**Generating signals for use in the dataset:**

This study aims to classify degraded signals from healthy ones. For this, the first task is to create a dataset consisting of multiple signals. The type of signal we are using in this study is an electroencephalogram or EEG. this is an electric signal generated by the brain. EEG signals can easily be generated using Python. They consist of a basic sine wave with random noises superimposed on it and some characteristics that resemble the brain waves. Multiple signals can be generated by changing the amplitude and the noise levels. It is important to note that the sampling frequency and the total time duration for the signal to be generated must remain the same in ensure smooth computation in the next step.

import numpy as np

import matplotlib.pyplot as plt

def generate\_dummy\_eeg(duration=10, sampling\_rate=1000, noise\_level=0.5):

time\_points = np.arange(0, duration, 1/sampling\_rate)

# Generate random noise

noise = np.random.normal(0, noise\_level, len(time\_points))

# Create a sine wave resembling brainwave patterns

theta = 8 # Frequency in Hz

sine\_wave = np.sin(2 \* np.pi \* theta \* time\_points)

# Combine noise and sine wave to simulate EEG signal

eeg\_signal = noise + sine\_wave

return time\_points, eeg\_signal

# Example usage

duration = 10 # seconds

sampling\_rate = 1000 # Hz

noise\_level = 0.5

time\_points, eeg\_signal = generate\_dummy\_eeg(duration, sampling\_rate, noise\_level)

# Plot the generated EEG signal

plt.plot(time\_points, eeg\_signal)

plt.title('Dummy EEG Signal')

plt.xlabel('Time (seconds)')

plt.ylabel('Amplitude')

plt.show()

To generate degraded signals, the same procedure is to be used followed by augmenting the data with Gaussian noise. Varying the magnitude of the mean and standard deviation of the Gaussian noise, multiple degraded signals can be produced. In this study, 8 sets of healthy signals and 8 sets of degraded signals will be used.

**Feature extraction:**

EEG signals are time domain and dynamic. Applying machine learning to raw EEG data can not only take a lot of computational time, but as the literature suggests, it also compromises the classification accuracy. It is important to convert the EEG data into frequency domain using the Welsh method and obtain the Power Spectral Densities (PSDs). This step also ensures increasing the data set as now each density is considered as an input value, thus increasing the input from 8 to 100s or 1000s, depending on the sampling frequency.

**building the model:**

For the task of binary classification, the most used models according to the literature are Support Vector Machines (SVM), Artificial Neural Networks (ANN), Ensemble models, and K nearest neighbor (KNN). The objective is to use the four algorithms to build four models. The dataset of both, degraded and health signals will be divided into:

* 60% training
* 20% testing
* 20% validating

since we are using the PSDs as input, the data can further be divided into parts to be used as input to the models at different stages. For example, the first stage is to select the two best models from the four according to their testing accuracies. Instead of using the entire training data (which is 60% of the original dataset), we only use 20% of it. Similarly, we use 20% of the testing data during the first stage.

In the second and the final stage, we will be using the remaining 80% of both training and testing data for the two models. Since the task at hand is binary classification, the activation function in all models to be used will be the sigmoid function. Sigmoid function is used for binary classification when the output must be a predicted probability. This is why it computes a value between 0 and 1.

trial 1