical impulses and are active all the time, even during asleep. This activity shows up as wavy lines on an EEG recording.

An EEG is one of the main diagnostic tests for epilepsy. An EEG can also play a role in diagnosing other brain disorders.



Classification of Epilepsy

ILAE (International League Against Epilepsy) Classification

- Class 1: Completely seizure free; no auras
- Class 2 : Only auras; no other seizures
- Class 3 : One to three seizure days per year; +/–auras
- Class 4 : Four to 12 seizure days per year + aura
- Class 5 : Daily seizures +/– auras
- Class 6 : + 100% increase of baseline seizure days +/– auras

Our Dataset

- Our dataset contain 500 Epilepsy patients with 4095 channels of their EEG signal. Which ranged from class 1 to 5 according to their severity of epilepsy level.
- We will focus only on class 3 and class 4 (mild and severe epilepsy)

Our Aim

- Identify mild (class 3) from severe (class 4) epileptic patient based on their EEG signals, Which was archived by transforming the EEG signal to FFT (fast fourier transform) and extract features from it.

Data Normalization

- The StandardScaler is a data preprocessing technique used in machine learning to transform features by scaling them to have a mean of 0 and a standard deviation of 1.

- This normalization process helps in making features comparable and ensures that they are on the same scale, which can improve the performance of certain algorithms and make the model training process more stable.

Principal component analysis

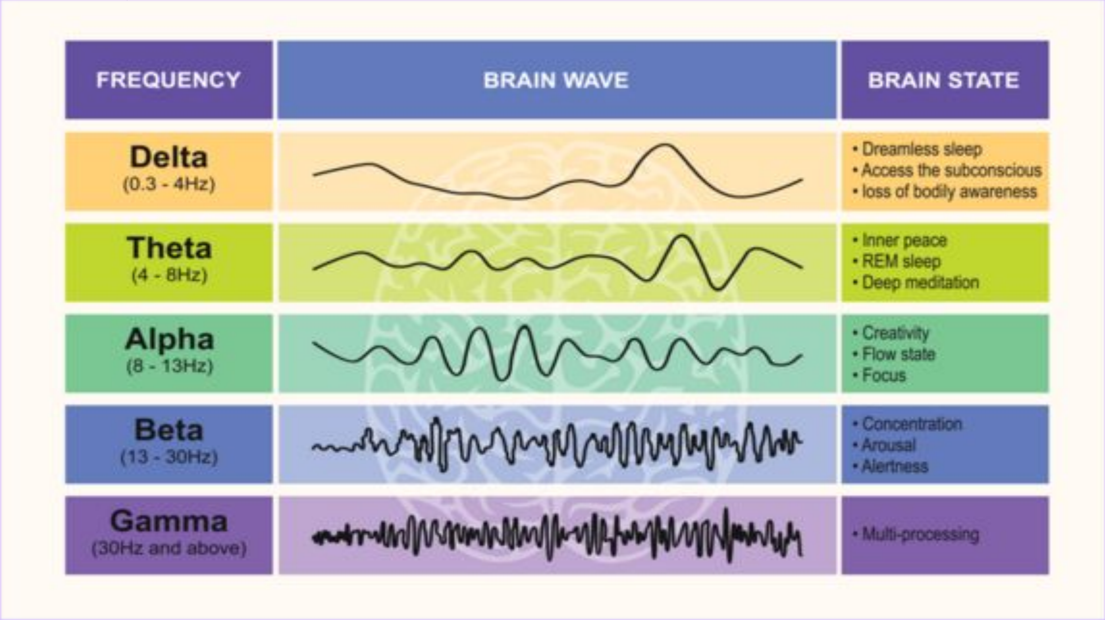
- In outlier detection we used the technique call PCA, In simple words, this technique takes a dataset, simplifies it, and tries to reconstruct the original data.

- It compares the reconstructed one with the original and sees the differences, Setting the difference as a standard to Dimensionality Reduction, Anomaly Detection, Reconstruction Error, and Threshold to identify outliers.

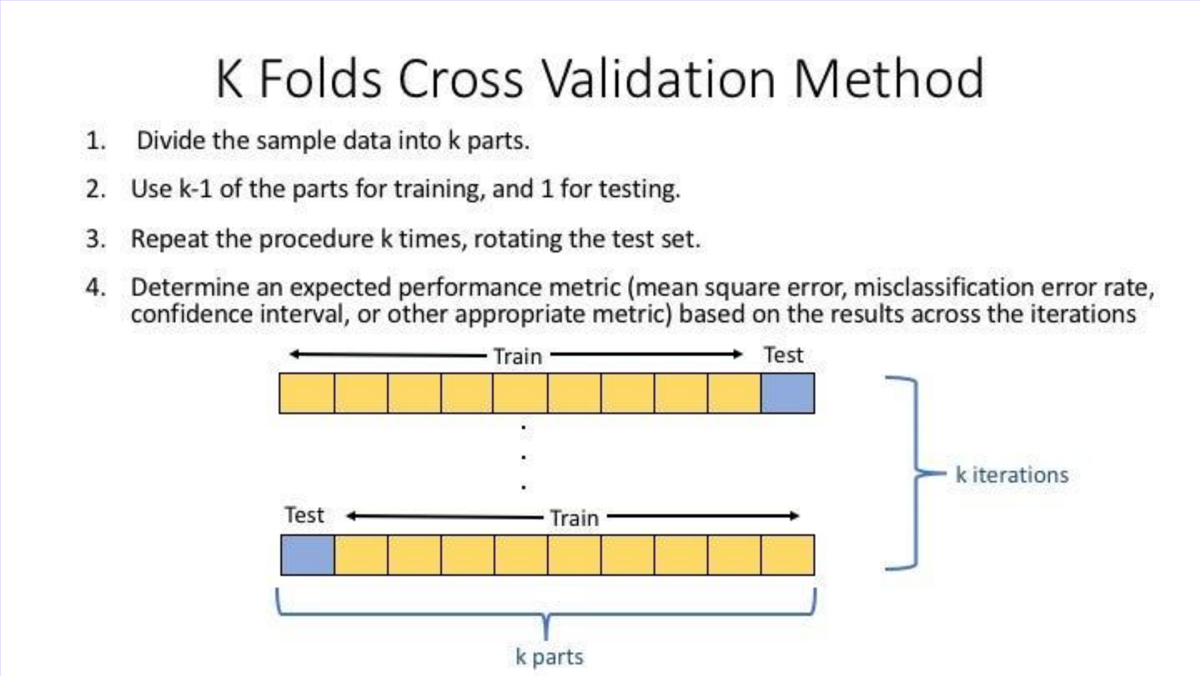
- Using this technique We identified 2 outliers in our dataset.

```
pca = sklearn.decomposition.PCA(n_components=0.9999)

X_pca = pca.fit_transform(X)

X_ori = pca.inverse_transform(X_pca)

anomaly_score = np.abs(X.to_numpy() - X_ori).sum(1)

threshold = np.quantile(anomaly_score, 0.99)

anomalous_ids = np.argwhere(anomaly_score >
threshold).squeeze()
anomalous_ids
```

Frequency bands we focused on

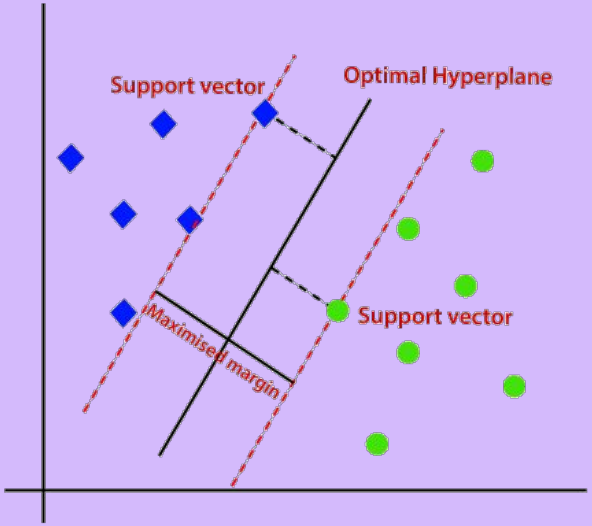


- We will use the psd (power spectral density) to know how is the power is distributed across different frequency band)



- It provides more accurate assessment by average the evaluation metrics across k-folds.
- (In our case we will use k=5)

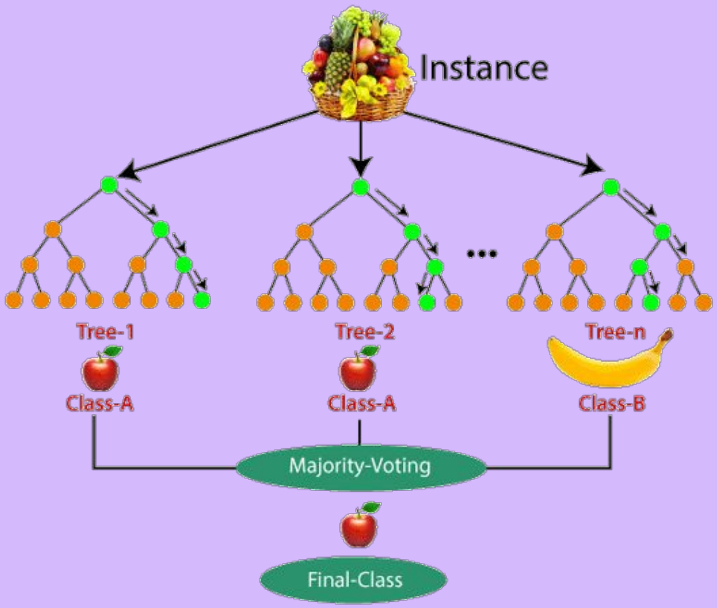
Classifiers We used



SVM (linear)

- A support vector machine (SVM) is a supervised machine learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N–dimensional space.

- The main idea behind SVM is to find the hyperplane that best separates the classes in the feature space. A hyperplane is a decision boundary that separates the data into different classes.



Random Forest (Ensemble)

An Ensemble Random Forest is a machine learning algorithm that combines the predictions of multiple decision trees to make more accurate and robust predictions. Each decision tree in the forest independently learns from a random subset of the training data and features, and then the final prediction is made by averaging or voting on the predictions of all the trees. Random Forests are commonly used for tasks like classification and regression."

Naïve Bayes

$$\underbrace{P(c|x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x|c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c)$$

Naive Bayes is a machine learning algorithm based on Bayes' theorem, used for classification tasks. It assumes that features are independent of each other, which means that the presence of one feature does not affect the presence of another.

specificity and sensitivity.

But why specificity and sensitivity?

– Sensitivity (True Positive Rate): Measures the proportion of the actual positive cases (severe epileptic patients) that are correctly identified by the classifier as positive

– Specificity (True Negative Rate): Measures the proportion of the actual negative cases (mild patients) that are correctly identified by the classifier as negative

– While the accuracy gives us the overall performance of the algorithm, it may not reflect the situations in which different types of errors have different consequences.

Conclusion

	SVM (linear)	Random Forest (Ensemble)	Naïve Bayes
Sensitivity	0.81	1.0	1.0
Specificity	0.89	0.66	0.26
Accuracy	0.66	0.875	0.725

- After comparing the results, we concluded with the Random forest is the best choice with the default hyperparameters given it has the overall highest results.