PROJECT INFORMATION			
Report Description:	Labeling		
Professor:	Prof. Gady Agam	**Tools used/work done:**	1. Labeling 2. Initial Feature Extraction and training
Report prepared by:	Noviya Balasubramanian		
HAWK ID:	A20541236		
Report no:	8	**Report Date:**	10/11/2024

Timeline:

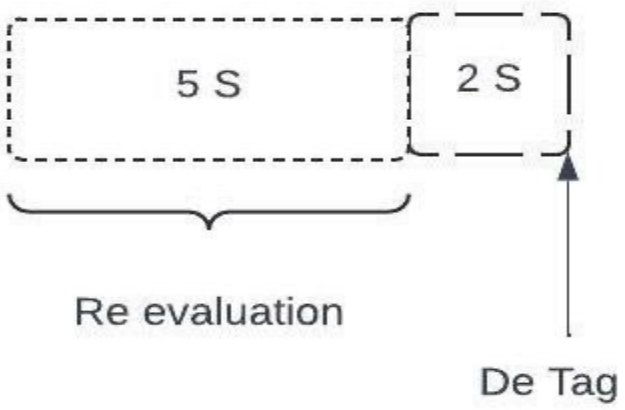
1. **First 6 Weeks: Literature Review, Data Access, Preprocessing, Problem Statement Definition**
2. **Week 7 (Oct 4): Data Preprocessing Completion, MARA Exploration in MATLAB - Completed for 33 subjects**
3. **Week 8 (Oct 11): Labeling, Feature Extraction and Classification - Initial Training**
4. Week 9 (Oct 18): Classifier Selection and Initial Training
5. Week 10 (Oct 25): Classifier Optimization and Validation
6. Week 11 (Nov 1): Multimodal Analysis
7. Week 12 (Nov 8): Fusion or Comparison Analysis Scope
8. Week 13 (Nov 15): Final Testing
9. Week 14 (Nov 22): Model Evaluation
10. Week 15 (Nov 29): Report Preparation (Buffer)
11. Week 16 (Dec 6): Report Submission

Topic: Classification of Cognitive States Using EEG and Physiological Signals: Impasse, Aha!, Uncertainty

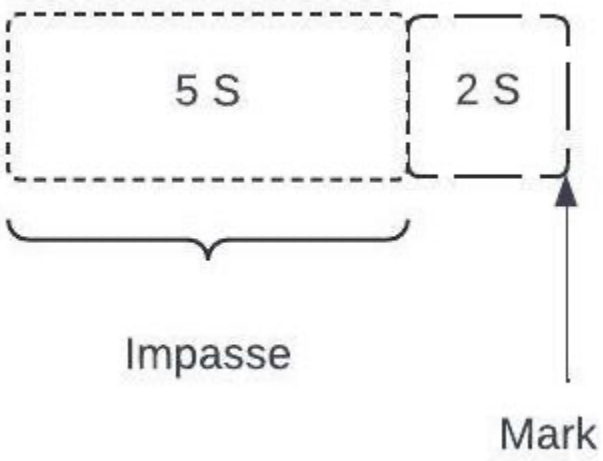
As planned:
Labeling, Feature Extraction and Classification - Initial Training

Methodology for labeling the data
Overall time taken = 5-6 hours

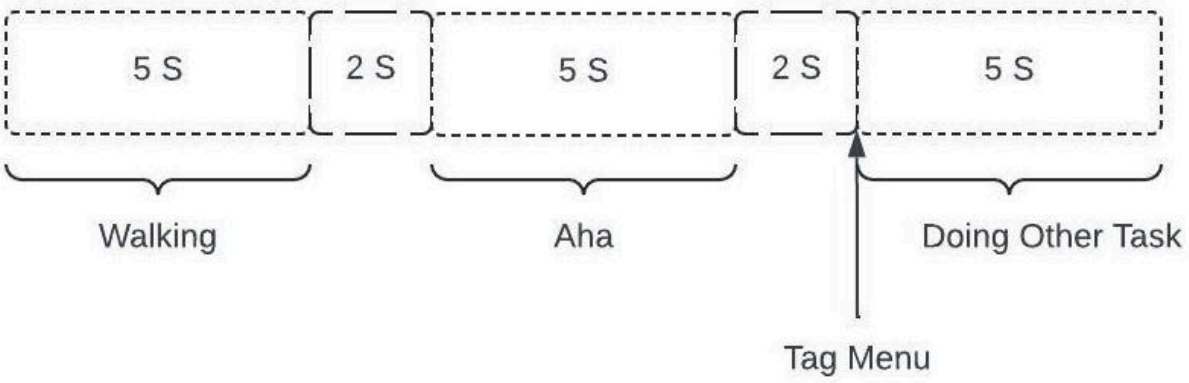
1. Data Setup: Accessed Cleaned data stored in Google Drive using colab
2. Labeling Process:
    a. Use a time frame of 5 seconds to label each cognitive state.
    b. Link specific events (e.g., tagging and untagging) to predefined labels:
        i. “Re-evaluation” for detagging events.



        ii. “Impasse” for button press events.



iii. Aha": 7 to 2 seconds before the "TagHandMenuPumpTime" event.
iv. "Walking": 14 to 9 seconds before the event.
v. "Doing Other Task": From the event time to 5 seconds after the event.



c. Each state is labeled for a duration of 5 seconds surrounding the respective event
d. The labeling starts 7 seconds before the event time and ends 2 seconds before the event time. This creates a labeled interval of 5 seconds (from -7 seconds to -2 seconds) to account for the cognitive processes leading up to the event.

3. Assumptions:
    a. Indices of the labeled data are rounded to the nearest multiple of 8 to ensure alignment with the structure of the recorded data.
    b. Any remaining "Unknown" labels are without specific cognitive events.

4. Data Processing for Each Subject:
    a. The start index is determined by calculating the closest index to the event time (converted to milliseconds) and rounding to the nearest multiple of 8, adjusting it to start 7 seconds prior to the event. The end index is also adjusted to end 2 seconds after the event time.

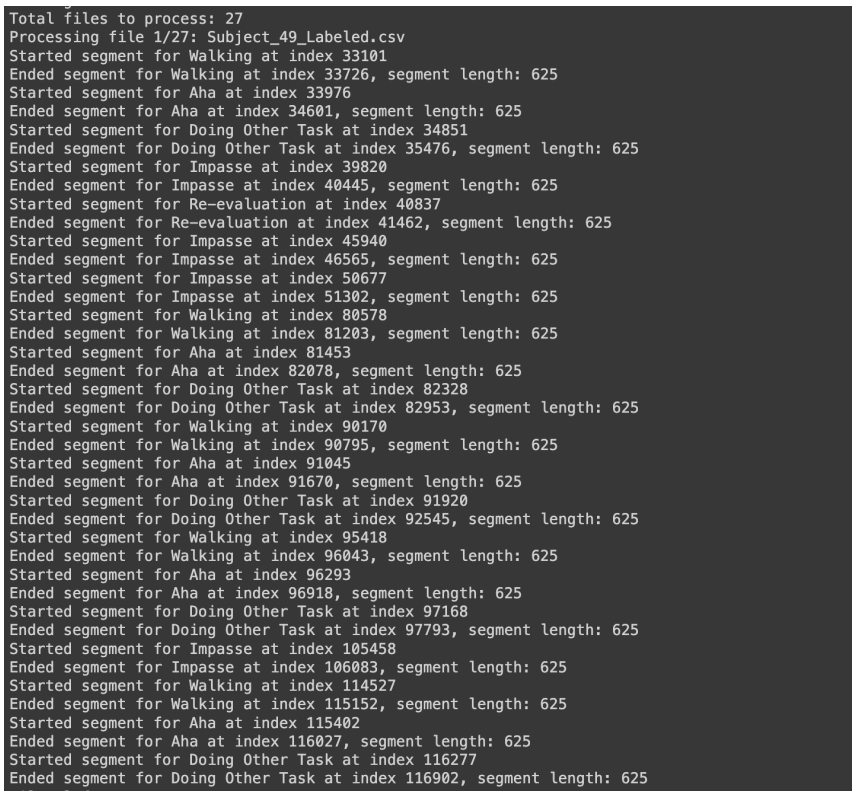
Each cognitive state is categorized based on a specific 5-second window to ensure the temporal relevance of the data.

To extract segments of EEG signals based on the labeled cognitive states.

- iterated over each row of the DataFrame, checking the ‘Labels’ column for specific labels such as ‘Aha’, ‘Re-evaluation’, ‘Impasse’, ‘Doing Other Task’, and ‘Walking’. When it encounters one of these labels and no segment is currently active, it starts a new segment and stores the EEG channel data (columns 1 to 17) in a list.
- If it encounters the label ‘Unknown’, which signifies the end of the current segment, it appends the accumulated segment to the segments list and stores the corresponding label in the labels list.
- The function continues this process, appending new data to the current segment if the label matches, and returns the segments and their associated labels at the end.

The segments represent continuous blocks of EEG data for specific cognitive states, where:

- Each segment has a size of 625 data points (rows) for 16 EEG channels (columns).
- The 625 data points correspond to 5000ms of EEG data with each point recorded at 8ms intervals.
-



From 27 subjects Total segments extracted: 853

All the 853 segments are collated in one segment and then features are extracted.
With shape: (853, 625, 16)

```
{'Aha': 279, 'Doing Other Task': 276, 'Walking': 261,
'Re-evaluation': 23, 'Impasse': 14}
```

Training data shape: (682, 625, 16)
Test data shape: (171, 625, 16)

>> Spectogram -> CNN -> flatten ->
>> Use min, max features
>> Class imbalance - increase the weight, Augment

Built a CNN

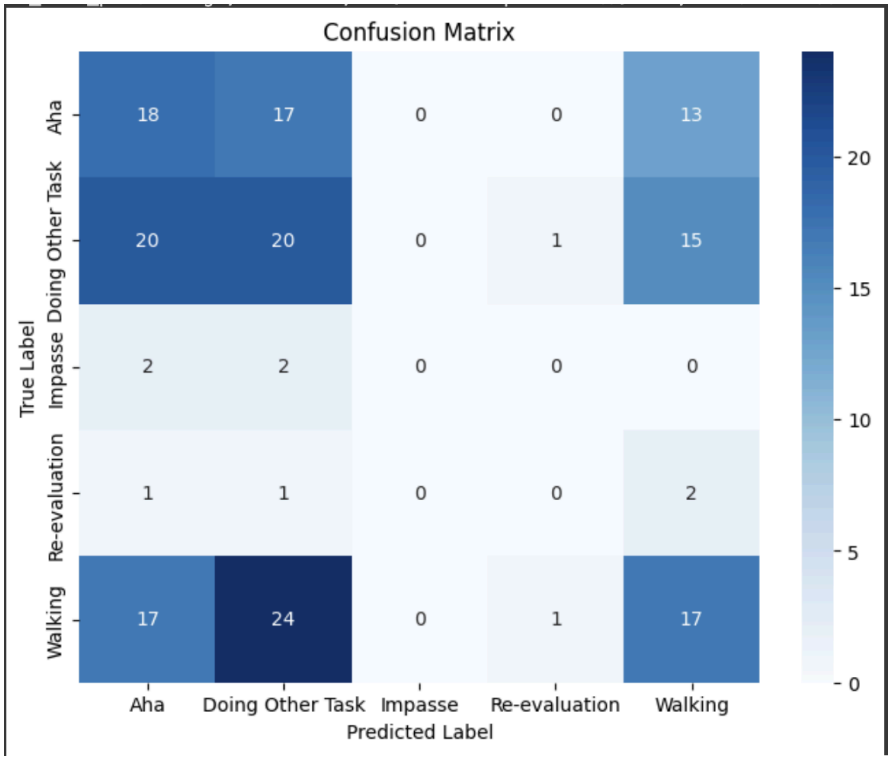
- Input: Shape (X_train.shape[1], X_train.shape[2])
- Conv1D(64 filters, kernel_size=3, ReLU)
- MaxPooling1D(pool_size=2)
- Conv1D(128 filters, kernel_size=3, ReLU)
- MaxPooling1D(pool_size=2)
- Flatten
- Dense(128 units, ReLU)
- Dense(y_train.shape[1], Softmax)

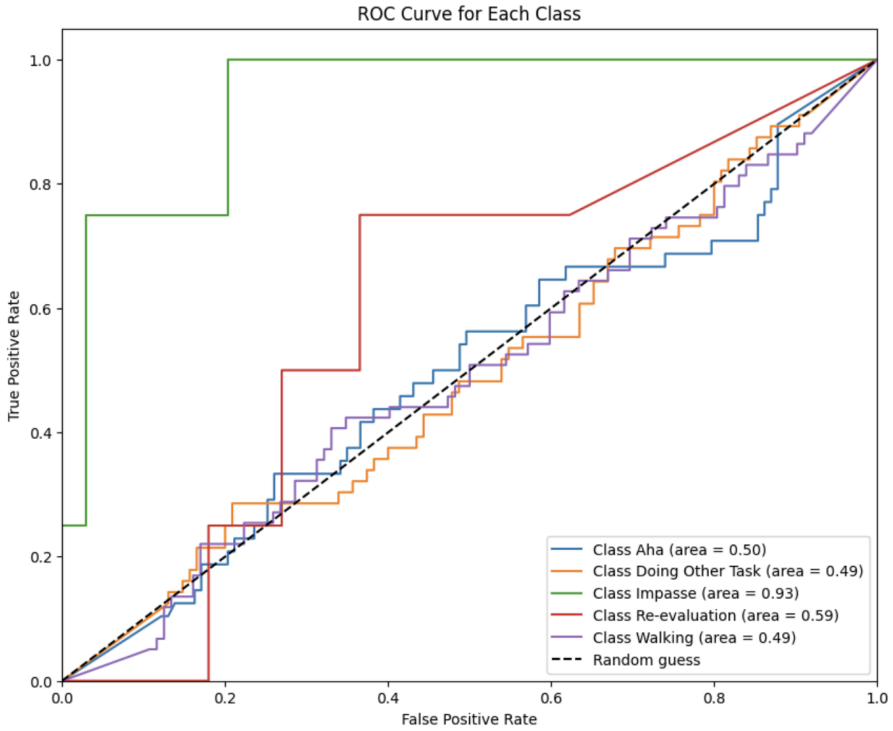
```
Evaluating the model...
6/6 ━━━━━━━━━━━━━━━━━━━━ 1s 70ms/step - accuracy: 0.3089 - loss: 131.3217
Test Loss: 100.6965
Test Accuracy: 0.3216
6/6 ━━━━━━━━━━━━━━━━━━━━ 1s 152ms/step

Classification Report:
              precision    recall  f1-score   support

           0       0.31      0.38      0.34        48
           1       0.31      0.36      0.33        56
           2       0.00      0.00      0.00         4
           3       0.00      0.00      0.00         4
           4       0.36      0.29      0.32        59

    accuracy                           0.32       171
   macro avg       0.20      0.20      0.20       171
weighted avg       0.31      0.32      0.32       171
```


Confusion Matrix
True Label
Predicted Label
Aha
Doing Other Task
Impasse
Re-evaluation
Walking
18 17 0 0 13
20 20 0 1 15
2 2 0 0 0
1 1 0 0 2
17 24 0 1 17
0
5
10
15
20


ROC Curve for Each Class
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class Aha (area = 0.50)
Class Doing Other Task (area = 0.49)
Class Impasse (area = 0.93)
Class Re-evaluation (area = 0.59)
Class Walking (area = 0.49)
Random guess

In [1]Huang, Min, et al. "Feature Representation for Meditation State Classification in EEG Signal." 2021 11th international conference on information technology in medicine and education (ITME). IEEE, 2021.

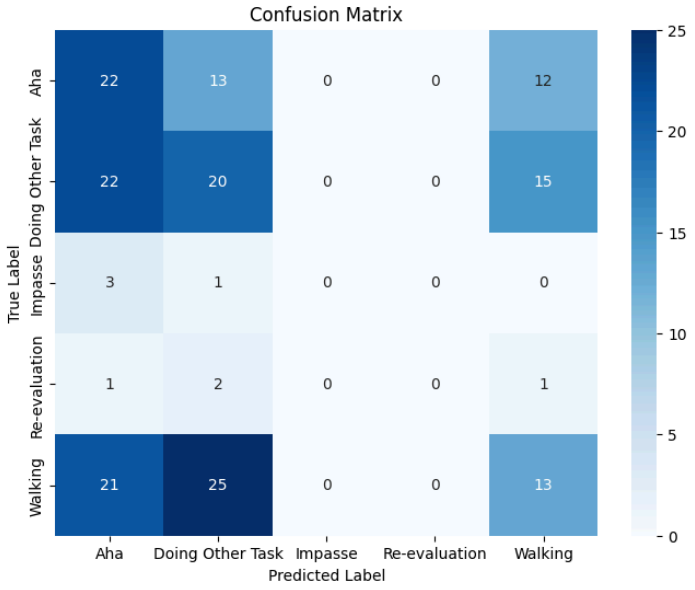
Feature Extraction:

Original Power Features: Analyze the power spectrum of EEG signals in specific frequency bands (delta, theta, alpha, beta, gamma) to capture the brain's electrical activity.

Power Ratio Features: Calculate ratios between the powers of different frequency bands to identify relative changes that may indicate different mental states.

Multi-Modal Features: Combine all extracted features into a single 15-dimensional feature set, aiming to capture comprehensive information about the EEG signals.
Classification: Use a Random Forest classifier to categorize the EEG data into the two meditation states based on the extracted features. This ensemble method aggregates predictions from multiple decision trees for improved accuracy.

Combining 2 features:
Cross-validated accuracy: 0.36
Confusion Matrix:



Classification Report:

	precision	recall	f1-score	support
Aha	0.32	0.47	0.38	47
Doing Other Task	0.33	0.35	0.34	57
Impasse	0.00	0.00	0.00	4
Re-evaluation	0.00	0.00	0.00	4

Walking	0.32	0.22	0.26	59
accuracy			0.32	171
macro avg	0.19	0.21	0.20	171
weighted avg	0.31	0.32	0.31	171

While doing individually:
opf - 0.29
prf - 0.34

```
BANDS = {
   'delta': (0.5, 4),
   'theta': (4, 8),
   'alpha': (8, 13),
   'beta': (13, 30),
   'gamma': (30, 50)
}
```

Next:

1. **Class Imbalance**
2. **Feature Engineering**
3.