MENTAL STATE CLASSIFICATION USING EEG

CDSS Final Project

Official Intelligence Team 4 Members

Name	Section	Bench number
Ahmed Mohamed Abd ElRaouf	**1**	**7**
Adham Mohamed	**1**	**9**
Shrouk shawky	**1**	**46**
Omar Saad ElGharbawy	**2**	**3**
Amr Mohamed	**2**	**7**

Mental State Classification using EEG

Introduction

Problem definition

Automatic detection of mental states, whether related to cognition or emotions, has numerous potential applications across various domains, including healthcare, education, neuroscience, and robotics. The ability to accurately detect an individual's mental state in real-time can provide valuable insights into their cognitive or emotional processes, which can be used to improve their well-being and performance. In **healthcare**, mental state classification can be used to monitor and diagnose mental health conditions such as depression, anxiety, and dementia. In **education**, it can be used to assess students' engagement and attention levels, allowing for personalized learning experiences. In **neuroscience**, it can be used to study brain function and its relationship with mental states. In **robotics**, mental state classification can be used to improve human-robot interaction by enabling robots to adapt their behavior and responses based on the individual's mental state. Overall, the ability to autonomously detect mental states has the potential to revolutionize various fields and improve the quality of life for individuals.

We can detect these signals using EEG devices but most EEG recording systems are expensive medical devices. However, there are some EEG recording systems that are affordable and can be used in a consumer market.
The Muse headband is an affordable and flexible tool that candetetct EEG signals for mental state classification. This headband is designed to detect and record brainwave activity using four EEG sensors (TP9, AF7, AF8, TP10) placed on the forehead and behind the ears. The Muse headband has the advantage of being relatively inexpensive compared to other EEG devices on the market, making it accessible to a wider range of researchers and practitioners.In addition to its low cost, the Muse headband is also highly flexible and portable, allowing for use in a variety of settings. Its wireless connectivity and small size make it easy to use in both laboratory and real-world settings, such as classrooms or clinical settings.
Overall, the Muse headband offers a cost-effective and versatile solution for mental state classification. Its portability and flexibility make it a useful tool for a range of applications, from research to clinical settings, and its affordability makes it accessible to a wider range of users.

Methodology

Our methodology about mental state classification using EEG includes: searching for data, data processing, feature extraction, feature selection, different models evaluation, and model training :

1. Searching for data: The first step is to search for data that can be used to train the model. There are a number of databases that contain EEG data that can be used for this purpose.

2. Data processing: Once the data has been found, it needs to be processed. This involves cleaning the data and removing any noise. The data may also need to be normalized so that the features are on a comparable scale.

3. Feature extraction: The next step is to extract features from the data. Features are measurements of the EEG signal that can be used to distinguish between different mental states. There are a number of different features that can be extracted from EEG data, such as Frequency domain Features and time domain features.

4. Feature selection: Once the features have been extracted, it is important to select the most important features. This can be done using a variety of methods, such as statistical tests and machine learning algorithms to reduce the data complexity and improve the model performance.

5. Different models evaluation: Once the features have been selected, it is important to evaluate different models. This can be done by training the models on the data and then testing them on a held-out set of data. The models can be evaluated using a variety of metrics, such as accuracy, precision, and recall.

6. Model training: Once the best model has been selected, it can be trained on the entire dataset. The model can then be used to classify the mental state of new EEG data.

Literature review

The brain is one of the most significant body parts. It is the human central nervous system, and it is made up of both grey and white matter in addition to being contained within the skull and spinal cord. Brain is the primary center of regulation and manages all bodily functions, including breathing and heartbeat. It is a very intricate system that displays vibrant spatiotemporal dynamics throughout the entire body.

The Electroencephalogram which is abbreviated as (EEG) is clinically used to investigate brain disorders or to diagnose various brain functionalities. The first attempt at measuring this brain disorder activity was done by British Physician Richard Caton in the year 1875. Electroencephalographic record is one of the most important tools for the study of the brain electrical activity and for the diagnosis of neurological diseases.

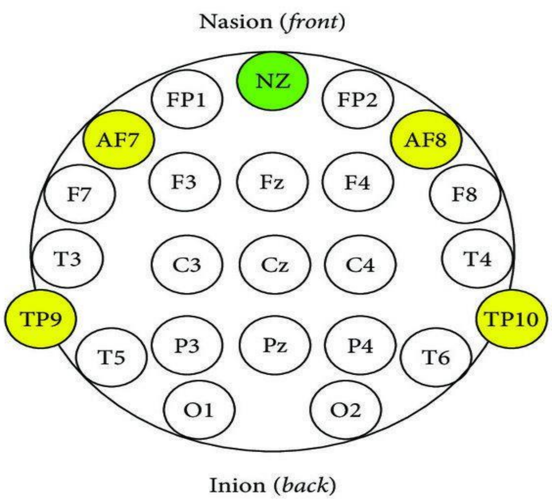
There are various research papers that use EEG signals to classify diseases or the mental state of the brain. In 2017 Nitendra Kumar et al proposed a wavelet based feature extraction technique to EEG signals to classify presence of epileptic seizure, they used ANN and SVM classifiers, ANN classifier achieved accuracy 96% while SVM achieved accuracy 85.46% [1].

In 2018 Damodar Reddy Edla et al used EEG signal to classify the mental state of the brain then translates these results into actions and commands for various applications, they used real time data and built a Random Forest model that achieved 75% accuracy [2].

In 2018 Jordan J. Bird et al proposed a feature-based approach utilizing short-term windowing of EEG data from five signals to classify three mental states: neutral, relaxed, and concentrated, they test different classifiers and found that Random Forest give them the best accuracy 87.16% [3].

Data set and Feature Extraction

The data set provided by the source was collected using *Muse Headband* which is a commercial EEG sensing device with five dry-application sensors, one used as a reference point (NZ) and four (TP9, AF7, AF8, TP10) to record brain wave activity.



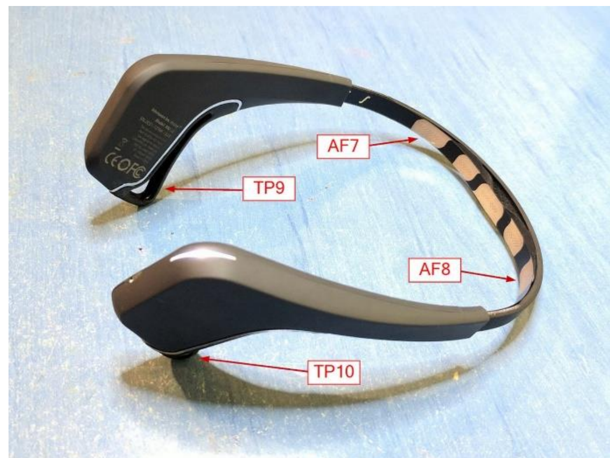


Fig. 1. The International 10-20 EEG Electrode Placement Standard Highlighted in yellow are the sensors of the *Muse Headband*. The NZ placement (green) is used as a reference point for calibration

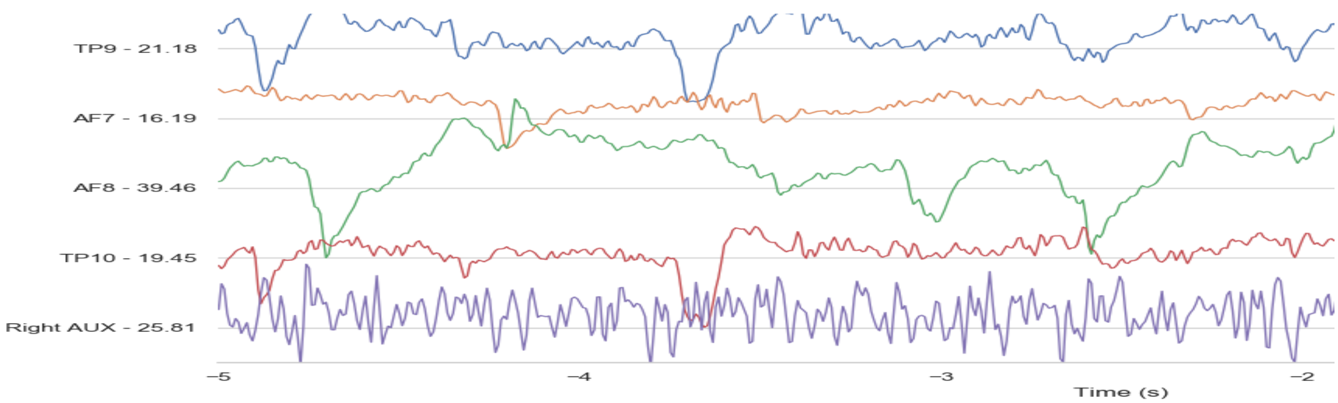


Fig. 2. Example of a live EEG stream of the four *Muse* sensors, Right AUX did not have a device and was discarded due to it simply being noise. This live feed graph has a Y-Axis of measured microvolts at t=0 on each sensor, and an X-axis detailing the time reading.

We have a dataset with four individuals and sessions lasting one minute for each class of mental state in order to train and test different methods. Each session includes the signal data that was collected from the four electrodes and the time-stamp of each sample data point.
The features extracted from the data rely on statistical techniques, time-frequency based on fast Fourier transform (FFT), Shannon entropy, max-min features in temporal sequences, log-covariance and others. Thus, the total number of feature values extracted from these signals is 450 values.

I. Preprocessing

After we made some observations about the timestamps of each dataset, we noticed some inconsistencies in the sampling periods (especially *subject b - relaxed - 2*)
In order to solve this issue, we interpolated the datasets with sampling frequency 200Hz

II. Features extraction

All features proposed to classify the mental states are computed in terms of the temporal distribution of the signal in a given time window. i.e. all features are computed within this instant. An overlap of 0.5 second is used when moving the window, i.e. the temporal window 1 (w1) starts at 0 sec. and finishes at 1 sec.; w2 starts at 1.5 sec. and finishes at 2.5 sec.; w3 starts at 2 sec. and finishes at 3 sec.; w4 starts at 2.5 sec. and finishes at 3.5 sec., and so on.

1-Statistical Features:

We are using a set of classical statistical features since they are useful with proven efficiency to complement a set of multiple features in order to recognize patterns in time series.
These features include:
Given the sets of windows of 1 sec.
- The Mean value
- The Standard Deviation.
- Statistical Moment of 3rd and 4th order {skewness, k = 3; kurtosis, k = 4}.

Given a time window of 1 sec. for each time window, we split the time window by 2.
such w/2 = 0.5 sec. and w = 1 sec., resulting in two sequences of data at ~125 Hz

- Mean Derivative :

$$uv = (\mu\,(w) - \mu\,(w/2))/2$$

- Maximum and Minimum values

$$max(t) = (max(w) - max(w/2))/\ 2 \qquad min(t) = (min(w) - min(w/2))/2$$

The next temporal features are extracted after splitting the initial time window of one second into 4 batches of 0.25 sec.

- Mean {μ1, μ2, μ3, μ4},
- Max {max1, max2, max3, max4}
- Min {min1, min2, min3, min4}

Then we compute the 1D Euclidean distance among all four mean values,
$\delta\boldsymbol{\mu}12 = |\ \mu1 - \mu2|$, $\delta\boldsymbol{\mu}13 = |\ \mu1 - \mu3|$, $\delta\boldsymbol{\mu}14 = |\ \mu1 - \mu4|$, $\delta\boldsymbol{\mu}23 = |\mu2 - \mu3|$, $\delta\boldsymbol{\mu}24 = |\mu2 - \mu4|$,$\delta\boldsymbol{\mu}34 = |\mu3 - \mu4|$
The same for the Max and Min

2-Shannon Entropy:

For every time window of 1 sec we calculate Shannon entropy which is given by

$$h = -\sum Sj \times log(Sj)\ j$$

where h is a feature computed in every time window of 1 sec. and Sj is each element (normalized) of this temporal window. Then, given the same time window.

3-log-energy entropy:

given the same time window, we split into two to compute the log-energy entropy as follows.

$$loge = \sum log\ (Si\ 2\ )\ i + \sum log(Sj\ 2\ )\ j$$

where i represents an index for the elements of the first sub window (0 - 0.5 sec.) and j represents an index for the second sub window (0.5 - 1 sec.).

4-Frequency domain Features:

The frequency domain features for each 1 sec window are computed as follows:

- Power Spectral Density (PSD): The PSD is the square of the magnitude of the Fast Fourier Transform (FFT) coefficient. It is calculated for each window and represents the distribution of power in the frequency domain of the signal.
- Mean PSD: The average of the PSD values for the window.
- Minimum PSD: The minimum value of the PSD for the window.
- Maximum PSD: The maximum value of the PSD for the window.

- Spectral Entropy: Spectral entropy is a measure of the complexity of the signal spectrum. It is calculated by normalizing the PSD and computing the Shannon entropy of the normalized distribution. Entropy is a measure of the disorder or randomness of a system.

- Spectral Edge Frequency (SEF): SEF is the frequency below which a specified percentage of the total spectral power of the signal is contained. In this code, SEF is computed by finding the frequency at which the cumulative sum of the normalized PSD is equal to half of its maximum value.

- Peak Frequency: Peak frequency is the frequency at which the PSD is maximum. It represents the dominant frequency component in the signal. It is calculated by finding the frequency at which the PSD is maximum.

- Frequency Band Filtering: To extract the features related to specific brain wave frequencies $\{\alpha, \beta, \theta, \delta, \gamma\}$, each frequency band is filtered out of the signal spectrum. This is done using band-pass filters with a specific frequency range for each band. For each filtered frequency band, the same statistical features used above:

Mean: The average value of the filtered signal in the frequency band.
Standard Deviation: A measure of how much the filtered signal deviates from its mean.
Skewness: A measure of the asymmetry of the filtered signal distribution.
Kurtosis: A measure of the "peakedness" of the filtered signal distribution.

- Phase synchronization: For each pair of electrodes, we calculated the Phase synchronization which is a measure of the synchronization of the phase angles of two EEG signals at different electrode sites, computed from the phase differences of the two signals in the frequency domain. This is done for each band of the brain wave frequencies $\{\alpha, \beta, \theta, \delta, \gamma\}$

Feature selection Algorithms

The objective of feature selection is to eliminate data that does not contribute to the model's performance and only adds unnecessary computational burden. In this study, six datasets were created using various algorithms while keeping the same data points. However, each dataset had a reduced number of attributes selected by the Filter algorithm.

Filter methods are a type of feature selection algorithm that selects features based on their statistical properties, independent of any specific machine learning algorithm and it is less computationally expensive than other feature selection algorithms. These methods use a scoring function to rank the features and select the top-ranking features for the model.

The filter-based feature selection algorithm that we used includes four methods: Variance, Correlation, Mutual information filter, and univariate ROC AUC. Let's take a closer look at each of these methods:

Variance: This method selects features based on their variance or spread in the dataset. Features with low variance are likely to be constant or nearly constant, and thus carry little or no information. Therefore, variance-based feature selection removes such features. The variance threshold is set manually, and all features with a variance below the threshold are removed.

Correlation: This method selects features based on their correlation with the target variable. Highly correlated features are likely to be redundant, and therefore, one of them can be removed without affecting the performance of the model. The correlation coefficient between each feature and the target variable is calculated, and features with a low correlation coefficient are removed.

Mutual Information Filter: This method selects features based on their mutual information with the target variable. Mutual information is a measure of the amount of information shared between two variables. Therefore, features with high mutual information with the target variable are likely to be important for the model. The mutual information score is calculated for each feature, and the features with the highest mutual information score are selected.

Univariate ROC AUC: This method selects features based on their ability to discriminate between the positive and negative classes of the target variable. The ROC AUC (Receiver Operating Characteristic Area Under Curve) score is calculated for each feature using one decision tree per feature, to predict the target, then make predictions and ranks the features ROC-AUC metric, and the features with the highest ROC AUC score are selected.

Each of these filter-based feature selection methods has its advantages and limitations. For example, the variance method is simple and computationally efficient, but it may remove informative features if their variance is low. The correlation method is useful for identifying redundant features, but it does not consider the interaction between features. The mutual information method is effective in identifying informative features, but it may not work well with high-dimensional data. The univariate ROC AUC method is useful in selecting features that discriminate between classes, but it may not work well with imbalanced data.

Since, each method has its strengths and weaknesses, and the choice of method depends on the specific problem and the characteristics of the dataset.

So, in our study we tried different filters to get the best multiple filters combination to reduce the risk of overfitting and improve the performance of our model.

First, we applied basic filters by removing the Constant and Duplicated features across our data set and the calculated variance of each feature and selected 99% threshold to remove quasi constant features that have low variance or variability in their values, so this helped in reducing computational complexity and overfitting by reducing number of features from 450 to 406 without reducing the model performance.

We tried applying the correlation filter to remove features that are highly correlated with each other this reduces the number of features to 288 but it reduced the accuracy by 1 % because it does not consider the interaction between features so we will not use it.

Then we applied Mutual information filter and ranked the features according to their mutual information scores.

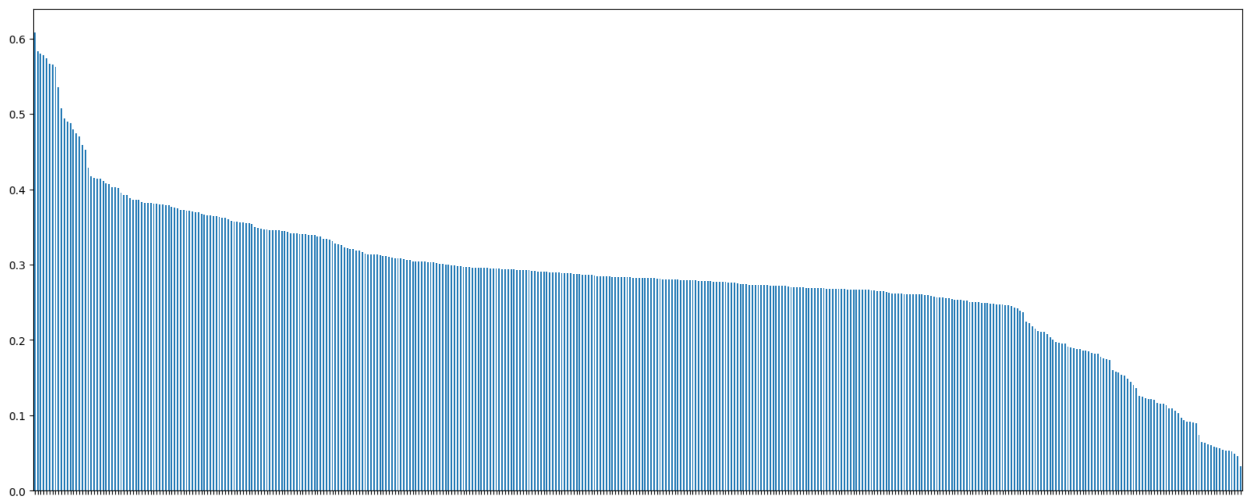


Fig. 3. Sorted best Mutual Information features.

Then we tried selecting the k best mutual information variables, so we tried different K variables to select the best k value for computation and model performance.

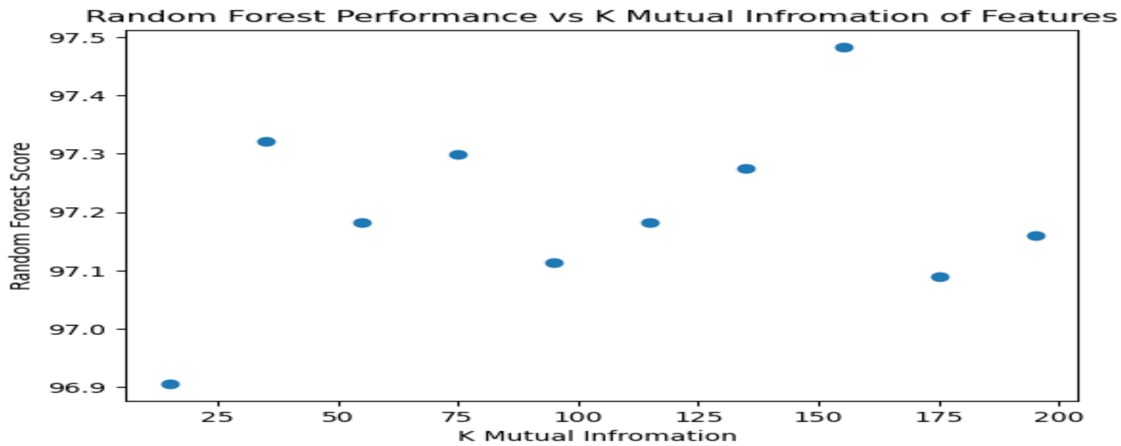


Fig. 4. Random forest performance of different select k mutual information feature

So from the graph we selected 135 features which represent the highest performance and less number of features from the 406 features.

Finally we applied the Univariate ROC AUC filter on the selected 135 features and selected the top 35 features.

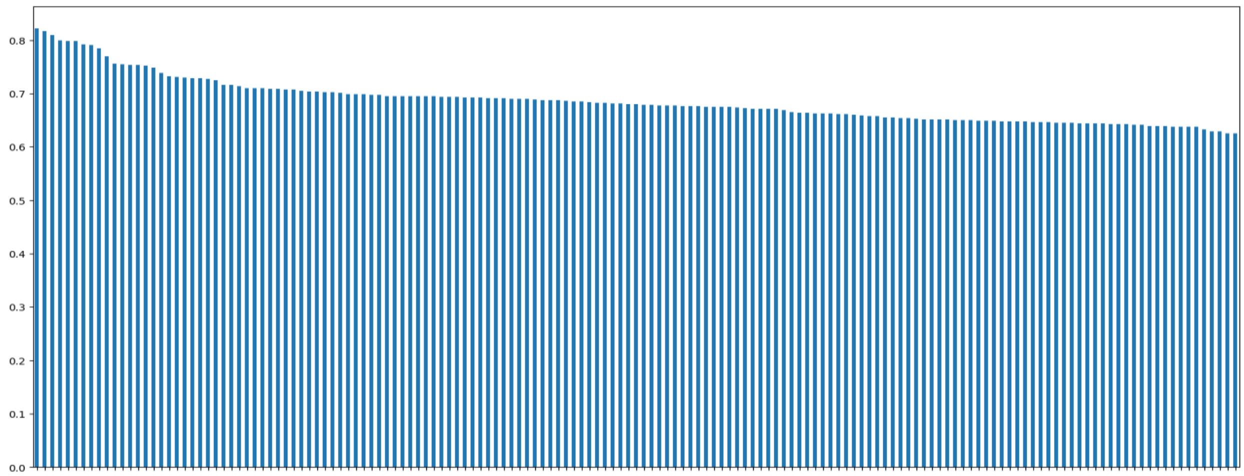


Fig. 5. Sorted best univariate ROC_AUC features

After all of this we successfully selected the best 35 features from the all 450 features to reduce the effect of overfitting and reduce the model complexity without reducing the model evaluation but also improved the model evaluation from 97.5 to 97.6 % with fit time of 0.66 seconds instead of 2.65 seconds.

From feature selection we noticed that The top selected feature is : Gamma, Beta,& Alpha waves and this is logically because :

- Gamma Waves : Represents a pattern of neural oscillation in humans with a frequency between 25 and 140 Hz, and it is used by doctors as an evidence evidence of achieving peak concentratio, and it is engaged with solving a problem and studying.
- Beta Waves : Neural oscillation in the brain with a frequency range of between 12.5 and 30 Hz and it is associated with active, busy or anxious thinking and active concentration.
- Alpha Waves : Neural oscillations in the frequency range of 8–12 Hz and it is associated with relaxed and calm mental states, such as meditation or relaxation.

Machine learning Algorithms

In this study, we aimed to predict brain status from EEG signals using machine learning. To prepare the data for analysis, we applied a categorical encoding to convert the relaxed, neutral, and concentrating classes into numeric classes 0, 1, and 2, respectively, as the classes represent levels of brain status that can have an order. We then selected several models based on the literature review, including: Naive Bayes, Support Vector Machine, Decision Tree, Random Forest, and KNN.
We evaluated the performance of each model using several techniques. First, we plotted a confusion matrix for each model to summarize the actual and predicted classifications of the model on the data.

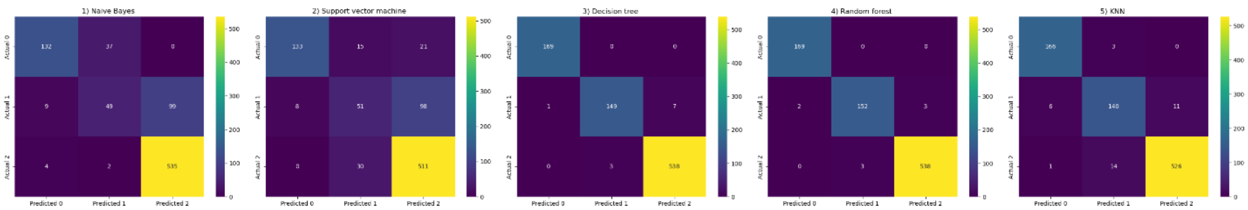


Fig. 6. Different models confussion matrixes.

To prevent overfitting and obtain a more reliable estimate of the model's performance, we used a 10-fold cross-validation technique. From the results, we found that the Random Forest model had an average accuracy of 97.4% across the folds, indicating that it was able to generalize well to new data.

We also calculated several performance metrics, including precision, recall, accuracy, F1 score, negative predictive value, specificity, and sensitivity. From these metrics, we found that the Random Forest model performed the best, with an accuracy of 97.6% and a precision of 97.4%.
We also plotted ROC curves for each model, which provided a comprehensive visualization of the trade-off between sensitivity and specificity. We used a one-vs-all ROC curve to handle multi-classes and found that the Random Forest model performed the best with an AUC of 1, 0.99, and 1 for each class.

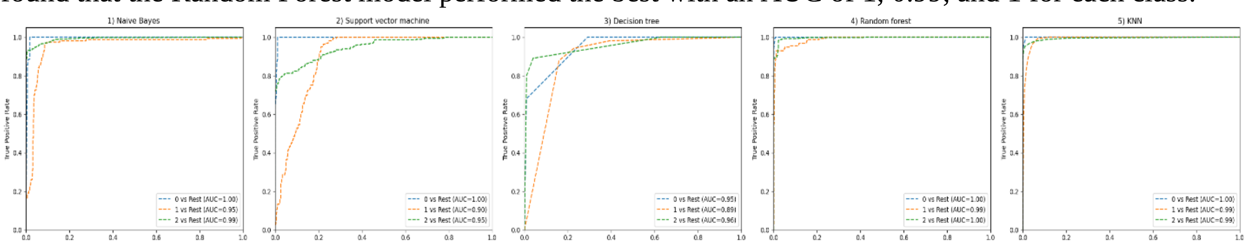


Fig. 7. Different models ROC Curves.

Overall, our study shows that Random Forest is an effective machine learning model for predicting brain status from EEG signals. Our findings highlight the importance of using a categorical encoding for ordered classes and evaluating model performance using various techniques, such as the confusion matrix, performance metrics, ROC curves, and cross-validation. These practices can help ensure that machine learning models are reliable and can generalize well to new data.

For new testing, we will need to use Muse EEG data that has been preprocessed and segmented into windows, as described in our methodology and extract the selected features from the windows of the four channels it to enter it the the model to predict the mental state.

Different models and feature selection accuracy

TABLE I. TABLE TO SHOW ACCURACY OF TRAINED MODELS

Dataset	Model Accuracy %				
	Naive Bayes	SVM	Decision tree	Random forest	KNN
All Data	77.7	78.7	96.1	97.5	86.5
Basic Filter	77.5	78.7	96.3	97.5	86.5
Basic Filter + Uncorrelated	74.8	**58.6**	93.7	96.3	85.8
Basic Filter + Mutual information (k = 35)	80	78.7	96.4	97.3	90.6
Basic Filter + Mutual information (k = 155)	79.1	82.8	95.6	97.48	87.6
Basic Filter + Mutual information (k = 155) + ROC_AUC (k = 35)	80.6	76.6	96.2	**97.6**	92.6
Basic Filter + ROC_AUC (k=35)	80	75.6	95.8	97.37	92.8

TABLE II. FIT TIME & NUMBER OF ATTRIBUTES SELECTED BY FILTERS OF THE ORIGINAL 450 ATTRIBUTES

Attribute Selection Filter	**No. of attributes selected**	**Fit Time**
All Data	450	2.65
Basic Filter	406	2.63
Basic Filter + Uncorrelated	288	2.32
Basic Filter + Mutual information (k = 35)	35	0.62
Basic Filter + Mutual information (k = 155)	155	1.45
Basic Filter + Mutual information (k = 155) + ROC_AUC (k = 35)	**35**	**0.66**
Basic Filter + ROC_AUC (k=35)	35	0.62

Here we can observe that the best combination of Selection filters is the basic filter which include removing (Constant, Quasi-Constant and Duplicated) features then selecting the best 135 Mutual information features then selecting the best 35 Univariate ROC-AUC features which reduces the all number of features from 450 to 35 so it will reduce computation power and model fitting time with improving the model accuracy by 1%.

Project Block Diagram

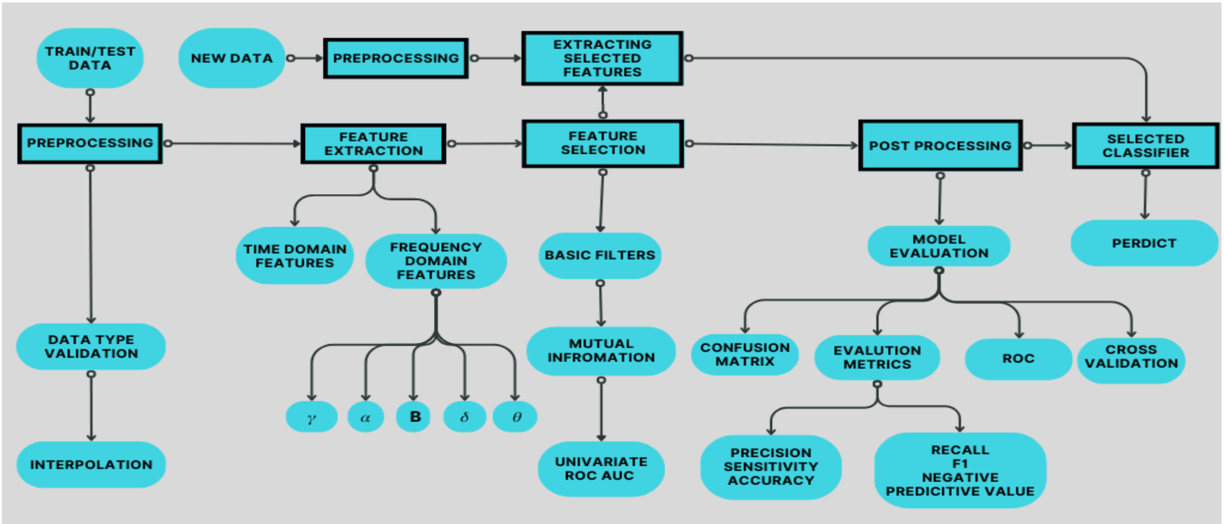


Fig. 9. Block Diagram.

Results and Discussion

The present study aimed to classify three different mental states (relaxing, neutral, and concentrating) using EEG data collected from the Muse headband. The EEG data was preprocessed and analyzed to extract relevant features for classification. The selected 35 features were then used to train and test different machine learning models using 10-fold cross validation, including SVM, KNN, decision tree, random forest, and naive Bayes, and comparing between them using different evaluation metrics including precision, recall, accuracy, ROC-AUC Curve and confussion matrix. The best-performing model was found to be the random forest with an accuracy of 97.6%.

The high accuracy of the random forest model suggests that the selected features were informative and that the model was able to effectively distinguish between the different mental states. Also because of its ability to handle complex and non-linear relationships between the features and the target variable. The random forest model is known for its robustness, ability to handle high-dimensional data, and resistance to overfitting, which may have contributed to its superior performance in this study compared to other models.

In terms of limitations, the study had a relatively small sample size, and the results may not be generalizable to other populations or contexts. Additionally, the study only classified three mental states, and it is possible that additional mental states or subcategories may require different features or models for accurate classification. Future studies could explore the use of more complex models or additional features to further improve the accuracy of mental state classification using EEG data.

In comparison to previous studies we achieved higher accuracy than the previous paper we followed [3], which reported an accuracy of **87.16%** . We reached this by selecting more effective features for each frequency band which are (Power spectral density, Spectral entropy, Spectral edge frequency, Spectral peak frequency and Phase synchronization) and using suitable feature selection techniques, our methodology achieved a higher accuracy of **97.6%**. These findings suggest that our approach has the potential to improve the accuracy of mental state classification using EEG data and could have important implications for various applications such as mental health monitoring and brain-computer interfaces."

In conclusion, the present study demonstrates the potential of the Muse headband and machine learning techniques for accurate classification of mental states using EEG data. The high accuracy of the random forest model suggests that this approach could be useful for various applications, such as mental health monitoring and brain-computer interfaces. However, further research is needed to validate these findings and explore the potential of this approach for other mental states and populations.

References

[1] Kumar, N., Alam, K., & Siddiqi, A. H. (2017). Wavelet Transform for Classification of EEG Signal using SVM and ANN. *Biomedical and Pharmacology Journal*, *10*(4), 2061-2069.

[2] Edla, D. R., Mangalorekar, K., Dhavalikar, G., & Dodia, S. (2018). Classification of EEG data for human mental state analysis using Random Forest Classifier. *Procedia computer science*, *132*, 1523-1532.

[3] Bird, J. J., Manso, L. J., Ribeiro, E. P., Ekart, A., & Faria, D. R. (2018, September). A study on mental state classification using eeg-based brain-machine interface. In *2018 international conference on intelligent systems (IS)* (pp. 795-800). IEEE.