Classification of EEG Signals

Project Report

SUBMITTED IN PARTIAL FULLFILLMENT
REQUIREMENT FOR THE AWARD OF DEGREE OF

Bachelor of Technology

(Computer Science & Engineering)

SUBMITTED BY:

Parth Kulshreshtha(210C2030159)

UNDER THE SUPERVISION OF

Mr. Arpit Bhardwaj

BML MUNJAL UNIVERSITY™

FROM HERE TO THE WORLD

Department of Computer Science and Engineering
School of Engineering and Technology

November 2022

Candidate’s Declaration

I hereby certify that I have completed the course “Project-1” in five months at School of Engineering and Technology and worked on the project entitled “Classification of EEG Signals” in partial fulfilment of requirements for the award of Degree of Bachelor of Technology in name of the Department of Computer Science and Engineering, having 210C2030159, is an authentic record of my own work carried out during a period from August, 2022 to December, 2022 under the supervision of Mr. Arpit Bhardwaj.

Parth Kulshreshtha (210C2030159)

Supervisor’s Declaration

This is to confirm that, to the best of my knowledge, the candidate's above statement is accurate.

Mr. Arpit Bhardwaj

Abstract

EEG is used to detect and measure brain activity. It aids in determining whether the person has epilepsy or is having a seizure. Data collected by processing EEG signals is generally classified into alpha, beta, gamma, delta, and theta frequency bands. In this project, I'm using a single channel EEG device called MindWave Mobile to analyse data from a single human subject. A full description of how to gather the data is provided, including the use, proper placement of the EEG equipment, and the setting in which the data is recorded. The dataset generated after data collection comprises several superfluous features, which are deleted during the preprocessing stage, and the model is then applied to 11 features.

This project's work revolves around the directions of forward, backward, left, and right. Machine Learning models used for classification on the dataset include LSTM, KNN, and Random Forest.

Acknowledgement

I'd like to express my heartfelt gratitude to my teacher, Mr. Arpit Bhardwaj, for providing us with the opportunity to complete this wonderful project on the topic "Classification of EEG Signals" and present it in front of him and other panel members. The research and report writing were completed with the utmost sincerity and honesty.

I would also like to extend my sincere thanks to Mr. Asif Iqbal for his invaluable guidance, without which this project would have been impossible to complete, as well as some of our seniors whose additional input was extremely beneficial in completing this project.

List of Figures

List of Table

List of Abbreviations

Abbreviation	Full Form
EEG	Electroencephalography
Hz	Hertz
KNN	K Nearest Neighbours
LSTM	Long Short-Term Memory

Table of Contents

Introduction

Millions of cells make up the brain, half of which are neurons. Synapses, which serve as entry points for inhibitory or excitatory activity, are used to connect these neurons in a dense network. Any synaptic activity produces a subtle electrical impulse known as a postsynaptic potential. When thousands of neurons fire in unison, they create an electrical field strong enough to travel through tissue, bone, and skull.

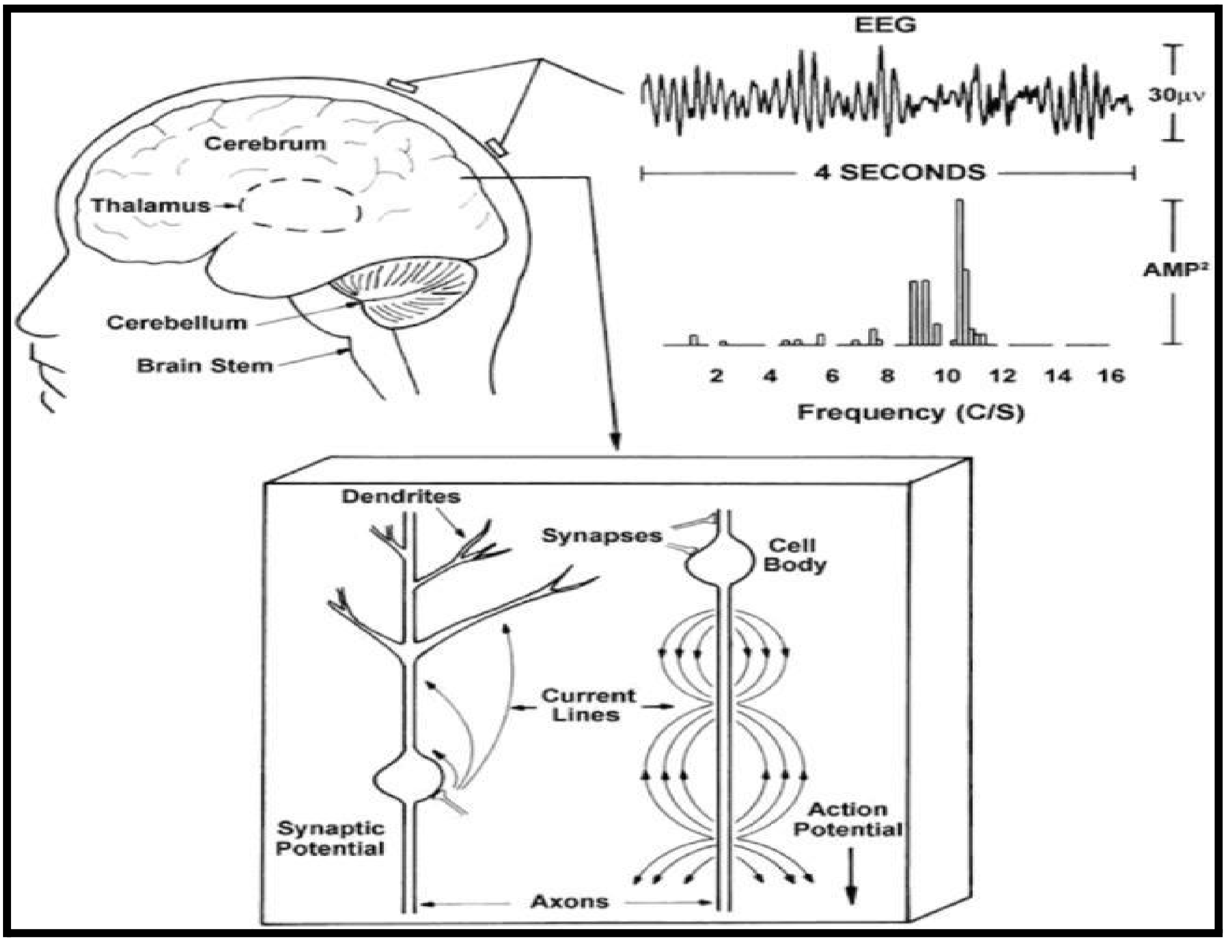

Fig. 1: Capturing EEG [1, p. 5]

It can eventually be measured on the head's surface using EEG. It is a test that uses metal discs called electrodes that are pinned to the scalp to assess electrical activity (in volts) in the brain. These charges are then amplified, and they can be shown as a graph on a computer screen or as a recording that can be printed out on paper.

One of the primary diagnostic procedures for epilepsy is an EEG. Other brain disorders can also be identified using an EEG. Although we are able to get some information about the brain and it’s complex structure in the form of EEG signals, research is still going on to extract more information from the EEG signals and use it in multidisciplinary fields like Neuromarketing, Psychology and Neuroscience, Brain Computer Interface(BCI).

Literature Review

Kumar, JS., & Bhuvaneswari, P. (2012). Analysis of Electroencephalography (EEG) Signals and Its Categorization - A Study

This paper has provided insight into the structure of the brain in different motor and sensory states such as eye movement, lip movement, remembrance, attention, hand clenching, and so on.

J. Kumar and P. Bhuvaneswari have described the EEG data collection process in great detail. They have described the machine used to collect data as well as the electrodes; their placement is also important for data collection.

They have also mentioned about the various steps involved in pre-processing the data such as data filtering, feature extraction, classification. They have also described that how the classification is done into Delta, Theta, Alpha, Beta and Gama frequencies ranging from 0.1Hz to 100Hz.

Sepp Hochreiter, P. (1997). Long Short-term Memory

Through this paper a lot of information and insight was gathered in the working of the LSTM model, which is a type of an RNN model. Hochreiter mentioned the need of LSTM model as a solution to the vanishing gradient problem. In RNN model the output depends on the input and the output of the previous state but sometimes for making the correct prediction, we need the output of the state which not in the recent past.

To solve the problem for the correct predictions Hochreiter mentioned about the three gates – input, output and a forget gate. With the help of these gate the information was controlled and stored to correctly predict the results.

Leo Breiman, P. (1996). Bagging Predictors

Breiman introduced the Bagging method in his paper. It's an Ensemble learning method made up of set of set of classifiers. In his paper he explained how a data point can be taken more than once in a model and then how these models are trained independently. Using this method resulted in substantial gain in the accuracy of both the Classification and Regression Models.

Material in this paper helped in understanding the Random Forest Model as it uses both the Bagging Method and Feature Randomness to create a forest of Decision Trees.

Tong-Yuen Chai, P. (2010). Classification of Human Emotions from EEG Signals using Statistical Features and Neural Network

This paper explained about the application of EEG in classification of Human Emotion. Chai used 6 test subjects to gather the data of human emotions on different stimuli. After classification they got five main features which were Anger, Sad, Surprise, Happy and Neutral.

Technique used in acquiring the data was a 64-channel biosensor. Chai explained in details how this sensor work i.e. how 62 electrodes are there for the EEG and the remaining 2 are for EOG i.e. for detection of eye blink and eye movement.

Chai used the BPTT neural network, it’s a type of a RNN. He explained how he took different combination of emotions like first he took only 2 emotions which resulted in the highest percent of accuracy of 97.5% which was followed by 97.2% accuracy by taking the combination of 3 emotions.

Methodology

Data was collected with the help of a device named “Mind Wave Mobile” made by the NeuroSky company. It’s an EEG headset which is equipped with the biosensor technology that makes it easy for us to gather the EEG data from our brain. A very useful feature of this device is that it gathers data and then do the basic pre-processing after which we can fetch the data in a .csv format.

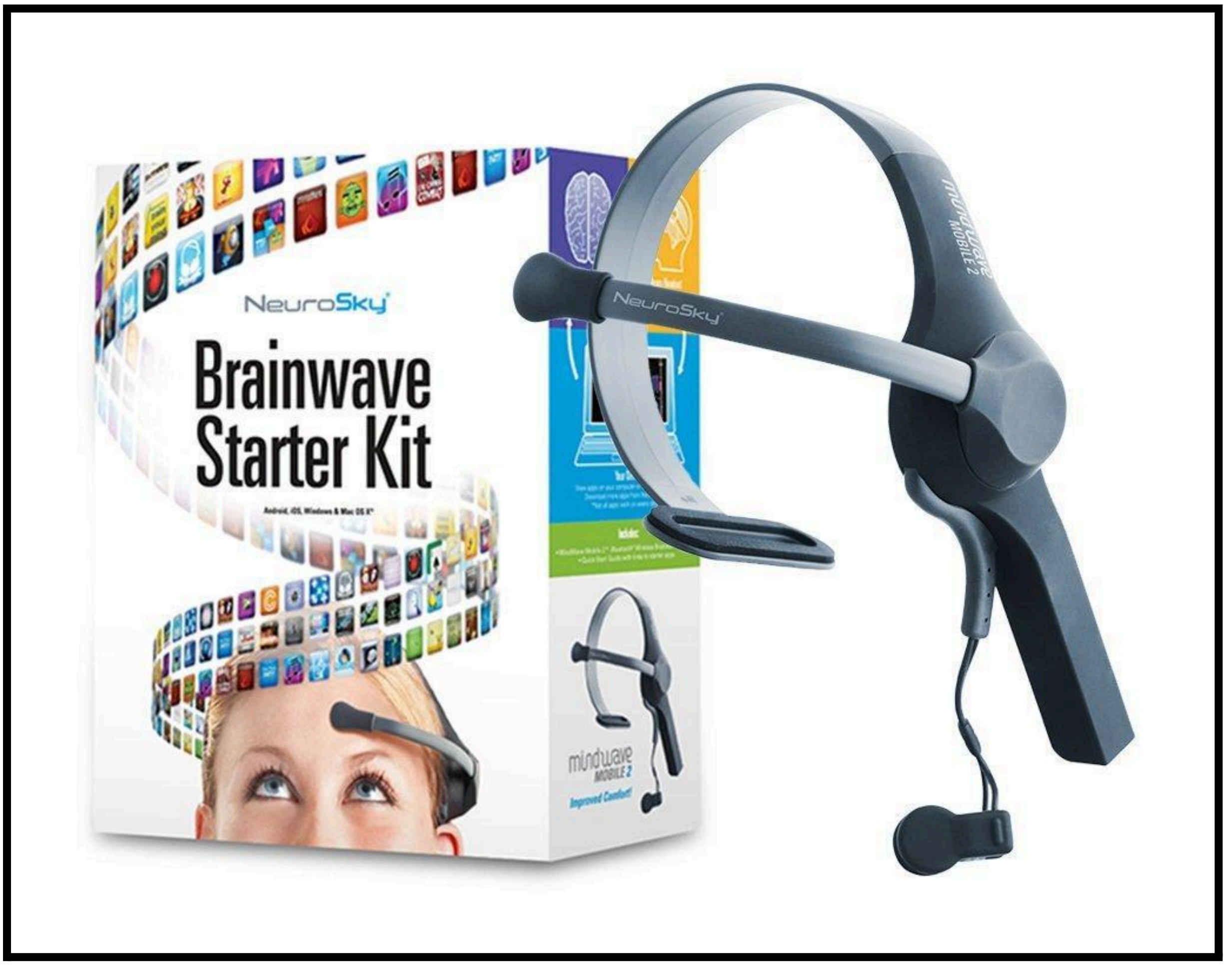

Fig. 2: MindWave Mobile [2]

Process of data collection is as follows:

1. Data is collected from a single human subject.
2. Subject is asked to sit in a room with minimal noise and disruption so that we can maintain data quality.
3. MindWave Mobile headset is mounted on the subject and correct placement of the headset is also necessary.
4. The ground of the headset is placed on the ear clip and the EEG sensor is placed on the end of the arm which is placed on the forehead.

Fig. 3: Subject during data collection

5. The headset connects to the phone via Bluetooth, and some parameters must be set by the user in order for the data to be recorded.
6. At the time of data collection, the subjects consider one of four directions: forward, backward, left, or right.
7. The subject is asked to sit for an hour and think about the four mentioned directions one at a time and data is stored in a .csv file.
8. The data is recorded for one minute at one-second intervals, yielding 60 rows of data in a single session.

9. This yields an average of 20 csv files, 5 in each direction, for a total of 1200 rows of data.

Once the process of data collection is completed then different Machine Learning and Neural Network Model are used for Classification

- **LSTM**

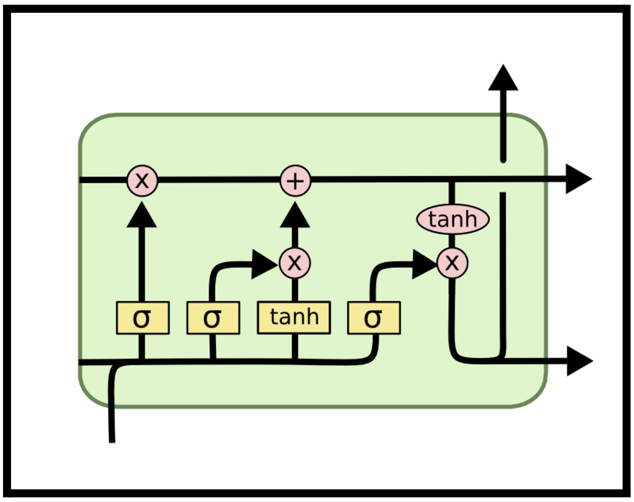

Fig. 4: Gated LSTM Cell

The data is first processed, basic algorithms for standardisation, outlier detection, etc. results in important features which are fed to the model. Then different layers are added to train the model and improve its accuracy.

- **KNN**

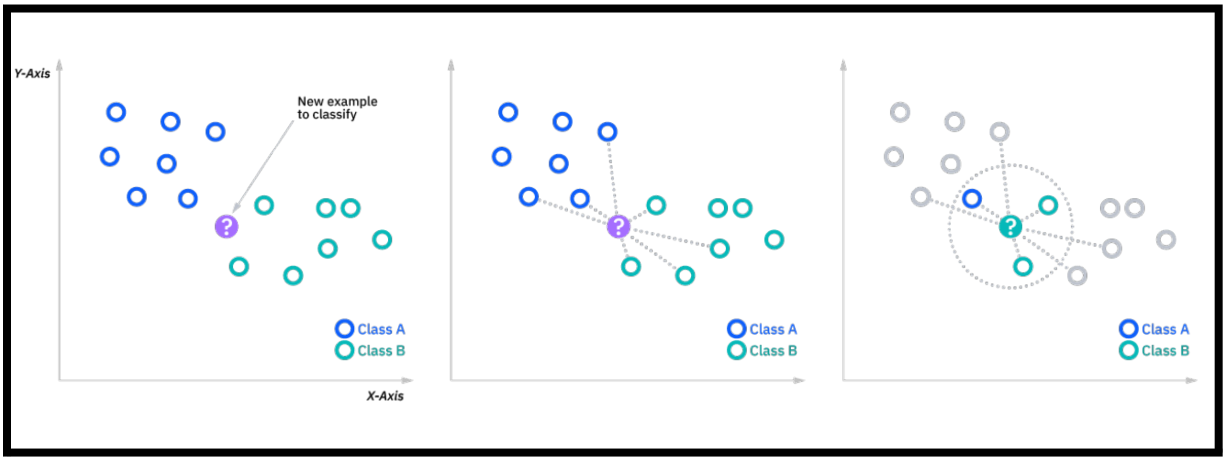

Fig. 5: KNN Diagram

The k-nearest neighbours algorithm, or KNN for short, is a supervised learning classifier that employs proximity to produce classifications or predictions about how a particular data point will be grouped. Although it can be applied to classification or regression issues, it is commonly employed as a classification algorithm because it relies on the idea that comparable points can be discovered close to one another. A class label is chosen for classification problems based on a majority vote, meaning that the label that is most frequently represented around a particular data point is used.

- **Random Forest**

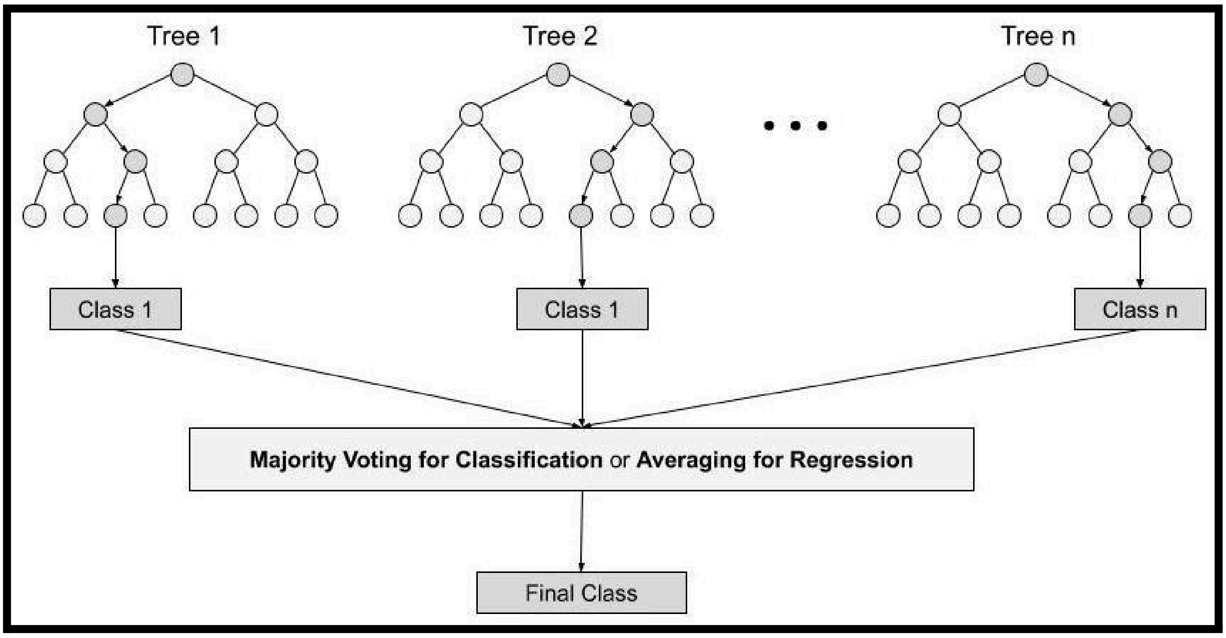

Fig. 6: Random Forest Diagram

A huge number of distinct decision trees work together as an ensemble in a random forest. The term "ensemble" refers to a class of methods that combine several learning algorithms to provide predictions that are more accurate than those produced by any one of the individual learning algorithms used in the ensemble. Every single tree in the random forest spits out a class forecast, and the classification that receives the most votes becomes the prediction made by our model.

Data Description

Over the course of four months, data was collected from a single subject in the directions forward, backward, left, and right. The data set contained 15,459 rows in total. We are now working with 11 columns in this project after removing unnecessary columns.

This is how the data set appears:

	attention	meditation	delta	theta	alphaLow	alphaHigh	betaLow	betaHigh	gammaLow	gammaMid	direction
0	50	63	47335	16755481	26686	6106	7214	2678	4946	1295	Forward
1	41	53	807962	381706	24030	11299	17020	22975	8633	2995	Forward
2	53	54	74205	31596	23686	4736	5837	11982	2429	3667	Forward
3	48	38	1154845	62690	10877	13968	9558	10730	6538	3292	Forward
4	66	29	4595	20664	5238	3855	4256	7509	14710	3759	Forward

Fig. 7: Data Set

Feature	**Description**
attention	Attention level of brain at the time of recording
meditation	Meditation level of brain at the time of recording
delta	(0.5 - 2.75Hz) frequency band
theta	(3.5 - 6.75Hz) frequency band
alphaLow	(7.5 - 9.25Hz) frequency band
alphaHigh	(10 - 11.75Hz) frequency band
betaLow	(13 - 16.75Hz) frequency band
betaHigh	(18 - 29.75Hz) frequency band

gammaLow	(31 - 39.75Hz) frequency band
gammaMid	(41 - 49.75Hz) frequency band
direction	One of the directions being thought at the time

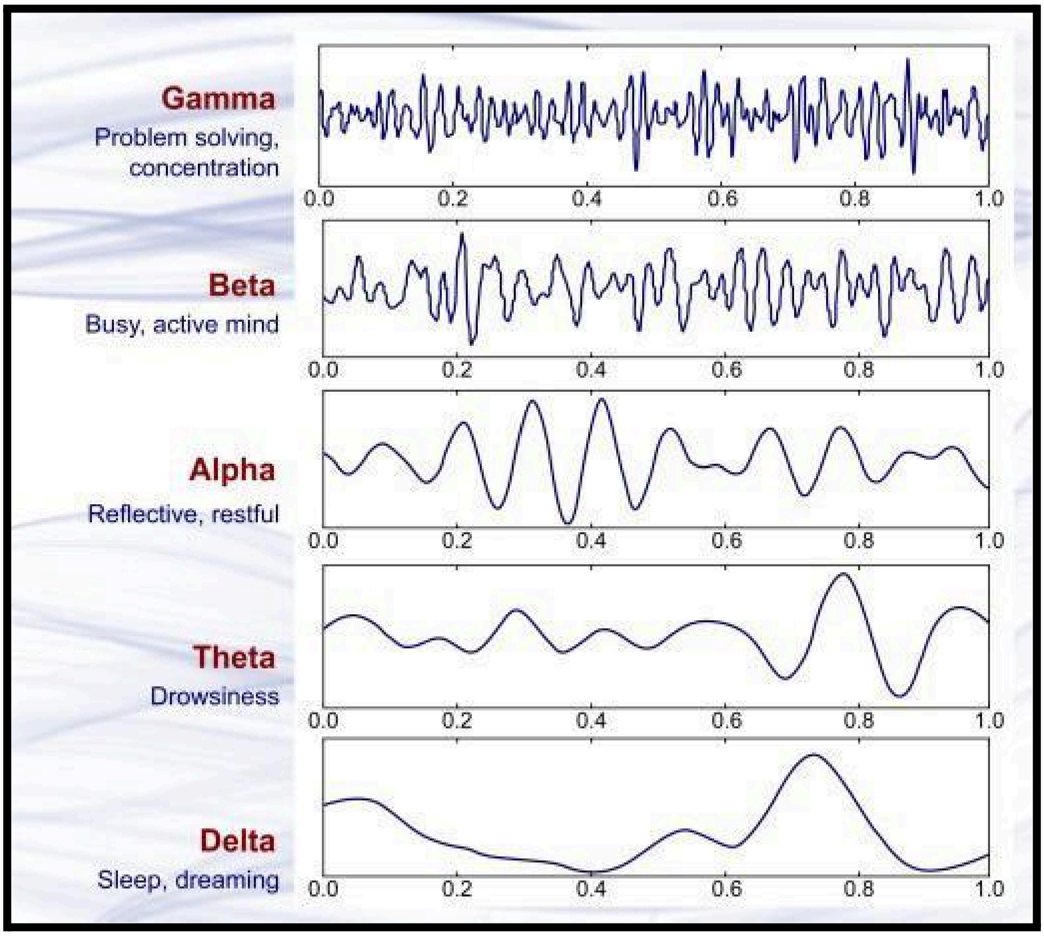

Fig. 8: Human Brain Waves

EEG Frequency Ranges:

- **Delta:** The depth of sleep is assessed by measuring delta waves. The delta rhythm becomes stronger as sleep depth increases.

- **Theta:** Theta is associated with a number of cognitive functions, such as memory encoding and retrieval as well as cognitive workload.
- **Alpha:** The alpha waves take over whenever we close our eyes and focus on becoming calm.
- **Beta:** Beta frequencies are stronger over the motor areas while we prepare or carry out motions of any body part.
- **Gamma:** Gamma serves as a carrier frequency to facilitate data transfer between brain regions and reflects attentive focusing.

Result and Discussion

We used three models in the project: LSTM, Random Forest, and KNN. The accuracy of prediction that we have obtained is:

Model	Accuracy
KNN	27.49%
LSTM	26.04%
Random Forest	2.04%

The current project has a lot of room for advancement and improvement. First and foremost, we can expand the dataset by including more subjects. We can also improve the models by using various techniques such as finding the hyperparameter and tuning it with cross validation and grid searches.

The project's next stage would be to collect data using multichannel EEG devices and work with the data by writing the words instead of thinking them and also taking visual stimuli into account.

Reference

1. Nunez PL, Srinivasan R. *Electric fields of the brain: the neurophysics of EEG.* Oxford: Oxford university Press; 2006.
2. NeuroSky. 2015 MindWave. Retrieved from https://store.neurosky.com/pages/mindwave
3. Fulvio, TD. (2019). Using Machine Learning to Categorise EEG Signals From The Brain to Words. Retrieved from URL: https://towardsdatascience.com/using-machine-learning-to-categorise-eeg-signals-from-the-brain-to-words-728aba93b2b3
4. Craik, A., He, Y., & Vidal, J. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. Retrieved from: iopscience.iop.org
5. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeg
6. https://imotions.com/blog/what-is-eeg/
7. https://www.ibm.com/topics/knn
8. https://hk9663.medium.com/working-with-1-channel-eeg-data-from-neurosky-mindwave-mobile-2-a3d28a3244fd
9. https://www.sciencedirect.com/science/article/pii/S1877050918308482
10. https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74