**CS597 - Classification of Cognitive States Using EEG Signals: Impasse, Aha!, Uncertainty**

**By Noviya Balasubramanian (A20541236)**

**Prof. Gady Agam, Prof. Ivan Mutis, Prof. Kristina Bauer**

**Illinois Institute of Technology Chicago, Illinois, USA**

# **Overview:**

# This comprehensive study focuses on classifying cognitive states using EEG signals, with a particular emphasis on Impasse, Aha!, and Uncertainty. The research utilizes a unique dataset collected from 65 participants who underwent an OSHA safety examination in a mixed-reality environment. The project encompasses various aspects of EEG signal processing and classification, including advanced preprocessing techniques, innovative feature extraction methods, and state-of-the-art machine learning approaches. The study employs a range of methodologies, from traditional summary features to deep learning techniques like EEGNet, E2FNet, Tsception and autoencoders, to enhance the accuracy of cognitive state detection.

# **Introduction:**

Electroencephalography (EEG) has emerged as a powerful tool for studying cognitive states, offering insights into the complex workings of the human brain. This non-invasive technique measures electrical activity produced by neurons, providing a window into various mental processes and states. EEG's high temporal resolution allows researchers to capture rapid changes in brain activity, making it particularly suitable for investigating dynamic cognitive phenomena.

The study of cognitive states through EEG has gained significant traction in recent years, with researchers exploring a wide range of mental processes, including attention, memory, decision-making, and problem-solving. Of particular interest are states such as Impasse, Aha!, and Uncertainty, which play crucial roles in cognitive processes and problem-solving scenarios. These states represent different phases of cognitive engagement and can provide valuable insights into how individuals approach and resolve challenges.

Advancements in EEG signal processing and machine learning techniques have greatly enhanced our ability to classify and interpret these cognitive states. Researchers now employ sophisticated algorithms to extract meaningful features from EEG data, allowing for more accurate and nuanced classification of mental states. This progress has opened up new avenues for understanding human cognition and has potential applications in fields ranging from education to human-computer interaction.

This comprehensive study focuses on classifying cognitive states using EEG signals, with a particular emphasis on Impasse, Aha!, and Uncertainty. The research utilizes a unique dataset collected from 65 participants who underwent an OSHA safety examination in a mixed-reality (MX) environment. As project engineers, subjects were tasked with identifying safety violations in a virtual construction site, while their EEG signals were recorded using an OpenBCI Mark IV headset across 16 channels.

The project encompasses various aspects of EEG signal processing and classification, including advanced preprocessing techniques, innovative feature extraction methods, and state-of-the-art machine learning approaches. The study employs a range of methodologies, from traditional summary features to deep learning techniques like EEGNet and autoencoders, to enhance the accuracy of cognitive state detection. By combining these diverse approaches, the research aims to provide a comprehensive framework for understanding and classifying complex cognitive states in real-world problem-solving scenarios.

# **Literature Review: EEG Signal Processing and Classification**

The process of analyzing and classifying EEG signals has undergone significant advancements, particularly in the areas of denoising, feature extraction, and classification using deep learning techniques.

Denoising Techniques

A variety of denoising techniques have been explored to enhance the quality of EEG data. Blind Source Separation (BSS), particularly Independent Component Analysis (ICA), decomposes EEG signals into their components, identifying and removing artifacts such as eye blinks and muscle movements. Advanced ICA methods, such as ERASE, offer enhanced artifact removal capabilities. Principal Component Analysis (PCA) is another common technique that reduces the dimensionality of EEG data by focusing on components with the highest variance, although it is less effective for specific artifact types compared to ICA. Canonical Correlation Analysis (CCA) is particularly effective in distinguishing brain activity from muscle artifacts by maximizing the correlation between EEG and artifact signals. Discrete Wavelet Transform (DWT) decomposes EEG signals into frequency components, which helps in artifact identification and removal. High-Order Spectral Analysis and Nonlinear Dynamic Analysis, such as using Lyapunov exponents, provide additional insights into complex patterns and are particularly effective for emotion detection and neuro-marketing applications [1].

Feature Extraction Methods

Various feature extraction methods have been employed to improve the performance of EEG signal classification models. Time-frequency analysis, which combines time and frequency domains, is commonly used for extracting features like mean power and frequency for different brainwave frequencies (alpha, beta, theta, delta). In one approach, EEG data is segmented into 10-second intervals, with 1-second segments for each brainwave frequency band extracted using Fourier transforms. These features, such as mean power and frequency, are then used as inputs for models like Long Short-Term Memory (LSTM) networks to classify different cognitive states [1].

Fractal and statistical features also enhance the discriminative power of EEG classifiers. For example, variance, band power (from alpha and beta bands), minimum energy, and time-sequence complexity (TSC) are used to capture the signal’s variability and dynamic properties. Fractal dimensions, such as the Roughness of Fractal Dimension (RFD), provide insights into the irregularity of the EEG signal, and Hjorth parameters measure signal activity, mobility, and complexity. Adaptive Autoregressive (AAR) parameters model variations in EEG signals over time [2].

Classification Techniques

Deep learning approaches, particularly convolutional neural networks (CNNs), have demonstrated superior performance in EEG signal classification tasks. For instance, one study employed a neural network model that used graph convolution layers to capture spatial relationships between EEG channels and a self-attentive layer to prioritize significant temporal features, leading to improved classification performance. This graph-based spatial analysis combined with attention mechanisms offers a robust method for classifying cognitive states from EEG data [3].

Recurrent convolutional neural networks (RCNNs) have also been applied to EEG classification, taking advantage of both convolutional layers and recurrent layers to capture temporal dependencies and spatial patterns. This architecture allows for the automatic extraction of features, followed by classification using a softmax output layer [4]. Additionally, combining EEG source imaging with CNNs enhances classification by integrating spatial information from brain regions associated with specific cognitive tasks. In one study, four regions of interest (ROIs) from the motor cortex were used to optimize the classification of motor imagery tasks, further enhanced by time-frequency analysis using Morlet wavelets [5].

Fusion of Multimodal Data

Fusion techniques that combine EEG data with other physiological measures, such as pupillometry, have gained attention for their ability to enhance classification performance. Feature-level fusion of EEG and pupillometry, where both modalities are processed together, has shown statistically significant improvements over single-modality approaches. This suggests the potential of multimodal data fusion for more accurate cognitive state detection, particularly in complex scenarios requiring real-time decision-making or effort monitoring [2].

# Data Description:

The dataset used in this study was created using a construction on-site OSHA safety examination designed as a problem-solving task within a mixed-reality (MX) environment [6]. Participants, acting as project engineers, inspected a virtual construction site modeled after small commercial buildings spanning three stories. During the walk-through safety check process, the subjects were tasked with identifying OSHA safety violations of varying complexity, with the design incorporating the most cited OSHA standards and frequent construction site hazards. Prior to the examination, participants received training to familiarize themselves with the task and tools involved. As shown in Fig.1 they interacted with virtual instruments, such as a measuring tape, to label violations within a 10-minute time limit. The MX environment, powered by the Hololens 2, allowed participants to navigate the virtual site by physically moving within a laboratory space, offering a seamless integration of virtual and real-world interactions. Multimodal data, including video, audio, and EEG signals, were collected during the task. EEG data, recorded with an OpenBCI Mark IV headset, utilized all 16 channels (FP1, FP2, F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8, O1, and O2) arranged according to the 10-20 system and sampled at 125 Hz. Timestamps synchronized across all data streams enabled precise mapping of cognitive states to specific actions in the virtual environment, creating a comprehensive dataset for analyzing cognitive states like Impasse, Aha!, and Uncertainty.

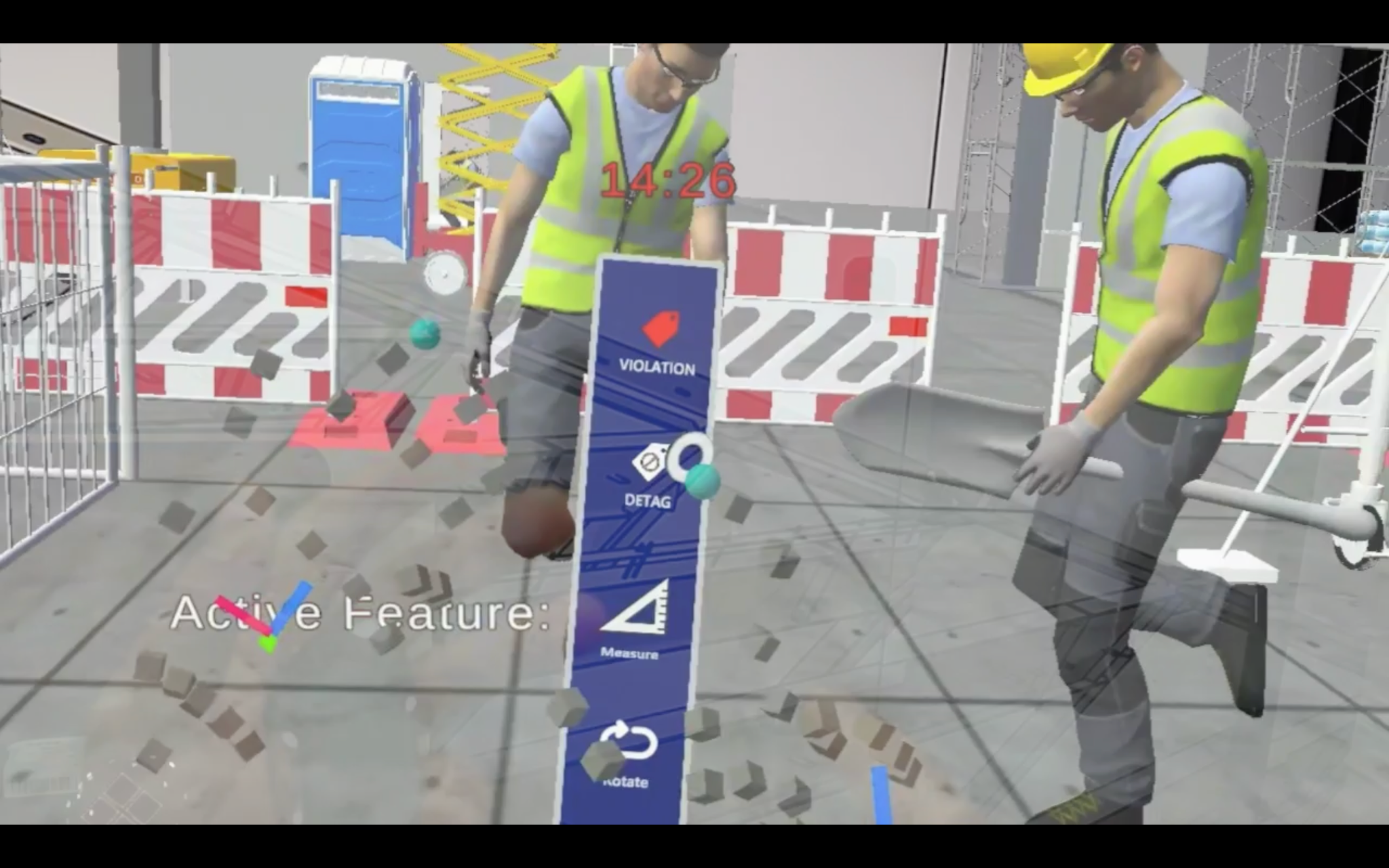


Fig.1: Screenshot of the exam video

## Dataset structure:

Our study involved 65 subjects who underwent CITI training before being granted data access. The experiment was designed to capture a comprehensive set of data points to analyze cognitive processes during various tasks. The following data were collected for each participant: Pupil training data, Video recordings of Empatica training sessions, Video recordings of the final examination, Raw EEG readings during three distinct phases:

- Final examination

- Baseline (while completing a sample quiz)

- OSHA training

Additionally, JSON files were generated to record crucial timestamps throughout the experiment. The "\_FinishJson.txt" file contains the StartTime, indicating when each subject began the exam. The "\_ExaminationJson.txt" file includes event-related timestamps such as:

- TagHandMenuPumpTime: When a subject tagged a component

- DetagHandMenuPumpTime: When a subject de-tagged a component

- BTButtonPressTime: When a subject marked a component for later analysis

The JSON files also record the start and end times for each section of the exam, which was divided into two parts: Section A and Section B. These timestamps play a crucial role in our data labeling process, allowing us to accurately associate EEG readings with specific events and actions during the examination. By utilizing these precise temporal markers, we can segment the EEG data and correlate it with subjects' cognitive states and decision-making processes throughout the experiment.

This comprehensive data collection approach, combined with the detailed timestamp information, enables a nuanced analysis of subjects' cognitive processes, decision-making patterns, and physiological responses across various stages of the experiment, from training to the final examination. The ability to accurately label EEG data based on these timestamps enhances the reliability and specificity of our subsequent analyses.

# Preprocessing:

To ensure high-quality EEG data for analysis, preprocessing was carried out using EEGLAB, following its recommended practices to maximize data integrity. Each subject’s EEG data, sampled at 125 Hz across 16 channels, was imported with channel locations for enhanced spatial understanding, ensuring accurate identification of source components. MARA, an an automated and efficient classifier extension, employs a binary linear classifier to distinguish between independent components (ICs) representing artifacts and those representing neuronal signals, enabling researchers to retain or reject components effectively. This approach has been shown to perform well in various experimental contexts, particularly in handling myogenic artifacts [7]. MARA was used to determine whether an IC is an artifact or a neuronal signal, thereby enabling researchers to retain or reject it.

In MARA, the data was filtered using a Finite Impulse Response (FIR) filter with a passband between 0.5 Hz and 50 Hz to remove low-frequency drifts and high-frequency noise effectively. This step ensured that the relevant frequency bands for cognitive state analysis were preserved, setting the foundation for accurate artifact detection and removal. Independent Component Analysis (ICA) was applied twice during preprocessing, as recommended by EEGLAB, to refine artifact removal and ensure cleaner data for subsequent analysis.

The ICA algorithm used was the extended infomax method (runica.m) in extended mode ('extended', 1) to identify both super-gaussian sources, such as neuronal signals, and sub-gaussian sources, including line noise and slow activity. This ensured that a broader range of signal types could be accounted for, improving the reliability of the decomposition. The component order returned by runica.m is based on the EEG variance accounted for by each component, with components listed in decreasing order of variance. Components with a lower order represented more significant data, including both neural and artifactual signals, crucial for informed artifact rejection. Artifact removal was guided by ICLabel, using the probability thresholds as shown in fig.2 to maintain data quality. Given the physical interaction required in the mixed-reality environment, particular attention was paid to removing muscle and eye movement artifacts, which are highly prevalent in such dynamic tasks. This rigorous approach ensured minimal interference from these artifacts, preserving the integrity of the neuronal signals. Fig.3 shows artifacts marked for removal.

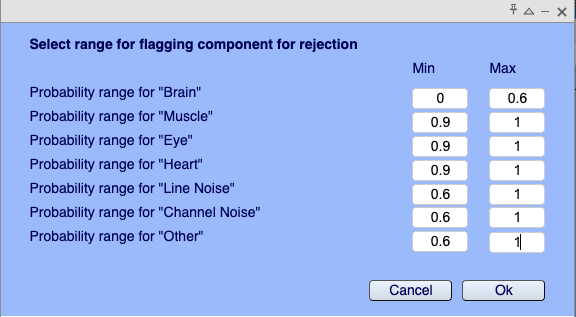


Fig.2: Range given in EEGLab to flag the artifacts

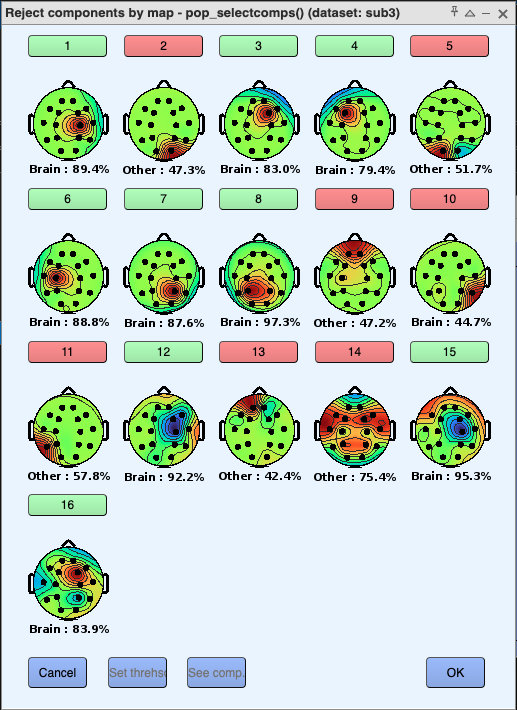


Fig.3: Components marked for rejection

The ICA decomposition was repeated on the cleaned data to enhance reliability and confirm the robustness of the preprocessing steps. Notably, running ICA multiple times can result in slight differences in the ordering, scalp topography, and activity time courses due to the random initialization of the weight matrix and data shuffling during training. Features that did not remain stable across runs were excluded from interpretation, acknowledging these as signs of ICA uncertainty, thus avoiding potential misclassification.

The final preprocessed data was exported in CSV format for labeling and feature extraction, ensuring a robust dataset for downstream analysis. This comprehensive preprocessing workflow provided a reliable foundation for further exploration of cognitive states and task performance.

# 

# 

# Labelling Cognitive States:

To accurately label the EEG data for our classification task, we developed a sophisticated labeling technique that utilizes the timestamps from the JSON files. This approach allows us to associate specific cognitive states with the corresponding EEG readings. The labeling process is implemented using Python and involves several key steps:

1. Time Conversion: We convert event times to milliseconds relative to the experiment's start time, ensuring precise temporal alignment.

2. State Labeling: The algorithm labels different cognitive states based on specific events:

- Impasse: Labeled in two different cases:

a) Using the finishClick time in Section A

b) Using the BTButtonPressTime

- Aha: Labeled for a period following the TagHandMenuPumpTime

- Walking: Labeled for a period preceding the TagHandMenuPumpTime

- Doing Other Task: Labeled for a period following the "Aha" state

- Re-evaluation: Labeled based on the DetagHandMenuPumpTime

3. Timestamp Usage:

- Impasse (Case 1): `finishClick` from `\_FinishJson.txt`

- Impasse (Case 2): `BTButtonPressTime` from `\_ExaminationJson.txt`

- Aha: `TagHandMenuPumpTime` from `\_ExaminationJson.txt`

- Walking: Derived from `TagHandMenuPumpTime`

- Doing Other Task: Derived from `TagHandMenuPumpTime`

- Re-evaluation: `DetagHandMenuPumpTime` from `\_ExaminationJson.txt`

4. Window Selection: To ensure robust labeling, we created four different sets of labeled data by varying the time window before the TagHandMenuPumpTime:

- Set 1: 1 second buffer

- Set 2: 2 seconds buffer

- Set 3: 3 seconds buffer

- Set 4: 4 seconds buffer

Depicted in fig:4.

5. Data Processing: The script processes each subject's data individually, applying the labeling function to the EEG time series.

6. Output Generation: Labeled data for each subject is saved as a CSV file, with an additional column for the assigned labels.

This multi-set approach allows us to experiment with different temporal windows, helping to determine the most appropriate labeling scheme for our classification task. By creating these four sets, we can analyze which time window best captures the cognitive states of interest, potentially improving the accuracy of our subsequent machine learning models.

The flexibility of this labeling technique enables us to fine-tune our approach, ensuring that we capture the most relevant EEG patterns associated with each cognitive state. This methodical labeling process forms a crucial foundation for our classification task, allowing us to train and evaluate our models on precisely labeled data.

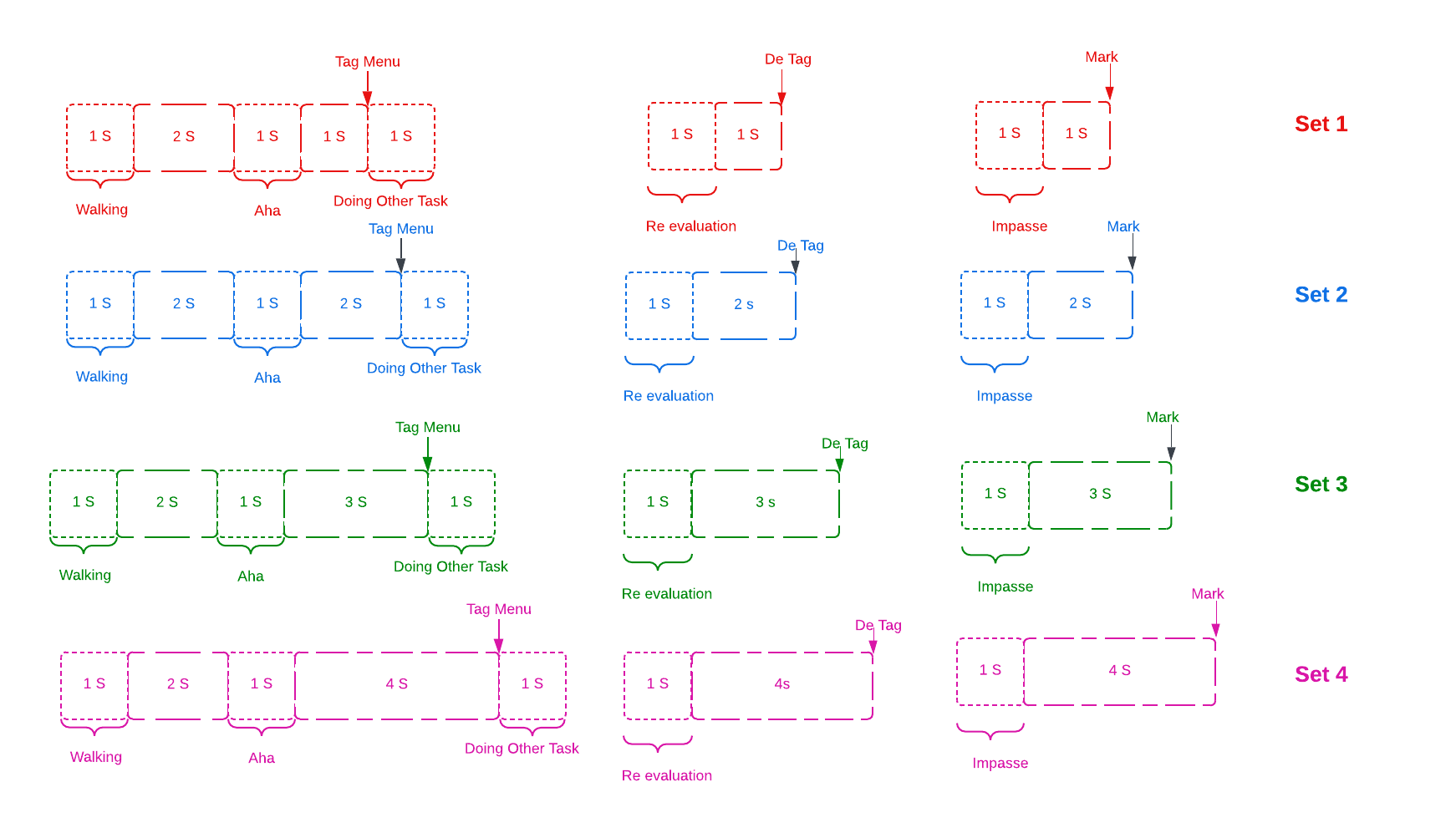


Fig.4: Labeling Methodology

General Labeling Methodology

After the labeling process, we extracted different sets of data into .npy format, each containing approximately 937 samples. Each data sample has a shape of (125, 16), where:

* 125 represents 1000 ms or 1 second of EEG data
* 16 represents the number of EEG channels

This format allows us to capture a full second of brain activity across all channels for each labeled cognitive state. The extraction of these specific time windows enables us to focus on the most relevant EEG patterns associated with each cognitive state.

# Augmentation:

Class imbalance in datasets poses a significant challenge for machine learning models, particularly in EEG-based classification tasks. This issue occurs when certain classes have substantially fewer samples than others, potentially leading to biased model predictions that favor the majority class while underperforming on minority classes. In our project, we encountered this problem with an EEG dataset for emotion recognition, where classes such as "Impasse" and "Re-evaluation" were significantly underrepresented compared to others.

The original dataset comprised 910 EEG segments, each with a shape of (125, 16), distributed across five classes: Walking (290 samples), Aha (290 samples), Doing Other Task (290 samples), Re-evaluation (25 samples), and Impasse (15 samples). This imbalance could potentially hinder the model's ability to accurately classify the underrepresented emotional states.

To address this class imbalance and enhance the dataset, we employed a data augmentation technique using Gaussian noise. This method involves adding noise generated from a normal distribution to the original EEG segments, particularly for the underrepresented classes. The augmentation process can be mathematically represented as:

The Gaussian noise, with a mean of 0 and a standard deviation of 0.05, is designed to mimic the statistical properties of the existing data. This approach ensures that the augmented data retains the underlying characteristics of the original signals while introducing variability to improve the robustness of the classification model.

After applying the Gaussian noise augmentation, the total number of segments increased to 1450, with a balanced distribution across all classes: Walking (290), Aha (290), Doing Other Task (290), Impasse (290), and Re-evaluation (290). The shape of each segment remained unchanged at (125, 16) fig.6.

This augmentation technique offers several advantages. Firstly, it is relatively simple to implement and computationally efficient. Secondly, when applied correctly, it maintains the underlying characteristics of the original EEG signals. Lastly, the addition of noise can help models become more resilient to variations in input, which is crucial in real-world applications where noise is prevalent.

However, it's important to consider some key factors when implementing this augmentation technique. The noise level, represented by the standard deviation of 0.05, should be carefully evaluated to ensure it doesn't distort the underlying signal characteristics. It may be beneficial to experiment with different noise levels to find the optimal balance between augmentation and signal preservation.

To maximize the benefits of data augmentation, it's crucial to validate that the augmented data improves model performance. This can be done by comparing the model's performance on the original imbalanced dataset versus the augmented balanced dataset. Proper cross-validation techniques should also be implemented to assess the model's performance accurately, especially when using augmented data fig.6.

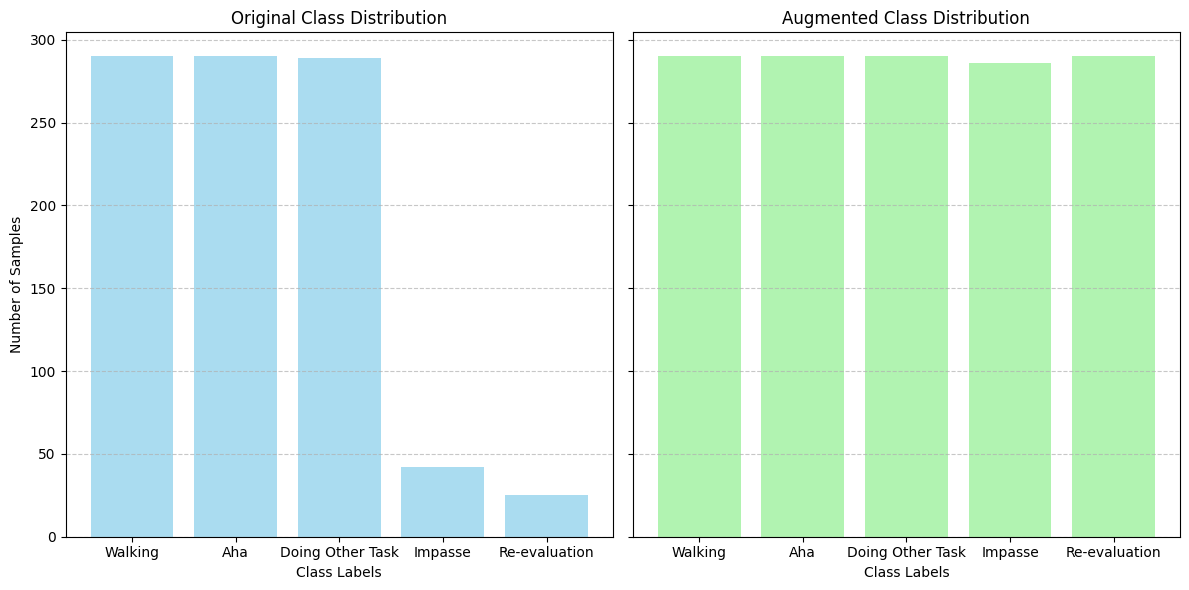


Fig.6: Class distribution before and after augmentation

# Feature Extraction Methodology

Feature extraction was a critical step in processing the EEG data, enabling the identification of patterns and characteristics relevant to cognitive states. A variety of feature extraction models were developed and applied to enhance the analysis, including summary features, a combined model, and multiple CNN-based models.

## Summary Features

Summary Features were extracted from the EEG signals to provide a comprehensive representation of the data. These features were categorized into Time Features, Frequency Features, and Wavelet Features, each offering unique insights into the EEG signal characteristics. This multi-faceted approach ensured that we captured both global and local properties of the signals, enhancing our ability to differentiate between cognitive states.

Time Features: Metrics such as mean, variance, first differences, and second differences were computed to capture the statistical properties of the EEG signals over time. For each of the 853 EEG segments across 16 channels, several time-domain features were calculated.

- Mean Features: These represent the average signal level for each channel, providing a measure of central tendency and reflecting overall activity patterns. The resulting feature set had a shape of (853, 16).

- Variance Features: Capturing variability within each channel, these features offered insights into fluctuations in brain activity and also resulted in a feature set with a shape of (853, 16).

- First Differences: The mean of absolute first differences was calculated to evaluate the dynamics of signal changes between successive time points. This feature indicated how the signal fluctuated over time and had a shape of (853, 16).

- Second Differences: Similarly, the mean of absolute second differences measured the average acceleration of these changes, identifying shifts in the rate of signal variation and yielding a feature set of (853, 16).

To create a comprehensive feature representation, all extracted time-domain features were concatenated into a single feature set referred to as combined time features as explained in [1]. This set incorporated mean features, variance features, first differences, and second differences, resulting in a robust feature representation with a shape of (853, 64). These combined features provided a holistic view of the time-domain characteristics of the EEG signals, enabling detailed analysis of brain activity. By summarizing key statistical and dynamic properties of the signals, these features served as the foundation for initial model training and classification tasks.

Frequency Features: Frequency domain features were derived from spectral analysis techniques applied to the EEG signals to capture energy distribution across frequencies. Using a sliding window approach with segments analyzed at a sampling rate of 125 Hz, we extracted several key frequency features:

- Mean Power: This feature represents the average energy across all frequencies by calculating the mean magnitude of the spectrogram.

- Standard Deviation of Power: This metric indicates variability in power across frequencies, providing insights into stability and fluctuations in EEG signals.

- Peak Frequency: The frequency with the highest average power was identified as peak frequency, highlighting dominant frequency components associated with cognitive processes.

- Frequency Bandwidth: Defined as the difference between the 25th and 75th percentiles of cumulative power, this feature reflects where most signal energy is concentrated.

These frequency features encapsulate spectral characteristics that serve as crucial inputs for downstream classification tasks.

Wavelet Features: Wavelet Features were obtained by applying wavelet transforms to decompose EEG signals into time-frequency representations. This approach enabled the capture of transient and localized features across various scales as mentioned in [2]. Key wavelet-based features included:

- Autoregressive (AR) Coefficients: Extracted using AR modeling to capture temporal dependencies within each EEG segment.

- Shannon Entropy (SE): This quantifies complexity in EEG data by offering insights into randomness and information content.

- Wavelet Variance: Using Discrete Wavelet Transform (DWT), variance in each frequency band was computed to highlight frequency-specific activity.

The methodology effectively combines temporal and frequency-domain information through AR-wavelet feature extraction. By capturing both localized time-domain changes and global frequency-domain patterns, these features provide a comprehensive representation of EEG data that facilitates accurate classification of cognitive states or neurological conditions.

In summary, through careful extraction and combination of time-domain, frequency-domain, and wavelet-based features from our EEG data—resulting in various shapes such as (853, 64) for combined time features and (1450, number\_of\_frequency\_bands) for frequency features—we establish a robust framework for analyzing cognitive states during our experimental tasks.

## Fusion Techniques

In the realm of EEG data analysis, fusion techniques play a crucial role in enhancing the performance of classification models by effectively combining diverse feature sets. These techniques allow for the integration of information from various sources, leading to a more comprehensive representation of the underlying cognitive states. In our study, we implemented three primary fusion strategies: concatenation of summary features, feature extraction using EEGNet, and an innovative autoencoder approach.

1. Concatenation of Summary Features

The first fusion technique we employed was the concatenation of summary features. This method involved combining multiple feature sets derived from different extraction processes into a single high-dimensional vector. Specifically, we integrated time-domain features, frequency-domain features, and wavelet-based features into one comprehensive representation.

This method combined:

* Time-domain features: Shape (1450, 64)
* Frequency-domain features: Shape (1450, 160)

The resulting concatenated feature vector had a shape of (1450, 224). By concatenating these features, we ensured that our model had access to a rich array of information that captured various aspects of the EEG signals. The time-domain features provided insights into the statistical properties and dynamics of the signals over time, while frequency-domain features highlighted energy distribution across different frequency bands. Wavelet-based features contributed additional temporal and spectral information by capturing transient changes in the signals.

The resulting concatenated feature vector allowed for a holistic view of the EEG data, enabling our classification models to leverage all available information effectively. This straightforward yet powerful method is commonly used in machine learning applications, as it preserves all original features while potentially enhancing model performance through increased dimensionality.

2. EEGNet Feature Extraction

To further refine our feature representation, we implemented the EEGNet architecture, a specialized convolutional neural network designed for processing EEG signals. EEGNet is known for its compact design and efficiency in extracting relevant patterns from raw EEG data. By using this model as a feature extractor, we aimed to capture complex spatial and temporal relationships within the EEG signals. The input to EEGNet consisted of raw EEG segments with a shape of (1450, 125, 16), representing 1450 samples, each with 125 time points across 16 channels.

The EEGNet model processed this data through several convolutional and pooling layers designed specifically for EEG signal processing. The output from the flattened layer was used as the learned feature representation, resulting in a feature set with a shape of (1450, 3945). This learned representation captured complex spatial and temporal patterns within the EEG signals that might not be evident in manually engineered features. [3]

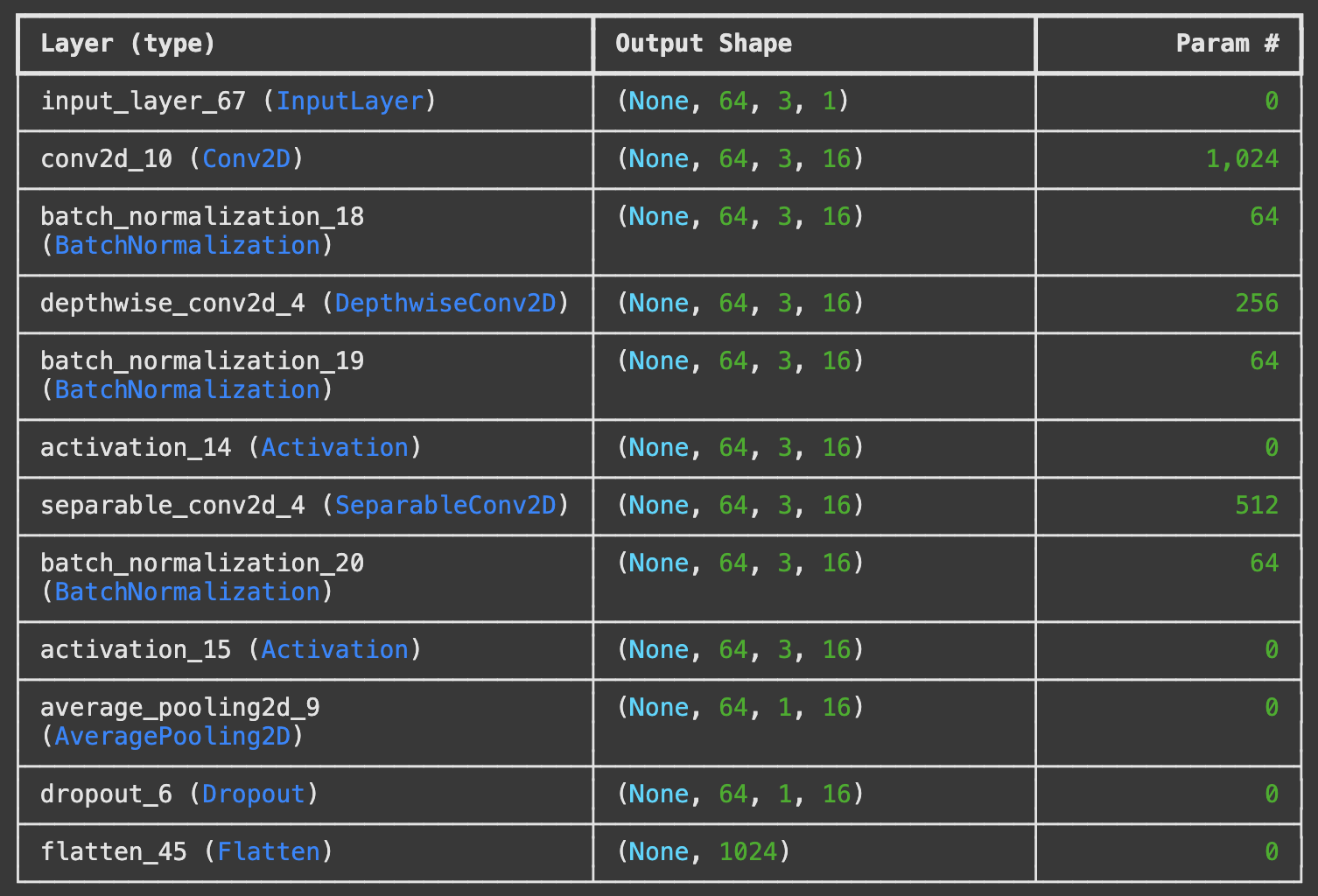


Fig.7: EEGNet model Summary

The architecture consists of several convolutional layers that learn to identify significant patterns directly from the input data. The output from a flattened layer of EEGNet serves as a learned feature representation, encapsulating essential characteristics of the EEG signals while reducing noise and redundancy.

One of the key advantages of using EEGNet is its ability to automatically learn relevant features without requiring extensive manual feature engineering. This deep learning approach allows for capturing intricate patterns that may not be easily discernible through traditional methods. By integrating these learned features into our classification framework, we enhanced our models' capacity to differentiate between cognitive states effectively.

3. Autoencoder Features

In addition to using EEGNet for direct feature extraction, we explored a novel hybrid approach that combined the powerful feature extraction capabilities of EEGNet with an autoencoder network for further refinement. This approach aimed to enhance the quality of the extracted features by reducing noise and redundancy, thereby improving the downstream classification performance. The autoencoder was trained to reconstruct the features extracted by EEGNet, leveraging its ability to compress data into a compact and informative representation.

The autoencoder comprises two main components: the encoder and the decoder. The encoder compresses the input data into a lower-dimensional bottleneck layer, which acts as the model’s core for learning a reduced but information-rich representation of the data. The decoder then reconstructs the input from this compressed representation, enabling the model to learn the most salient features while discarding less relevant or noisy details. By focusing on reconstruction accuracy, the autoencoder ensures that the essential patterns within the data are preserved.

The bottleneck layer is central to the autoencoder’s operation, as it serves as the final refined feature representation. This layer captures a compact summary of the input features while emphasizing the most significant and discriminative aspects of the data. The reduced dimensionality not only simplifies the feature space but also enhances generalization, making it easier for classifiers to identify meaningful patterns during training.

The architecture of the autoencoder was carefully designed to complement the features generated by EEGNet fig.7. It includes an input layer that matches the size of EEGNet’s flattened output (3945) and multiple convolutional layers for efficient encoding and decoding. The dense layers further refine the representation by enabling additional non-linear transformations, while the bottleneck layer provides the most compact form of the features. This multi-layered design ensures that the extracted features retain high discriminative power for classification tasks.

This approach was inspired by the work of Ameta et al., which demonstrated the effectiveness of autoencoders for improving feature quality in emotion classification tasks [1]. By incorporating a similar methodology, we aimed to refine the EEGNet features, thereby leveraging both deep learning and feature fusion techniques. Our method, fig.8 focused on minimizing redundancy and amplifying meaningful signals, which could significantly improve classification accuracy.

To evaluate the effectiveness of this hybrid approach, we extracted different feature sets, including EEG signal characteristics, PSD, deep features, and auto-encoded features. These features were fed into three classifiers—Random Forest, XGB, and KNN—for performance comparison. Among the results, the highest average accuracy for Valence was 82.10% using B features on the XGB classifier, while for Arousal, the highest accuracy was 81.33% with A features on the KNN classifier. These results highlight the potential of autoencoder-enhanced features for achieving robust classification performance.

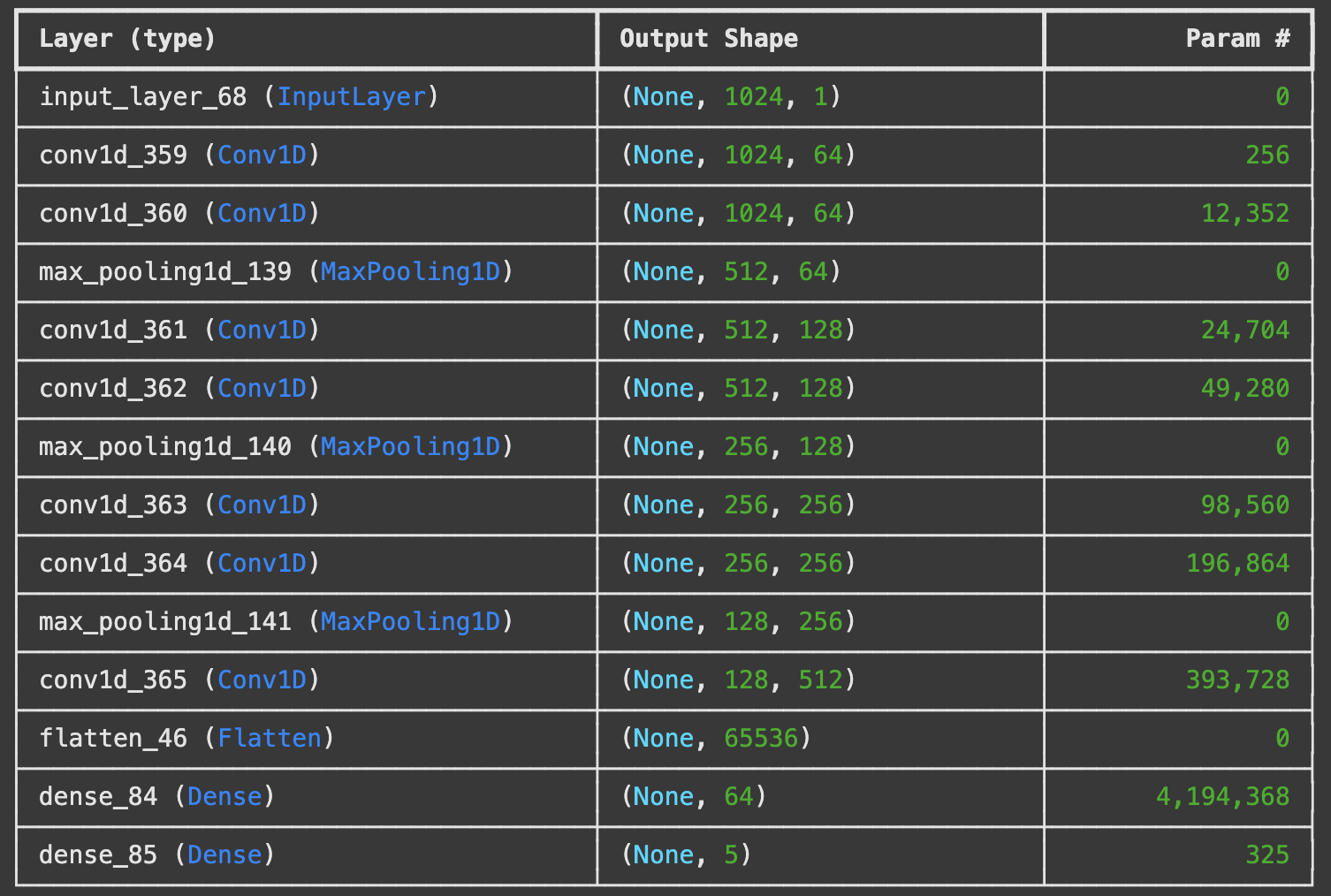


Fig.8: Autoencoder Approach

Using this fusion technique allows us to create an efficient representation that balances complexity and interpretability. The combination of learned features from EEGNet with the refined representations from the autoencoder offers a powerful way to capture intricate relationships within the EEG signals.

These fusion strategies not only improved model performance but also provided flexibility in how we approached classification tasks. As we move forward with our analysis and model training, these techniques will serve as foundational elements in our efforts to accurately classify cognitive states and understand their underlying neural correlates. The integration of diverse feature sets through these fusion methods represents a critical advancement in our ability to interpret complex EEG data effectively.

**Suggested Solution - TSception:**

The TSception model, suggested by another researcher, is a specialized neural network architecture designed for EEG signal classification, particularly effective for binary classification tasks. This model draws inspiration from the Inception architecture used in computer vision, adapting it for time series data like EEG signals. The key feature of TSception is its use of parallel temporal convolution layers with different kernel sizes (3, 5, and 7 in this implementation). This multi-scale approach allows the model to capture temporal patterns at various resolutions simultaneously. Each convolution layer uses 16 filters, enabling the extraction of diverse features from the input EEG data.

After the parallel convolutions, the features are concatenated, combining information from all temporal scales. This is followed by a MaxPooling1D layer with a large pool size of 20, which significantly reduces the temporal dimension while retaining the most salient features. The pooled features are then flattened and passed through fully connected dense layers for final classification. The architecture's strength lies in its ability to efficiently process EEG time series data by capturing both short-term and relatively longer-term temporal patterns in parallel. This multi-scale feature extraction is particularly useful for EEG data, where relevant information can exist at different time scales.

By employing this TSception model, especially for binary classification tasks, the research aims to leverage its unique temporal feature extraction capabilities to effectively distinguish between two classes of EEG signals, potentially improving classification accuracy and robustness in EEG-based brain-computer interface applications.

## **Proposed Solution - E2FNet: Spatio-Temporal FNet: A Hybrid Approach for EEG Signal Classification**

The Spatio-Temporal FNet , fig.9 represents a significant advancement in EEG signal classification by integrating spatial and temporal feature extraction pathways directly within the EEG data. This architecture addresses a key limitation in prior work, such as Jiang et al.’s E2FNet, which relied on fusing EEG and EMG features to enhance classification. Unlike E2FNet, which focused on temporal features from EEG and spatial features from EMG, Spatio-Temporal FNet achieves comprehensive feature extraction solely from EEG signals. This eliminates the dependence on additional modalities, resulting in a more streamlined and modality-independent approach.

The model architecture is built around three core components: spatial feature extraction, temporal feature extraction, and feature fusion, followed by a classification layer. The spatial feature extraction pathway employs Convolutional Neural Networks (CNNs) to analyze relationships between EEG channels. These CNNs include two convolutional layers with batch normalization, ReLU activation, and max pooling, which efficiently capture local and hierarchical spatial patterns. A global average pooling layer is incorporated to reduce spatial dimensions while preserving critical features, ensuring computational efficiency without compromising performance.

For temporal feature extraction, Spatio-Temporal FNet leverages Long Short-Term Memory (LSTM) layers, which are well-suited for modeling the sequential nature of EEG signals. The LSTM pathway captures long-term dependencies and temporal dynamics, allowing the model to recognize changes and transitions in cognitive states over time. Stacked LSTM layers further enhance the network’s capacity to model complex time-series patterns, making it particularly effective for EEG data.

The fusion of spatial and temporal features is a key innovation in this architecture. Outputs from the CNN and LSTM pathways are concatenated using the autoencoder to create a comprehensive representation of the EEG data. This fused representation combines structural spatial information with dynamic temporal characteristics, forming a robust foundation for the classification layers. The final dense layers, equipped with dropout regularization, prevent overfitting and enhance generalization. A softmax-activated output layer then classifies EEG signals into distinct cognitive states, such as *Impasse*, *Aha!*, *Re-evaluation*, and *Doing Other Task*.

This integrated approach provides several key contributions. It captures complex spatial relationships through CNNs while modeling temporal dynamics with LSTMs, ensuring a balanced analysis of EEG signals. By eliminating reliance on additional modalities like EMG, Spatio-Temporal FNet simplifies the process while maintaining high classification accuracy. Furthermore, its robust design enhances generalizability across tasks and datasets.

The Spatio-Temporal FNet demonstrates its potential for diverse EEG applications, particularly in multi-class classification tasks such as hand motion intention recognition and cognitive state monitoring. Future work could focus on improving cross-subject generalization, testing scalability on larger datasets, and exploring real-time deployment in brain-computer interface systems, paving the way for further advancements in EEG-based technologies.

Compared to Jiang et al.'s work , which separates spatial and temporal feature extraction using different data modalities (EEG for temporal and EMG for spatial), Spatio-Temporal FNet achieves comprehensive feature extraction directly from EEG data alone. This integrated approach eliminates reliance on additional modalities while maintaining high classification performance.

After experimenting with various combinations of CNNs and LSTMs, this hybrid architecture yielded the best results for five-class classification tasks. Its success demonstrates its potential for distinguishing between hand motion intentions using only EEG data. Future work could explore its scalability on larger datasets, cross-subject generalization capabilities, and real-time applications in brain-computer interfaces.

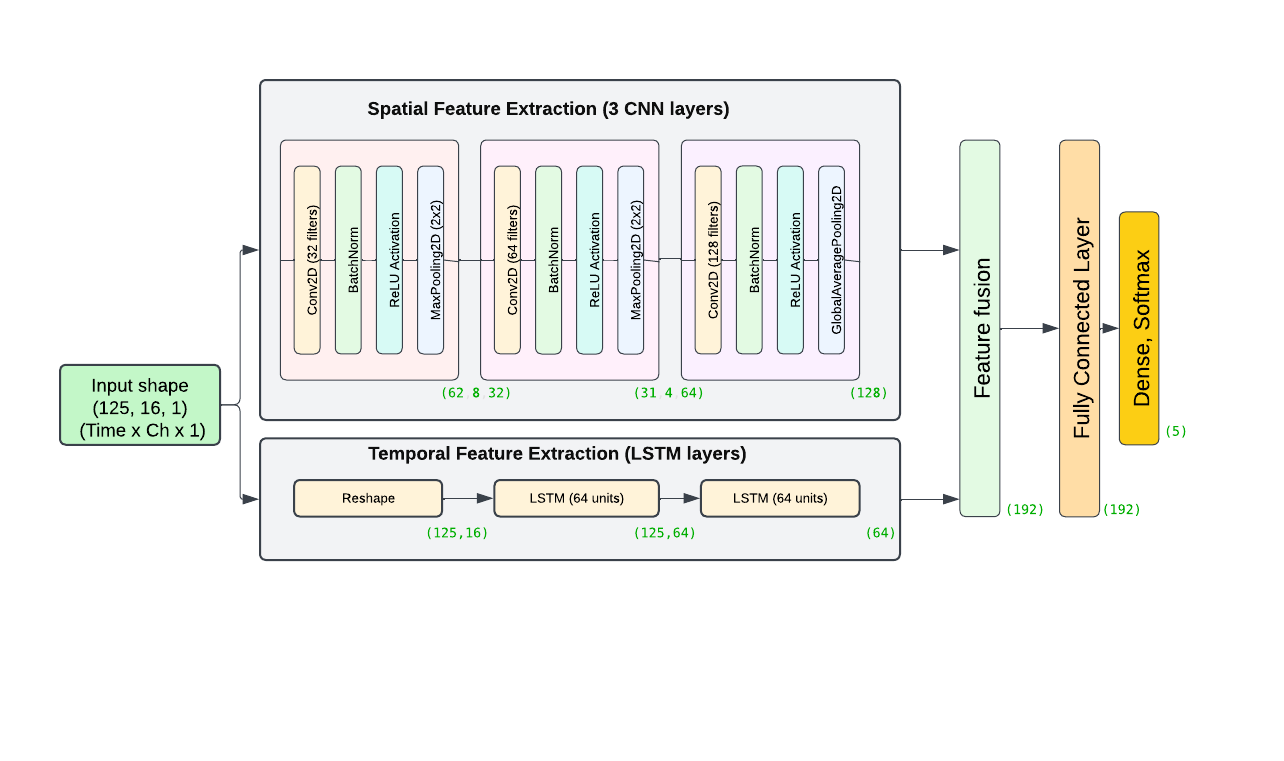


Fig.9: E2FNet: Spatio-Temporal FNet

# **Comparison Study:**

Cross validation:

The code snippet implements a stratified k-fold cross-validation strategy, which is a robust method for evaluating machine learning models, especially when dealing with imbalanced datasets. By using `StratifiedKFold` with `n\_splits=5`, the data is divided into 5 folds, ensuring that each fold maintains the same proportion of samples for each class as in the complete dataset. This approach helps mitigate bias and provides a more reliable estimate of the model's performance across different subsets of the data.

The `random\_state=42` parameter ensures reproducibility of the random splits, while `shuffle=True` randomizes the order of the samples before splitting[2]. The `for` loop iterates through each fold, creating training and testing sets (`X\_train`, `X\_test`, `y\_train`, `y\_test`) for each iteration. This method allows the model to be trained and evaluated on different subsets of the data, providing a comprehensive assessment of its generalization capabilities and reducing the risk of overfitting to a particular subset of the data.

​​from sklearn.model\_selection import StratifiedKFold

skf = StratifiedKFold(n\_splits=5, random\_state=42, shuffle=True)

for train\_index, test\_index in skf.split(data, encoded\_labels):

X\_train, X\_test = data[train\_index], data[test\_index]

y\_train, y\_test = encoded\_labels[train\_index], encoded\_labels[test\_index]

This set of X\_train, Y\_train,X\_test and Y\_test is used.

The classification of features across all four sets using various machine learning models—SVM, Random Forest, XGBoost, Gradient Boosting, KNN, and Logistic Regression—was conducted for 5 classes, each with a 1-second window.

Time features: Captured statistical properties and dynamics of EEG signals, including mean, variance, and signal changes. The best results obtained are detailed below:

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | Random Forest | 0.553633 | 0.527038 | 0.553633 | 0.537613 |
| 2 | Gradient Boosting | 0.560554 | 0.543848 | 0.560554 | 0.550425 |
| 3 | XBoost | 0.482759 | 0.453314 | 0.482759 | 0.466427 |
| 4 | Random Forest | 0.544828 | 0.533142 | 0.544828 | 0.537486 |

Frequency Features: Derived spectral characteristics like mean power, peak frequency, and frequency bandwidth from EEG signals. The best results obtained are detailed below:

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | XBoost | 0.572917 | 0.571039 | 0.572917 | 0.568508 |
| 2 | XBoost | 0.574394 | 0.559712 | 0.574394 | 0.565684 |
| 3 | XBoost | 0.551724 | 0.540166 | 0.551724 | 0.544609 |
| 4 | Gradient Boosting | 0.493103 | 0.464346 | 0.493103 | 0.475771 |

Fused Features:  
1. Concatenation of Time and Frequency Features: Combined time and frequency-domain features into a comprehensive representation to capture diverse signal aspects.

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | Random Forest | 0.570934 | 0.553388 | 0.570934 | 0.558619 |
| 2 | XGBoost | 0.565972 | 0.561388 | 0.565972 | 0.562153 |
| 3 | XGBoost | 0.5 | 0.475484 | 0.5 | 0.486508 |
| 4 | Gradient Boosting | 0.531034 | 0.514507 | 0.531034 | 0.520132 |

2. EEGNet: Extracted complex spatial and temporal patterns from raw EEG data using a specialized convolutional neural network.

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | Gradient Boosting | 0.574394 | 0.573359 | 0.574394 | 0.570075 |
| 2 | XBoost | 0.582759 | 0.577831 | 0.582759 | 0.580123 |
| 3 | Gradient Boosting | 0.570934 | 0.565367 | 0.570934 | 0.567757 |
| 4 | Gradient Boosting | 0.562069 | 0.565155 | 0.562069 | 0.559727 |

3. Based on the above results, sets 1 and 2 appear to be effective labeling methods. Therefore, the autoencoder approach was applied using the best-performing model for further analysis.

| SVM | 0.6655 | 0.675 | 0.66 | 0.6675 |
| --- | --- | --- | --- | --- |
| Random Forest | 0.7334 | 0.74 | 0.73 | 0.735 |
| XGBoost | 0.7852 | 0.79 | 0.78 | 0.785 |
| Gradient Boosting | 0.7438 | 0.75 | 0.74 | 0.745 |
| Logistic Regression | 0.741 | 0.745 | 0.735 | 0.74 |

CNN features comparison:

For 5 class classification, E2FNet showed the best result with Set 2 as Best almost in all classifiers:

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | Random Forest | 0.878472 | 0.87951 | 0.878472 | 0.878441 |
| 2 | Gradient Boosting | 0.920139 | 0.91907 | 0.920139 | 0.919141 |
| 3 | Gradient Boosting | 0.896552 | 0.898954 | 0.896552 | 0.896494 |
| 4 | Random Forest | 0.882759 | 0.882313 | 0.882759 | 0.882304 |

For 2 class classification:

TSception features for binary classification of Aha vs Not Aha, along with corresponding accuracy, precision, recall, and F1 score.

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | Gradient Boosting | 0.862069 | 0.8625 | 0.862069 | 0.862028 |
| 2 | Gradient Boosting | 0.895652 | 0.897515 | 0.895652 | 0.895494 |
| 3 | Gradient Boosting | 0.801724 | 0.802534 | 0.801724 | 0.801591 |
| 4 | Gradient Boosting | 0.862069 | 0.8625 | 0.862069 | 0.862028 |

TSception features for binary classification of Aha vs Not Aha, along with corresponding accuracy, precision, recall, and F1 score.

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | Gradient Boosting | 0.93087 | 0.931356 | 0.93087 | 0.930947 |
| 2 | Random Forest | 0.937093 | 0.937064 | 0.937093 | 0.936965 |
| 3 | Gradient Boosting | 0.944741 | 0.945444 | 0.944741 | 0.944533 |
| 4 | KNN | 0.926724 | 0.927598 | 0.926724 | 0.926279 |

E2FNet with Aha vs Not Aha: the Fusion of Spatio-Temporal features is highly effective for distinguishing between "Aha" and "Not Aha" states

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | SVM/Random  Forest/XGBoost/Lo  gistic Regression | 0.938596 | 0.945419 | 0.938596 | 0.938431 |
| 2 | SVM | 0.939655 | 0.946154 | 0.939655 | 0.939435 |
| 3 | SVM | 0.921053 | 0.931818 | 0.921053 | 0.920557 |
| 4 | Gradient Boosting | 0.920139 | 0.91907 | 0.920139 | 0.919141 |

E2FNet with Impasse vs Not Impasse: the Fusion of Spatio-Temporal features is highly effective for distinguishing between "Impasse" and "Not Impasse" states:

| Set Number | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- |
| 1 | Random  Forest/XGBoost/Gr  adient Boosting | 0.982456 | 0.983061 | 0.982456 | 0.982456 |
| 2 | All except Fully Connected Layer and Logistic  Regression | 0.982456 | 0.983051 | 0.982456 | 0.982451 |
| 3 | All classifiers | 0.974138 | 0.97541 | 0.974138 | 0.974121 |
| 4 | XBoost | 0.965517 | 0.965517 | 0.965517 | 0.965517 |

Study:  
Based on the results, sets 1 and 2 demonstrated better performance compared to sets 3 and 4 labels.

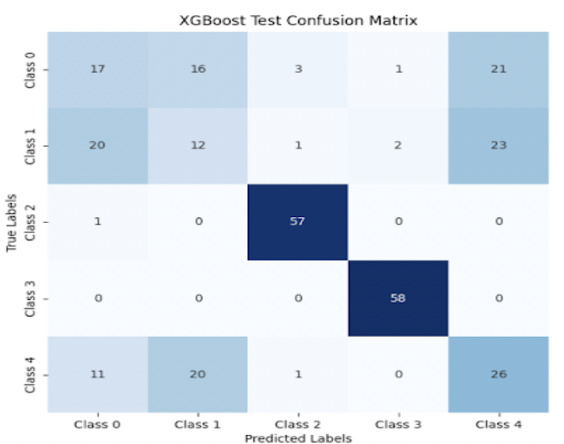
Comparing the models’ performance across 5-class classification in set 2 for further highlights their relative effectiveness.

| Number of classes | Window Size | Method | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | 1s | Time Features | Gradient  Boosting | 0.560554 | 0.543848 | 0.560554 | 0.550425 |
| 5 | 1s | Frequency  Features | XBoost | 0.574394 | 0.559712 | 0.574394 | 0.565684 |
| 5 | 1s | Time +  Frequency  Features  Concatenation | Random Forest | 0.570934 | 0.553388 | 0.570934 | 0.558619 |
| 5 | 1s | EEGNet Features | Gradient  Boosting | 0.582759 | 0.577831 | 0.582759 | 0.580123 |
| 5 | 1s | Autoencoder  Approach | XBoost | 0.7852 | 0.79 | 0.78 | 0.785 |
| 5 | 1s | E2FNeT | Gradient Boosting | 0.920139 | 0.91907 | 0.920139 | 0.919141 |

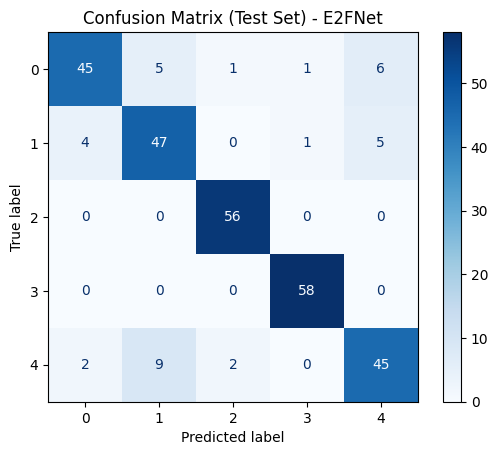
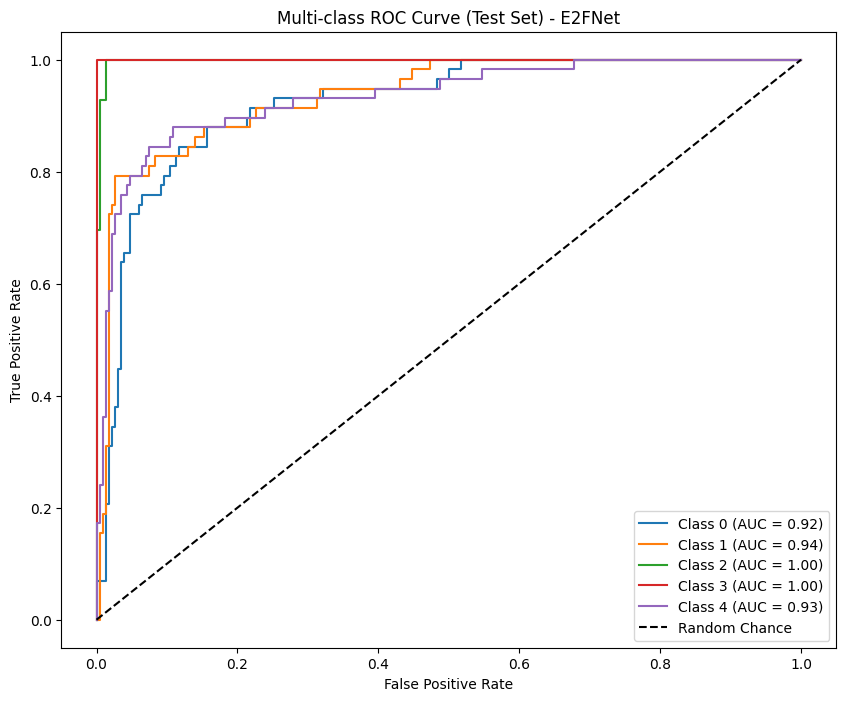
Similarly for 2-class classification on set 2 and 1:

| Method | Classification Task | Best Classifier | Best Accuracy | Best Precision | Best Recall | Best F1 Score |
| --- | --- | --- | --- | --- | --- | --- |
| TSception | Aha vs Not Aha | Gradient Boosting | 0.895652 | 0.897515 | 0.895652 | 0.895494 |
| TSception | Impasse vs Not Impasse | Gradient  Boosting | 0.944741 | 0.945444 | 0.944741 | 0.944533 |
| E2FNet | Aha vs Not Aha | SVM | 0.939655 | 0.946154 | 0.939655 | 0.939435 |
| E2FNet | Impasse vs Not  Impasse | Random  Forest/XGBoo  st/Gradient  Boosting | 0.982456 | 0.983061 | 0.982456 | 0.982456 |

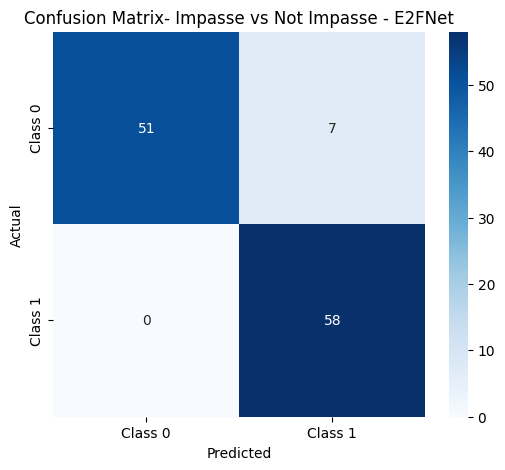
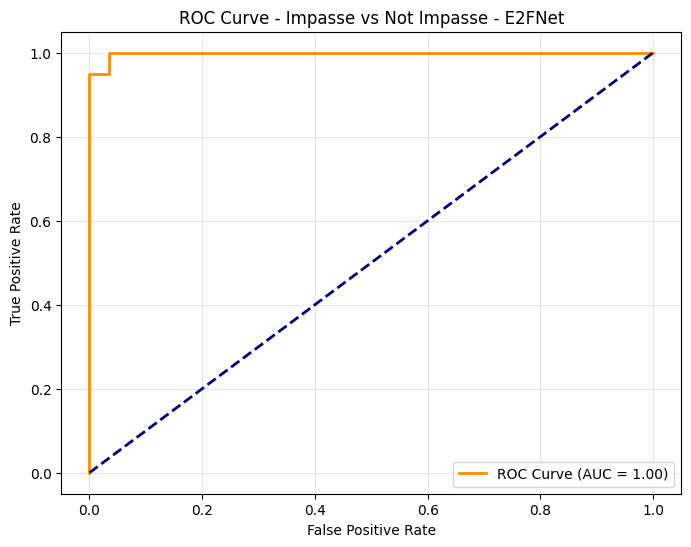
RoC Curve and Matrix:

Autoencoder: 

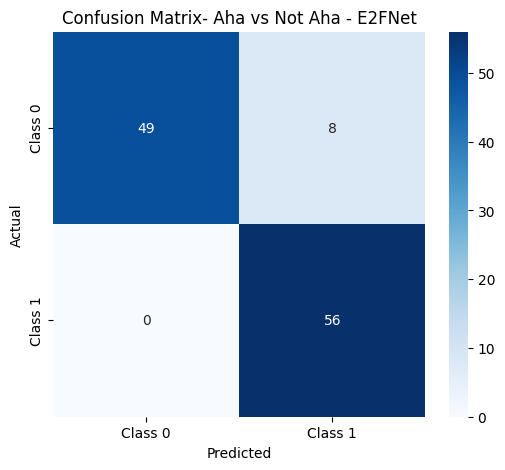
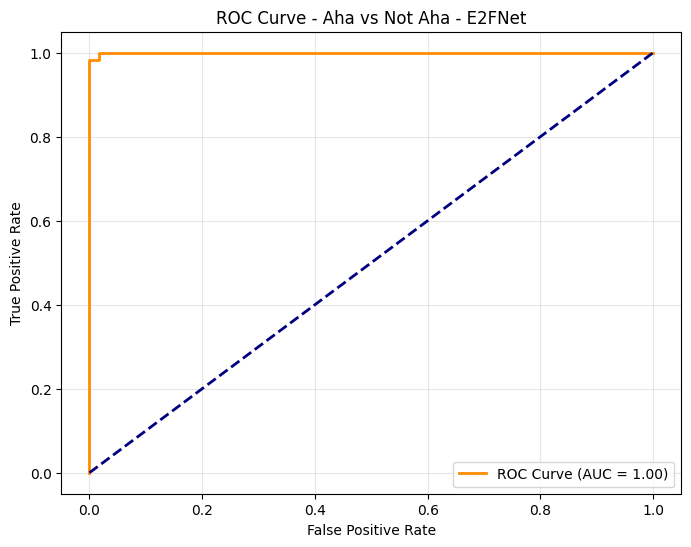
RoC and Confusion Matrix for the best classifier - E2FNet 5 class



RoC and Confusion Matrix for the best classifier - E2FNet 2 class - impasse vs Not Impasse



RoC and Confusion Matrix for the best classifier - E2FNet 2 class - Aha vs Not Aha



Conclusion:  
From the results, it is evident that summary features, when classified individually (e.g., time features, frequency features, or their concatenation), yield relatively lower accuracies. However, when fusion methods are employed—especially autoencoder-based fusion—the classification accuracy becomes significantly more appreciable. This highlights the importance of combining diverse feature sets and leveraging deep learning-based feature fusion techniques to achieve better performance in EEG data classification.

Furthermore, comparing the E2FNet and TSception models, it is clear that E2FNet demonstrates superior performance in both 5-class and 2-class classification tasks. Notably:

In 5-class classification, E2FNet achieves the highest accuracy of 0.9201 with Gradient Boosting, surpassing other methods like EEGNet and traditional feature concatenation approaches.

In 2-class classification tasks, E2FNet also outperforms TSception consistently. For instance:

* For “Aha vs Not Aha,” E2FNet paired with SVM achieves an accuracy of 0.9397, which is higher than TSception’s best result of 0.8957.
* For “Impasse vs Not Impasse,” E2FNet with Random Forest, XGBoost, or Gradient Boosting attains an impressive accuracy of 0.9825, significantly better than TSception’s 0.9447.

These findings suggest that E2FNet not only excels in leveraging fused feature representations for multi-class scenarios but also proves to be more effective than TSception for binary classification tasks. Therefore, E2FNet is a strong candidate for robust and accurate EEG data classification.

Future Work

In future iterations of this study, several improvements can be made to enhance the robustness and reliability of our results:

1. Subject-Independent Analysis: To address potential data leakage and improve generalizability, we should ensure that one subject's data does not overlap between the training and testing sets. This can be achieved by implementing a leave-one-subject-out cross-validation strategy, where models are trained on data from all subjects except one, and then tested on the held-out subject's data.

1. Combine physiological signals and EEG signals
2. Consider delay while clicking.

By incorporating these improvements, we can increase the validity of our results and gain more confidence in the generalizability of our cognitive state classification models to new subjects and scenarios.

References

[1] A. S. Chakkamallisery, S. T. Pelmo, T. Angsuwatanakul, Y. Pititheeraphab, T. Puttasakul, and T. Khemanuwong, “Mind to Motion: EEG-Based Classification of Motor Imagery and Actual Hand Movements Using LSTM Models,” *2023 15th Biomedical Engineering International Conference (BMEiCON)*, Tokyo, Japan, 2023, pp. 1-5, doi: 10.1109/BMEiCON60347.2023.10322025.

[2] Y. Narayan, “Hand Motion Identification Based on EEG Signals Classification,” *2021 2nd Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, 2021, pp. 1-7, doi: 10.1109/GCAT52182.2021.9587556.

[3] L. Chen and Y. Niu, “EEG Motion Classification Combining Graph Convolutional Network and Self-attention,” *2023 International Conference on Intelligent Supercomputing and BioPharma (ISBP)*, Zhuhai, China, 2023, pp. 38-41, doi: 10.1109/ISBP57705.2023.10061298.

[4] E. Popov and S. Fomenkov, “Classification of hand motions in EEG signals using recurrent neural networks,” *2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, Chelyabinsk, Russia, 2016, pp. 1-4, doi: 10.1109/ICIEAM.2016.7911620.

[5] L. Zhou, Q. Zhu, B. Wu, Q. Bing, and Z. Qian, “Classification of four-class motion imagination tasks based on EEG by combining EEG source imaging with convolution neural networks,” *2021 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, Hefei, China, 2021, pp. 1-4, doi: 10.1109/ICMIPE53131.2021.9698943.

[6] Liu, Xuanchang, and Ivan Mutis. "Chapter Cognitive Dynamics for Construction Management Learning Tasks in Mixed Reality Environments." (2023).

[7] Kyriaki, Konstantina, Dimitrios Koukopoulos, and Christos A. Fidas. "A Comprehensive Survey of EEG Preprocessing Methods for Cognitive Load Assessment." IEEE Access (2024).

[8] Lawhern, Vernon J., et al. "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces." *Journal of neural engineering* 15.5 (2018): 056013.

[9]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.

[10] Ameta, Durgesh, et al. “Comparative Evaluation of EEG Feature Fusion Techniques for Emotion Classification.” *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024.

[11]G. Jiang, K. Wang, Q. He and P. Xie, "E2FNet: An EEG- and EMG-Based Fusion Network for Hand Motion Intention Recognition," in \*IEEE Sensors Journal\*, vol. 24, no. 22, pp. 38417-38428, 15 Nov. 2024, doi: 10.1109/JSEN.2024.3471894.

[12] H. Chao and L. Dong, “Emotion Recognition Using Three-Dimensional Feature and Convolutional Neural Network from Multichannel EEG Signals,” in *IEEE Sensors Journal*, vol. 21, no. 2, pp. 2024-2034, 15 Jan.15, 2021, doi: 10.1109/JSEN.2020.3020828.

[13] Abdulghani, Mokhles M., Wilbur L. Walters, and Khalid H. Abed. “Enhancing the classification accuracy of EEG-Informed Inner Speech Decoder Using Multi-Wavelet Feature and Support Vector Machine.” *IEEE Access* (2024).