Medical Instrumentation

Signal Acquisition and BCI report

Emotions Classifier Team 4

Name	Section	B.N
Mohamed Elsayed Ali	**2**	**10**
Mohamed Elsayed Eid	**2**	**11**
Mariam Ahmed	**2**	**27**
Mahmoud Magdy	**2**	**23**
Mohamed Sayed Mosilhy	**2**	**16**

Introduction

The Muse EEG headband is a wearable device designed to monitor brain activity and provide real-time feedback to users. EEG stands for Electroencephalography, a method of monitoring electrical activity in the brain. Muse uses EEG technology to measure brainwave patterns, which can provide insights into a person's mental state, such as their level of relaxation or focus. Due to the complexity, randomness, and non-stationary aspects of brainwave data, classification is challenging with a raw EEG stream. So we use a script that will take EEG brainwaves and create a **statistical extraction** dataset through a sliding window approach.

Our Dataset

Data was obtained from two participants (one male and one female) for three minutes per state (positive, neutral, and negative). The TP9, AF7, AF8, and TP10 EEG locations were recorded using a Muse EEG headband and dry electrodes. Six minutes of resting neutral data are also collected.

Feature Extraction

Due to the temporal, auto-correlated nature of the EEG waves, single-point features cannot generally provide enough information for good rules to be generated by machine learning models. In this work, we follow the approach of extracting statistical features based on sliding time windows. More specifically, the EEG signal is divided into a sequence of windows of length one second, with consecutive windows overlapping by 0.5 sec, e.g., [(0−1 sec), (0.5−1.5 sec), (1−2 sec),. . .].

Assume that each 1-second time window contains a sequence $x = [x_1, \ldots, x_N]$ composed of N samples. Also, let $\mathbf{x}_{h1}$ and $\mathbf{x}_{h2}$ denote the first and second halves of the window, and $\mathbf{x}_{q1}$, $\mathbf{x}_{q2}$, $\mathbf{x}_{q3}$, and $\mathbf{x}_{q4}$ denote the four quarter-windows obtained by dividing the window into four (roughly) equal-sized parts, each composed of approximately N/4 samples.

In this work, the following statistical features were generated for each time window:

1. The sample mean and sample standard deviation of each signal.
2. The sample skewness and sample kurtosis of each signal.
3. The maximum and minimum value of each signal.
4. The sample variances of each signal, plus the sample covariances of all signal pairs.
5. The eigenvalues of the covariance matrix.
6. The upper triangular elements of the matrix logarithm of the covariance matrix.
7. The magnitude of the frequency components of each signal, obtained using a Fast Fourier Transform (FFT).
8. The frequency values of the ten most energetic components of the FFT.

Regarding the representation of the signals in the frequency domain using FFT, two specific aspects were taken into account: first, the DC component of the signals was filtered out prior to the application of the FFT, so the zero frequency of the component was always set as zero. This was done to prevent the offset from completely dominating the power spectrum, even though it carries no relevant information for the classification task. The second aspect is that frequencies in the range of (50 ± 1) Hz were also filtered out, to remove any contamination from the AC electrical distribution frequency, which could also skew the power spectrum of our signals. Each window receives as features the vector of quantities computed above for both itself and the window that immediately precedes it (1-lag window). Features from the 1-lag window that were clearly redundant due to the halfwindow overlaps were removed prior to the composition of the feature vector, namely the sample means, maximum and minimum values of xq3 and xq4, as well as their respective differences.

Choice of electrodes:

Electrodes selection was based on various reasons that all correlate to the analysis of eeg signals that are specific to emotions, so TP9, AF7, AF8, and TP10, were chosen for recording EEG data is not arbitrary; it is based on their specific advantages and relevance for various research and clinical purposes.

- Standardized EEG Placements: The electrode placements TP9, AF7, AF8, and TP10 follow the international 10-20 system, which is a standardized system for electrode placement in EEG studies. This system ensures consistency and comparability across different EEG studies.
- Coverage of Key Brain Regions: These electrode placements cover important regions of the brain. For example:
    - Frontal Lobe (AF7 and AF8): These electrodes are situated over the frontal lobe, which is associated with higher cognitive functions, personality, and emotions. Frontal lobe activity is crucial in understanding emotional processing and cognitive functions.
    - Temporal Lobe (TP9 and TP10): These electrodes are placed over the temporal lobes, which are involved in auditory processing, memory, and

emotion regulation. Temporal lobe activity is important for capturing emotional responses, especially in auditory or emotionally charged stimuli.

- Bilateral Symmetry: The choice of electrode pairs (e.g., TP9 and TP10) on opposite sides of the head allows for capturing bilateral brain activity, which is essential in understanding how emotions and cognitive processes are represented and regulated in both hemispheres.

Statistical Analysis Report of EEG Data

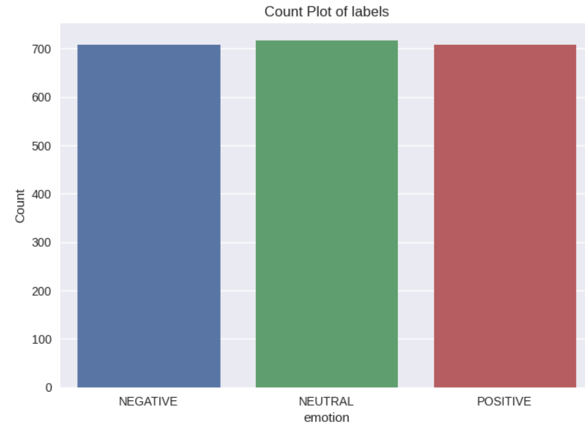
```
---------------------------------------------------

description:        # mean_0_a     mean_1_a     mean_2_a     mean_3_a     mean_4_a  \
count  2132.000000  2132.000000  2132.000000  2132.000000  2132.000000
mean     15.256914    27.012462  -104.975629    13.605898    24.150483
std      15.284621     9.265141   206.271960    16.874676    14.187340
min     -61.300000  -114.000000  -970.000000  -137.000000  -217.000000
25%       6.577500    26.075000  -195.000000     4.857500    23.600000
50%      14.100000    30.000000    14.950000    15.400000    25.200000
75%      27.700000    31.400000    29.600000    26.500000    26.800000
max     304.000000    42.300000   661.000000   206.000000   213.000000

        mean_d_0_a   mean_d_1_a   mean_d_2_a   mean_d_3_a   mean_d_4_a  ...  \
count  2132.000000  2132.000000  2132.000000  2132.000000  2132.000000  ...
mean      0.025378     0.052282     0.301655     0.036793     0.083567  ...
std      17.981796     8.509174    68.098894    17.010031    18.935378  ...
min    -218.000000  -255.000000 -1360.000000  -203.000000  -553.000000  ...
25%      -3.105000    -1.340000    -4.002500    -2.905000    -2.622500  ...
50%      -0.044600     0.132000     0.957500    -0.099750     0.146500  ...
75%       2.920000     1.540000     6.735000     2.535000     2.870000  ...
max     402.000000   257.000000  1150.000000   349.000000   444.000000  ...

         fft_740_b    fft_741_b    fft_742_b    fft_743_b    fft_744_b  \
count  2132.000000  2132.000000  2132.000000  2132.000000  2132.000000
mean    -22.938971   104.946111   -51.973647   -51.973647   104.946111
std     298.034311   212.532721   112.160233   112.160233   212.532721
min   -1180.000000  -921.000000  -504.000000  -504.000000  -921.000000
25%    -106.500000    -8.365000   -92.900000   -92.900000    -8.365000
50%      83.850000    12.150000   -21.800000   -21.800000    12.150000
75%     154.000000   177.000000    12.025000    12.025000   177.000000
max    1070.000000   843.000000  1490.000000  1490.000000   843.000000

         fft_745_b    fft_746_b    fft_747_b    fft_748_b    fft_749_b
count  2132.000000  2132.000000  2132.000000  2132.000000  2132.000000
mean     -6.934144    95.104886   -49.061255   -49.061255    95.104886
std     281.040552   203.194976   106.486317   106.486317   203.194976
min   -1160.000000 -1010.000000  -521.000000  -521.000000 -1010.000000
25%    -102.500000    -8.837500   -87.150000   -87.150000    -8.837500
50%      89.700000    13.400000   -24.100000   -24.100000    13.400000
75%     153.000000   149.250000    10.925000    10.925000   149.250000
max    1180.000000   888.000000  1670.000000  1670.000000   888.000000

[8 rows x 2548 columns]
```

RangeIndex: 2132 entries, 0 to 2131

Columns: 2549 entries, # mean_0_a to label

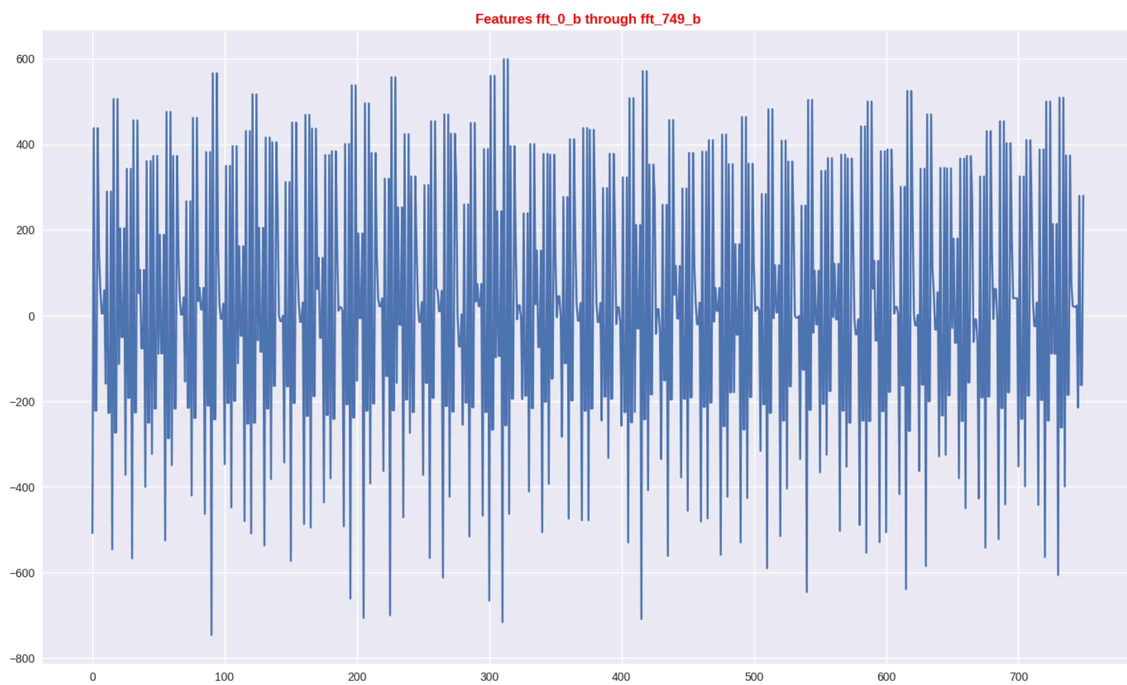
Data Exploration and preprocessing

Since the data is already preprocessed and undergone fast fourier transform, applying a bandpass filter would cause huge loss in the data that's already clean and not noisy, there are no duplicates nor there are null values, so our preprocessing step is to split the signal into categories, Positive, negative and neutral, it's worth mentioning that each of the emotions correspond to certain frequencies for example

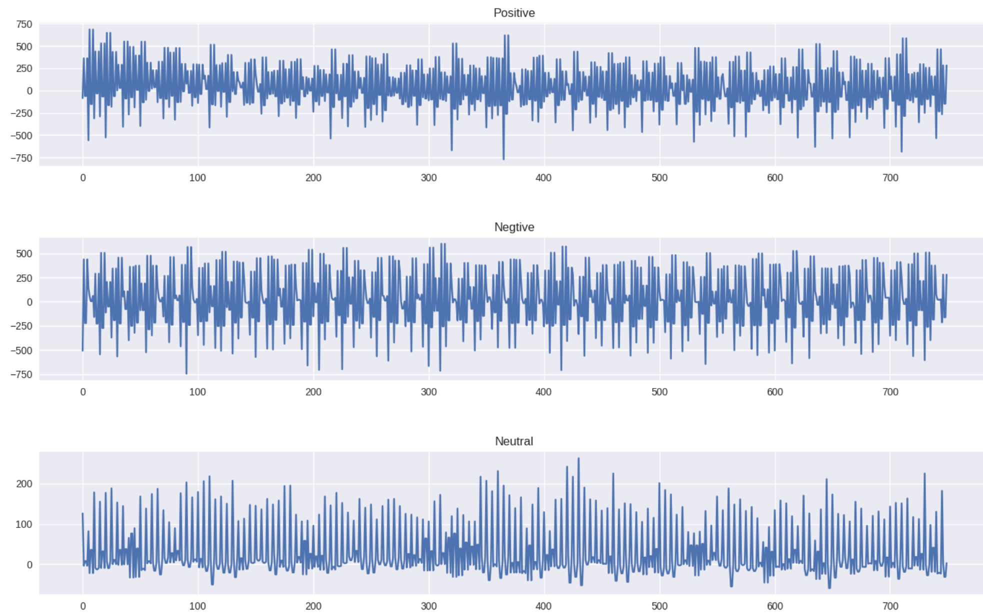


- positive emotions correspond to frequencies between 8-12 Hz (alpha band)
- negative emotions are in the 4-8 Hz (theta band)
- neutral emotions in the 12-30 Hz (beta and gamma bands).

Before Preprocessing



After Preprocessing



MLP Classifier

- The Multilayer Perceptron (MLP) classifier is a widely used neural network architecture for classification tasks.
- MLP classifiers are valuable for modeling complex non-linear relationships in EEG data, which is essential for capturing intricate emotional response patterns.
- EEG signals represent brain activity and often display intricate patterns that are challenging to analyze using traditional linear methods.
- MLP classifiers are advantageous in this context because they consist of interconnected layers of artificial neurons, enabling them to learn and represent complex features and relationships in EEG data.

- MLP classifiers are adaptable and capable of handling high-dimensional EEG data, making them well-suited for classifying emotional states.
- By training on labeled EEG datasets, MLP classifiers can effectively distinguish different emotional responses, providing insights into the neural basis of emotions.

```
Cross-Validation Scores: [0.95550351 0.97423888 0.98122066 0.97887324 0.96478873]
Mean Accuracy: 0.9709250035733528
Standard Deviation: 0.009546024170787648
```

Random Forest Classifier

Why use Random Forest for classifying EEG signals:

- **Ensemble Learning:** Random Forest is an ensemble method that combines multiple decision trees, which are capable of capturing different aspects of EEG data. This ensemble approach improves the robustness and accuracy of classification.
- **Feature Importance:** Random Forest can rank the importance of different EEG features, helping researchers identify the key variables that contribute to the classification of emotional responses or other cognitive states.
- **Robustness to Noise:** EEG signals can be noisy due to various factors like electrode placement and movement artifacts. Random Forest's ability to work with noisy data makes it a reliable choice for real-world EEG applications.
- **High-Dimensional Data:** EEG data is typically high-dimensional, and Random Forest is capable of handling this complexity, making it suitable for capturing intricate patterns within the data.
- **Reduced Risk of Overfitting:** Random Forest is less prone to overfitting compared to individual decision trees, which can be advantageous when dealing with limited EEG datasets.