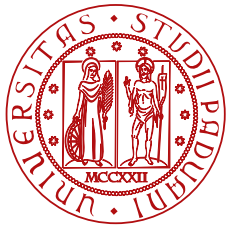


Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master Thesis in ICT FOR INTERNET AND MULTIMEDIA

Mitigating learning biases: A study on uncertainty quantification and instantial variability in machine learning models

Master Candidate

Alda Kola

Student ID 2071530

Supervisor

Prof. Roberto Corvaja

University of Padova

Co-supervisor

Prof. Giulia Cisotto

University of Trieste

Academic Year: 2023/2024

Graduation Date: December, 4th 2024

To my family.

To my parents, Valentin and Dila, to all their sacrifices.

To my sister Florinda, , my absolute favorite person, who has always been my light.

Abstract

Machine learning (ML) has become a significant driver of advancement in healthcare, enabling advanced solutions for diagnostics, personalized treatments, and decision-making support. In the study of brain dynamics, EEG microstate analysis stands out as a valuable technique for exploring brain activity, offering insights into cognitive functions and neural processes. However, inherent data variability and model limitations introduce uncertainty, posing significant challenges to the reliability of ML applications in this domain.

This thesis explores the impact of instantial variability on the accuracy and robustness of machine learning (ML) models in EEG microstate analysis. EEG microstates, brief patterns of brain activity, are crucial for understanding neural dynamics. Using a dataset of resting-state EEG recordings from 203 participants, key microstate features such as Global Explained Variance, Mean Durations, and Corrected Time Coverage were analyzed.

To simulate variability, probabilistic augmentation techniques were applied on the dataset and uncertainty-aware methods were used to classify microstates. Four classifiers: K-Nearest Neighbors (KNN), Augmented Support Vector Classifier (ACS), Augmented Gradient Boosting Classifier (ACG), and Weighted Sampling Forest (WSF) were evaluated under baseline and perturbed conditions and compared with the performance of a traditional ML model, Linear Support Vector Machine (LSVM). The ACS model consistently showed the highest performance, demonstrating the effectiveness of augmentation and uncertainty quantification in enhancing robustness. To further evaluate the robustness of these classifiers, perturbations were introduced to simulate real-world variability.

The findings emphasize the importance of variability-aware techniques in improving ML models for EEG analysis, paving the way for more reliable applications in clinical diagnostics and brain-computer interfaces. Future work should focus on expanding datasets, exploring deep learning approaches, and adapting methods to real-time applications.

Contents

List of Figures

List of Tables

List of Acronyms

UQ Uncertainty Quantification

IV Instantial Variability

MS Microstates

EEG Electroencephalography

EO Eyes-Opened

EC Eyes-closed

AV Analytical variation

BV Biological Variation

ML Machine Learning

Gev Global Explained Variance

KNN K-Nearest Neighbors

AUC Area Under the Curve

ACC Accuracy

LSVM Linear Support Vector Machine

CI Confidence Interval

BCIs Brain Computer Interfaces

1

Introduction

Machine learning (ML) has become a powerful and transformative tool across various domains, with its applications in healthcare proving to be particularly impactful. From diagnostic tools to personalized treatment recommendations, ML has demonstrated its potential to enhance decision-making and improve patient outcomes. In critical areas such as brain-computer interfaces (BCIs) and intensive care units (ICUs), ML models are being increasingly deployed to interpret complex biosignals like electroencephalography (EEG), providing real-time insights for clinical interventions [2][35]. However, translating ML advancements into real-world healthcare settings requires addressing challenges associated with data variability, model reliability, and uncertainty management [5][35].

Two primary types of uncertainty that affect ML predictions are: aleatoric uncertainty, which stems from inherent noise or variability in data, and epistemic uncertainty, arising from limitations in the model itself, such as insufficient training data or model misspecification [2][1]. Aleatoric uncertainty is common in EEG data due to factors such as equipment differences, environmental conditions, and subject-specific variability, while epistemic uncertainty can manifest when a model is applied to out-of-distribution data or unseen scenarios [5][35].

EEG, a non-invasive and cost-effective tool, provides high temporal resolution for monitoring brain activity [27]. It is widely used in neuroscience and healthcare, particularly for diagnosing neurological disorders and monitoring cognitive states. Despite its advantages, EEG data is inherently noisy and exhibits significant variability across and within subjects [13][10]. This variability

poses challenges for ML models, especially when these models are expected to generalize across diverse populations and experimental conditions [38][11].

A critical aspect of EEG data variability is instantial variability (IV), that has yet to be tested, captures within-subject differences arising from biological processes or measurement inconsistencies [7][35]. Unlike variability between subjects or populations, IV represents intrinsic patterns specific to an individual and can fluctuate based on physiological or psychological states [10]. IV poses a significant challenge for ML models, as it may lead to reduced robustness and inconsistent predictions even in controlled conditions [7][35]. Hence, addressing IV is essential for developing reliable ML frameworks, particularly for clinical applications.

EEG microstate analysis offers a useful case study to investigate the impact of IV on the reliability of ML models [18]. Microstates are brief, quasi-stable patterns of brain activity that last approximately 40-100 milliseconds [28]. These patterns reflect distinct functional states of the brain and can provide valuable insights into cognitive and neurological processes [28][10]. Metrics such as global explained variance (GEV), mean durations (MeanDurs), and corrected time coverage (TimeCov_corrected) are commonly used to characterize these microstates. However, these metrics are sensitive to both inter- and intra-subject variability [16].

This thesis work investigates the impact of instantial variability on the classification accuracy and robustness of ML models applied to EEG microstate analysis. To address the challenges associated with variability and uncertainty, a systematic computational framework was developed.

This study utilizes an open-source EEG dataset comprising resting-state recordings from 203 healthy participants across two conditions: eyes-open (EO) and eyes-closed (EC). The dataset includes detailed demographic information and was preprocessed to ensure high data quality. A key aspect, the variability in EEG microstate features was introduced using probabilistic representations such as Gaussian fuzzy labels and random labels. Hellinger and Mahalanobis distances were employed to quantify the impact of variability on data distributions. Four classifiers were used to address variability and uncertainty: Imprecise KNeighborsClassifier (KND-f), Augmented Support Vector Classifier (ACS-a), Augmented Gradient Boosting Classifier (ACG-a), and Weighted Resampling Forest (WSF-i). These models were designed to incorporate probabilistic augmentation and variability-aware techniques, and they showed more robustness

under perturbations compared to the traditional ML model Linear Support Vector Machine (LSVM). The robustness of the models was assessed using metrics such as Area Under the Curve (AUC), accuracy, and F1 score under both baseline and perturbed conditions. Stratified K-Fold cross-validation was employed to ensure reliable performance evaluation. Key findings include the effectiveness of variability-aware methods in improving model robustness and generalization. Among the tested classifiers, the ACS-a demonstrated the highest performance across all metrics under both baseline and perturbed conditions, highlighting the importance of integrating data augmentation and probabilistic methods in ML workflows for EEG analysis.

Through these objectives, this thesis work aims to address the research question:*"How does instantial variability influence the classification accuracy and robustness of a model?"* and contribute to the development of more reliable machine learning frameworks for clinical and neuroscientific applications.

The structure of this thesis work is organized as followed:

Chapter 2 - Background: This chapter provides a detailed overview of the foundational concepts in uncertainty quantification (UQ) and its relevance to machine learning and EEG analysis. It explains aleatoric and epistemic uncertainties, the importance of instantial variability (IV) in EEG research, and the challenges posed by cross-subject variability. Additionally, it introduces a comprehensive review of the current state of the art for addressing variability and uncertainty.

Chapter 3 - Methods: This chapter outlines the methodology used in this study. It starts with the dataset description and the preprocessing steps. It explains the variability simulation techniques, uncertainty quantification methods, and the four classifiers employed. The evaluation framework and perturbation analysis are also detailed to assess model robustness under real-world variability.

Chapter 4 - Results and Discussions: This chapter presents the results of the experiments, insights on the performance and a discussion on the impact of these results.

Chapter 5 - Conclusions and Future Works: This chapter summarizes the key findings. It highlights the strengths of the proposed methods, limitations and suggests future research directions.

2

Background

In machine learning (ML), particularly in applications such as healthcare, uncertainty quantification (UQ) plays a crucial role in evaluating the confidence of models' predictions. This is especially important when working with complex biosignals, where models must address both the inherent randomness in the data and the variability between subjects.[22].

Uncertainty in ML can be broadly classified into two primary types: aleatoric and epistemic uncertainty [2]. Aleatoric uncertainty, often referred to as data uncertainty, stems from the inherent variability in the data, including factors such as noise, multi-modality, and complexity [17]. This type of uncertainty is intrinsic to the data distribution and cannot be mitigated by enhancing the model, as it reflects the natural randomness or complexity present in the signal, such as noise in EEG recordings [5]. An example in healthcare is the variability introduced by differing equipment or protocols across clinical settings. [35].

Epistemic uncertainty, also known as knowledge uncertainty, is associated with the limitations of the model itself. It arises from factors such as insufficient or biased training data, model misspecification, or changes in the environment where the model is deployed, commonly referred to as concept drift [5]. Unlike aleatoric uncertainty, epistemic uncertainty can be reduced by increasing the diversity and quantity of training data, improving the model architecture, or adapting the model to new contexts [22].

In ML, aleatoric and epistemic uncertainty are often addressed independently. Predictive uncertainty is sometimes modeled by quantifying these two types separately, enabling researchers to evaluate the model's confidence in each

prediction. [17]. This is especially useful in clinical applications, where models deployed in real-world environments must perform robustly under diverse conditions, such as varying noise levels or unrepresentative data samples [2][35].

We hypothesize that in EEG research, UQ is essential for managing cross-subject variability and ensuring that models trained on limited or specific data can reliably interpret signals from new subjects. This is a crucial step for model robustness, as the underlying EEG signals may differ due to factors like the recording equipment, annotation variability, and subject-specific patterns [12].

On the other hand, recent research has been focused also on microstates [10][16]. EEG microstates represent brief, quasi-stable patterns of electrical brain activity that can reveal insights into brain dynamics, which can be seen in Figure 2.1 [3].

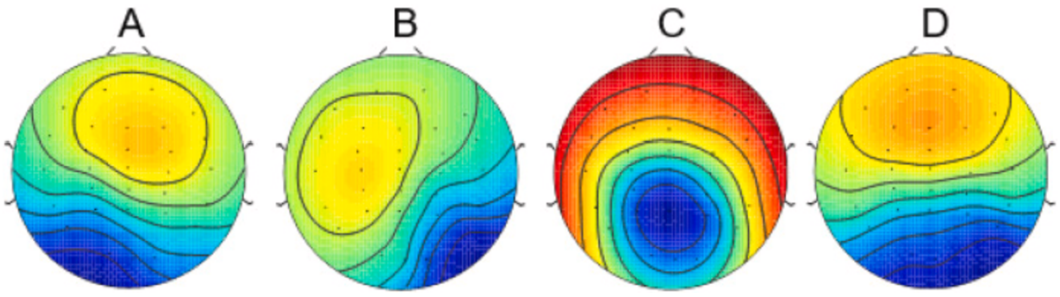


Figure 2.1: Topographical maps representing EEG microstates (A, B, C, D), illustrating the spatial patterns of electrical brain activity associated with distinct microstate classes [3].

These microstates, typically lasting between 40 to 100 milliseconds, reflect transient but stable topographical maps on the scalp, indicating synchronous neural activity across specific brain regions [16]. Microstate analysis, leverages EEG's high temporal resolution to segment brain activity into discrete, sequential microstates. Each microstate is thought to correspond to a distinct functional state of the brain, enabling researchers to capture rapid transitions in neural processes associated with various cognitive or sensory tasks, resting states, or sleep stages [10]. EEG microstate features:

- Gev (Global Explained Variance): Reflects the explanatory power of a microstate.
- MeanDurs (Mean Durations): The average duration of microstate segments.

- TimeCov_corrected (Corrected Time Coverage): The fraction of time covered by a specific microstate [16].

In this study, we hypothethize that microstates are particularly valuable in understanding both intra- and inter-subject brain variability. Intra-subject variability refers to changes in microstate patterns within the same individual under different functional conditions [31], such as sleep stages[6] or task demands. This variability captures how the brain dynamically shifts between different functional states based on internal or external stimuli [27]. Inter-subject variability, on the other hand, reflects differences in microstate patterns between individuals, which may be linked to unique cognitive or behavioral characteristics [27].

Campagner et al. (2023) have proposed IV to study data robustness and model reliability. Instantial variability refers to within-subject variability that stems from the inherent characteristics of an individuals biological processes or the measurement process itself, rather than from population differences or errors [7]. IV is often seen in the form of biological variation (BV)the natural distribution of a subjects physiological features and analytical variation (AV), which relates to fluctuations introduced by the measurement process or instrumentation.

IV could be studied in EEG signals, taken microstate classification as a case study, where the recorded signal can vary considerably due to intrinsic biological factors or technical factors, which could impact the robustness on the ML model.

The importance of IV in clinical data has been recognized, as it represents one of the most significant sources of uncertainty in medical data analysis [7]. Recent studies emphasize the need to incorporate IV into analytical frameworks, as it reflects essential individual differences that can impact data reliability and model robustness. By adequately accounting for pre-analytical, analytical, and within-subject biological variations, researchers can enhance model performance, ensuring more consistent and reliable outcomes across different subjects or conditions [7].

2.1 Uncertainty Quantification Methods

The goal of uncertainty quantification (UQ) techniques has been to quantify the reliability of machine learning (ML) models in real-world applications. Monte Carlo Dropout and Bayesian Neural Networks (BNNs) are two important

techniques for managing dataset shifts, improving prediction reliability, and calculating aleatoric and epistemic uncertainty [35][5][2]. Through the identification of model variety and the optimization of performance under distribution shifts, ensemble methods and adaptive data augmentation strategies, like the Uncertainty Estimation and Reduction Model (UNCER) framework, improve robustness [5][13]. Domain-guided transformations and latent space evaluations identify performance degradation in EEG models, while appropriate scoring rules such as Continuous Ranked Probability Score(CRPS) and quantile loss functions guarantee calibrated probabilistic forecasts [35][34]. Regardless of recent developments, the need for unified, effective frameworks for uncertainty estimates is reinforced by the difficulties in scaling these approaches for varied, high-dimensional data [35][5][13].

2.2 Instantial Variability

Current research emphasizes the importance of intra- and inter-subject variations as well as instantial variability in comprehending neural and biological systems. Variability in Stereo Electroencephalographic(SEEG) responses and task-induced EEG changes are used to improve diagnostic tools, such as identifying functional connectivity in epilepsy [21] and classifying mild cognitive impairment (MCI) with high accuracy [33]. These methods turn variability into a valuable diagnostic markers. Frameworks like Biological Variation Data Critical Appraisal Checklist (BIVAC) [1] ensure trustworthy estimates of within-subject (CVI) and between-subject (CVG) variability, increasing clinical use and reducing inconsistencies in study design and data reporting [1][30]. EEG reliability metrics analyses reveal stable intra-subject patterns but significant inter-subject differences, informing personalized interventions in both neuroscience and rehabilitation [27][39]. Resting-state EEG (rsEEG) research highlights the importance of considering external factors, such as time of day and pre-recording physical activity, to ensure accurate interpretation of metrics. Taking in consideration these factors demonstrates based on evidence, that time of day can significantly impact EEG measurements, with beta and gamma power increasing during afternoon sessions, likely reflecting circadian rhythms. Addressing these factors is important for reliable interpretations [36]. Brain-Computer Interfaces (BCIs) are systems that enable direct communication between the brain and external devices. These systems focus on addressing variability and im-

proving generalization across subjects. [36][26]. Another important aspect in this area of research is the role of microstate metrics, which are influenced by alpha rhythms. These metrics emphasize the importance of microstates in regulating neural networks, such as the default mode network (DMN) [10]. Instantial Variability (IV) is a critical source of uncertainty in clinical data analysis, reflecting variations within individual subjects that can influence the interpretation of outcomes. IV is made up of three key components: pre-analytical variation, analytical variation (AV), and biological variation (BV) [7].

Pre-analytical variation arises from factors such as patient preparation (e.g., physical activity, medication use or fasting) and sample handling (e.g., collection, storage, transport,). Even though this variation can be minimized (e.g., through standardized laboratory practices), it still plays a role in the uncertainty [7]. **Analytical variation (AV)** refers to the inherent uncertainty of measurement techniques, characterized by [7]:

- Random components (variance): Reflecting consistency between repeated measurements with the same instrument.
- Systematic components (bias): Highlighting discrepancies in values reported by different instruments.

The following definitions and methods described for AV and BV are based on the work of Campagner et al. (2023). AV is defined as:

$AV_p(x_i) = \frac{1}{m} \sum_{t=1}^{m} \text{StDev}(x_{pi}(t))$,

Biological variation (BV) refers to the natural fluctuations in a subject's biomarkers or features over time, independent of other variation sources. BV it is computed as $BV_p(x_i) = \sqrt{IV_p(x_i)^2 - AV_p(x_i)^2}$. To quantify IV, which is computed as $IV_p(x_i) = \text{StDev}(x_{pi})$, referenced individuals that are part of the experimental study are monitored over multiple time points. At each time step, the repeated measurements performed are done to isolate the AV component. The total IV is then calculated as the standard deviation of all feature values, with BV derived as the residual variability after removing AV. These values are typically expressed as coefficients of variation (e.g., CVT, CVA, CVI) relative to the features mean [7]. As reported in previous studies [7], the following metrics are computed:

- CVA (Analytical Variability): Variability arising from measurement or analytical errors.
- CVI(s) (Within-Subject Variability): Variability of a feature within the same subject, capturing intra-individual differences.

- CVI(m) (Between-Subject Variability): Variability of a feature across different subjects, representing inter-individual differences.
- CVI(ma) (Total Variability): A composite metric combining the above three sources of variability into a single measure.

$$CVT(x_i) = \frac{IV(x_i)}{\bar{x}_i}, \quad CVA(x_i) = \frac{AV(x_i)}{\bar{x}_i}, \quad \text{and} \quad CVI(x_i) = \frac{BV(x_i)}{\bar{x}_i},$$

where $\bar{x}_i$ represents the average value of x_i across all patients and time steps. These coefficients facilitate modeling of uncertainty for individual patients as a d-dimensional Gaussian distribution, $N_p(\hat{x}_p, \Sigma_p)$, where $\hat{x}_p$ is the patients homeostatic point and Σ_p is a diagonal covariance matrix with elements $\Sigma_{p,i,i} = CVT(x_i) \cdot \hat{x}_{p,i}$ [7].

These findings demonstrate understanding and managing variability improves robustness, personalization, and clinical relevance across different research domains.

2.3 EEG Microstate Analysis

The advancement of EEG microstate analysis has enhanced its application in clinical diagnosis [20]. By optimizing across dynamic data distributions, robust decoding frameworks like Distributionally Robust Optimization (DRO) enhance generalization across subjects and conditions [8], facilitating applications in Brain-Computer Interfaces (BCIs)[12]. Alpha-band oscillation-modulated microstate measures highlight their role as biomarkers for cognitive processes by reflecting interactions with neural networks, such as the default mode network (DMN) [10]. Changes in microstate dynamics, including duration, coverage, and occurrence, particularly in microstates B and D, are influenced by alpha-band power, emphasizing the role of alpha oscillations in shaping microstate behavior and their potential in understanding functional brain networks. [10]. Microstate analysis is further improved by narrowband decomposition, which demonstrates frequency-specific patterns that perform better in clinical predictions and behavioral state classification than conventional broadband techniques [16]. Last but not least, a data-driven methodology advances reproducibility and facilitates cross-research integration by enhancing microstate study comparability through the use of geographic similarity analysis and meta-microstate maps

[24]. Together, these techniques boost resilience, and increase the usefulness of EEG microstate analysis in both clinical and scientific contexts.

3

Methods

In this chapter it is outlined the implemented methodology to assess variability, uncertainty, and robustness in machine learning models applied to EEG microstate data. The primary goal of the analysis is to model and quantify different sources of variability and incorporate these into machine learning workflows to evaluate their performance under realistic perturbations. The methods described are designed to address the challenges of uncertainty that are present in clinical data.

To achieve this, a computational framework was developed that integrates probabilistic augmentation, advanced classification techniques, and statistical evaluation. Key features of EEG microstates included Global Explained Variance (GEV), Mean Durations (MeanDurs), and Corrected Time Coverage (TimeCov_corrected). Variability in these features was introduced using probabilistic representations like GaussianFuzzyLabel and RandomLabel, and uncertainty was assessed using metrics such as Hellinger and Mahalanobis distances.

This chapter details the framework for simulating variability, training models, and evaluating their performance under uncertainty. The methodologies described here form the basis for understanding the interplay between variability, uncertainty, and machine learning in EEG microstate classification, offering insights into improving model robustness and generalizability.

3.1 Dataset Description

The dataset used in this thesis is derived from an open-source, validated, repository of resting-state electroencephalography (EEG) recordings [4]. It offers a strong basis for examining brain dynamics since it contains data from 203 healthy participants in both eyes-open (EO) and eyes-closed (EC) settings. In order to participate in the Mind-Brain-Body study, participants had to meet strict inclusion requirements and ethical guidelines. In-depth demographic data comprises two cohorts: Younger adults: Age range 20-35 years (N = 153; 45 females; mean age = 25.1 ± 3.1 years). Older adults: Age range 59-77 years (N = 74; 37 females; mean age = 67.6 ± 4.7 years). Detailed protocol, inclusion criteria, and preprocessing steps are described in more detail in the literature [4]. Figure 3.1 shows the Power Spectral Density of the EEG signal across frequencies, using all channels. The topographical maps of synchronized neural activity of the microstates, can be seen in Figure 3.2 [16].

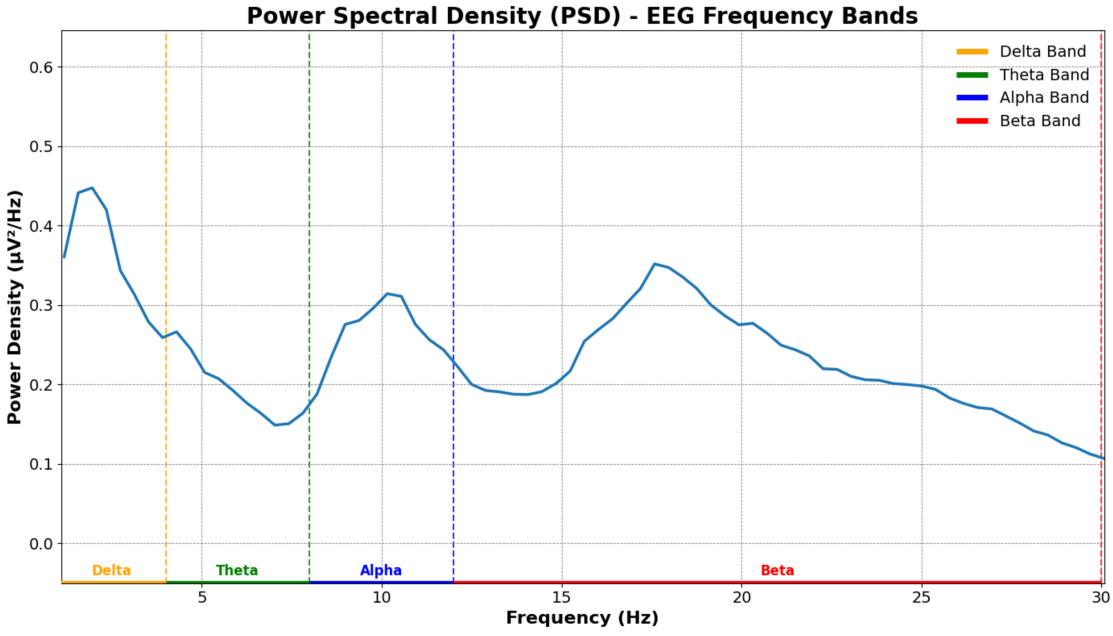


Figure 3.1: Power Spectral Density of the EEG Frequency Bands

3.1.1 Preprocessing

Comprehensive preprocessing was applied by the experts who made the dataset available to ensure consistency and quality of the data. The EEG signal

was resampled at 250 Hz and filtered between 1 and 45 Hz with an eighth-order Butterworth filter. Outlying channels identified and removed after visual inspection, and noise/artifact segments were excluded [4]. There was implemented the Independent Component Analysis (ICA) using the Infomax algorithm, which removed components related to eye blinks, movements, and cardiac artifacts [4]. Dimensionality Reduction was performed with Principal Component Analysis (PCA), which retained components explaining 95% of variance. EEG data were filtered into five traditional frequency bandsbroadband (1-30 Hz), delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), and beta (15-30 Hz)using zero-phase, non-causal bandpass finite-impulse response filters [4].

A Hamming window was employed for all filters to minimize border effects, ensuring optimal passband ripple and stopband attenuation [4]. Final preprocessing steps performed, included channel interpolation of missing/bad channels using spherical spline methods. Data was then re-referenced to the average signal across electrodes, further resampled to 100 Hz for compatibility with microstate analysis [16]. Given its high temporal resolution and comprehensive preprocessing, this dataset serves as the foundation for examining EEG microstate dynamics and variability under various settings. Reliable features are then extracted for further modeling and analysis. [16]

3.1.2 Microstates and Their Classification

Here we present microstates and their classification with the baseline model (linear SVM) without any perturbations. This study used EEG microstate segmentation to explore brain activity patterns across different frequency bands (broadband, delta, theta, alpha, beta) and behavioral conditions (eyes-open EO and eyes-closed EC). Microstates, which are brief and stable topographical maps of synchronized neural activity. The identified microstates were labeled (A, B, C, D, C') based on commonly recognized spatial patterns [16]: left-right diagonal (A), right-left diagonal (B), anterior-posterior (C), fronto-central maximum (D), and occipito-central maximum (C'). To investigate how frequency bands influenced these patterns, the EEG data were filtered into specific bands, and the topographies for each condition were visualized using 'mne.viz.plot_topomap'[16]. This visualization highlighted differences between EO and EC conditions, with alpha-band activity being particularly prominent during EC [16]. The classification model used to distinguish between eyes-open

(EO) and eyes-closed (EC) conditions based on EEG microstate features is a Linear Support Vector Machine (SVM). The feature set consisted of the three key classical microstate parametersGlobal Explained Variance (GEV), Time Coverage (TimeCov), and Mean Durations (MeanDurs) calculated for five distinct microstate maps (A, B, C, D, and C) [16]. The model was separately trained and tested on data from the alpha frequency band and broadband signals. The Linear SVM model was evaluated using a 10-fold cross-validation procedure across data from 203 subjects. The dataset was split into training, validation, and test sets using an 80-10-10 split, with stratified sampling to maintain class balance. Hyperparameter tuning was performed on the training set using a grid search approach to optimize the regularization parameter C and the tolerance *tol*. The final model's performance was assessed on the test set using metrics such as accuracy and the area under the ROC curve (AUC), with confidence intervals calculated to account for variability. [16].

3.2 Simulating Variability in Microstates: Artificial Perturbations of Microstate Features

3.2.1 Simulating Perturbations

To closely create real-world conditions where models encounter variability, perturbations were applied to the baseline features to simulate variability as close to reality as possible:

Mean Perturbation: A fixed decrement (e.g., 0.01) was subtracted from the mean values of each metric to simulate reduced performance due to noise or other external factors. Standard Deviation Perturbation: A fixed increment (e.g., 0.001) was added to the standard deviation to reflect increased uncertainty in the measurements. These changes created the perturbed metrics (*_p_m and *_p_s), which represent the model's performance under perturbation.

For the Accuracy metric, the Pooled variance was computed to combine the variability from both baseline and perturbed conditions [7]. The formula used was:

$$\text{Pooled Variance} = \frac{2 \cdot \text{Baseline Variance} + 2 \cdot \text{Perturbed Variance}}{6 - 2}$$

This calculation accounted for the contributions of variability from both

conditions, providing a comprehensive measure of the overall variance. To calculate the differences in Accuracy between baseline and perturbed conditions, Confidence intervals were calculated using the pooled variance. These intervals quantify the uncertainty in performance changes and provide a statistical measure of the robustness of the models under perturbation. The bounds were calculated as:

$$\text{Lower Bound} = (\text{Baseline} - \text{Perturbed}) - 2.05 \cdot \sqrt{\frac{2 \cdot \text{Pooled Variance}}{3}}$$

$$\text{Upper Bound} = (\text{Baseline} - \text{Perturbed}) + 2.05 \cdot \sqrt{\frac{2 \cdot \text{Pooled Variance}}{3}}$$

To better understand the impact of perturbations, scatter plots with error bars were generated for each metric (AUC, Accuracy, and F1 Score). These visualizations compared baseline and perturbed metrics for all models, highlighting key differences. The error bars for the plots were computed as:

$$\text{Confidence Interval (CI)} = 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation, and n is the sample size.

3.2.2 Quantifying Perturbations

To assess variability and uncertainty in EEG microstate features, two key metrics are applied: Hellinger Distance and Mahalanobis Distance [7]. The two metrics, Hellinger and Mahalanobis distances were chosen because they complement each other in assessing different aspects of variability [9]. These measures provide a robust framework for comparing probabilistic labels, modeled as Gaussian distributions, and quantifying differences in feature distributions [7] [9].

This metric quantifies the similarity between probability distributions, which is particularly useful for comparing probabilistic representations of features (e.g., GaussianFuzzyLabel). It accounts for the differences in both means and variances of the distributions, making it suitable for understanding the overall shift between two feature distributions.

The **Hellinger Distance** metric quantifies the similarity between probability

distributions, which is particularly useful for comparing probabilistic representations of features (e.g., Gaussian Fuzzy Label). Measures the similarity between two probability distributions, scaled between 0 (identical distributions) and 1 (maximally dissimilar distributions). It accounts for both the mean and variance of the distributions, making it suitable for understanding the overall shift between two feature distributions [29].

$$H(P, Q) = \sqrt{1 - \mathrm{BC}(P, Q)}$$

where:

$$\mathrm{BC}(P, Q) = \frac{\sqrt{\det(\Sigma_1)^{1/2} \cdot \det(\Sigma_2)^{1/2}}}{\sqrt{\det\left(\frac{\Sigma_1+\Sigma_2}{2}\right)}} \cdot \exp\left(-\frac{1}{8}(\mu_1 - \mu_2)^\top \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\mu_1 - \mu_2)\right)$$

The **Mahalanobis Distance** measures the separation between a point and a distribution, considering the correlations among variables. It is widely used in multivariate analysis, including classification and clustering, due to its ability to account for the covariance structure of the data. The Mahalanobis distance is particularly effective in identifying outliers and understanding the structure of multivariate data. [15]. In the context of EEG analysis, both distances have been employed to assess variability and uncertainty in EEG microstate features. For instance, the Mahalanobis distance has been utilized to measure the separation between different EEG states, considering the covariance among EEG features. Similarly, the Hellinger distance has been applied to compare the similarity between probability distributions of EEG features, providing insights into the variability of brain activity patterns.

$$D_M(P, Q) = \frac{1}{2}\left(\sqrt{(\mu_1 - \mu_2)^\top \Sigma_1^{-1} (\mu_1 - \mu_2)} + \sqrt{(\mu_1 - \mu_2)^\top \Sigma_2^{-1} (\mu_1 - \mu_2)}\right)$$

Each feature in the datasets, such as Global Explained Variance (GEV) and Mean Durations (MeanDurs), was assigned a Gaussian Fuzzy Label, where the mean reflects the expected value of the feature and the standard deviation quantifies its variability. These values were either derived from dataset-specific variability factors or based on assumptions from relevant literature [7]. By using this approach, each feature was represented not as a fixed value but as a distri-

bution, capturing the inherent uncertainty due to biological and measurement inconsistencies [7].

The purpose of creating these two datasets was because it allowed the quantification of the deviation between the perturbed dataset and the original dataset. This helped to evaluate how well augmentation techniques introduced realistic variability while preserving the representativeness of the data. Second, the datasets provided a framework to assess the effectiveness of machine learning methods in handling variability, with probabilistic labels enabling the simulation of real-world conditions [7]

The implementation of probabilistic labels [7] involved sampling values based on the predefined means and standard deviations of each feature. This sampling process supported the calculation of the Hellinger Distance, which quantified the differences in distributions between the datasets (non-perturbed and perturbed) and Mahalanobis Distance measured multivariate deviations between the datasets, accounting for correlations between features [7][9]. Together, these metrics provided insights into how variability affected feature distributions and how well the augmentation techniques maintained the integrity of the data [7]. This approach underscores the importance of probabilistic representations in capturing the uncertainty inherent in EEG microstate features and ensuring robust model performance in the presence of variability.

3.3 Classification Models with Implementation

This section describes advanced machine learning classifiers and methods designed to address variability and uncertainty in datasets, particularly those involving probabilistic or imprecise features like EEG microstate characteristics. The methods implemented include probabilistic classifiers, kernel-based approaches, and ensemble techniques tailored for uncertainty-aware classification. The chosen models are based on algorithms proposed in literature that also investigated methods to manage imprecise data [7]. Four classifiers are implemented to evaluate variability-aware classification [7]:

- KND-f (ImpreciseKNeighborsClassifier): Uses a custom Mahalanobis distance metric for variability-aware, distance-based classification [40].
- ACS-a (Augmented SVC): A support vector classifier trained on augmented data for robust classification [7].

- ACG-a (Augmented Gradient Boosting Classifier): Tree-based ensemble learning method incorporating probabilistic augmentation for enhanced robustness [19].
- WSF-i (Imprecise Forest): Ensemble classifier that trains decision trees on resampled datasets, addressing variability in data distribution [32] [14].

The K-Nearest Neighbors (KNN) algorithm was extended by incorporating custom distance metrics and probabilistic predictions, enabling robust classification in the presence of uncertainty. The classifier is initialized with parameters such as the number of nearest neighbors (k) and a custom distance metric tailored to uncertain feature spaces. For each test point, the classifier calculates distances using the specified custom metric, identifies the k-nearest neighbors, and predicts the most frequent label among them. This method is suitable for handling noisy data and scenarios where uncertainty in the similarity metric impacts classification accuracy.

Two kernel methods for probabilistic features are utilized. Mean Embedding Kernel computes similarity by sampling multiple values from probabilistic labels (e.g., Gaussian Fuzzy Label) and calculating average pairwise similarities. The process involves sampling n-values from probabilistic labels and using kernels such as radial basis function (RBF), polynomial, or sigmoid to compute pairwise similarities. The results are then averaged to produce an overall similarity score. The other kernel used is RBF Embedding Kernel. It is a kernel designed to account for both the means and variances of probabilistic labels, leveraging Gaussian properties to enhance similarity computations.

The Imprecise Forest Classifier builds upon ensemble learning techniques to handle uncertainty or imprecise labels by training multiple decision trees on resampled data. The classifier is initialized with parameters such as the base estimator (e.g., ExtraTreeClassifier) and the number of trees. Each tree is trained on either the entire dataset or resampled subsets, capturing diverse perspectives of the data. This allows the model to incorporate variability during the training phase. Predictions from all trees are aggregated, with the final class being assigned based on the highest aggregated score.

These methods address challenges in EEG microstate analysis, related to variability and uncertainty in feature measurements. These approaches are implemented to address the robustness of model, ensuring that variability and uncertainty are effectively managed. **Cross-Validation:** A StratifiedKFold cross-validation approach with three splits was implemented to ensure balanced class

representation across training and test sets. This method was chosen to maintain class proportions, especially important given the variability in EEG microstate data. [7]. For each fold, both the training and test datasets were augmented using the Data Generator method. This ensured the incorporation of realistic variability into the data to simulate real-world conditions. The augmented training data was used to train the models. Each model learned from probabilistic representations of feature values, enhancing its robustness to variability: The models were evaluated on augmented test data to quantify their performance under baseline and perturbed conditions. Metrics were computed across all folds to provide mean values and standard deviations, highlighting performance consistency and robustness.

The computational environment for this study was Google Colab, utilizing Python version 3.10.12 for implementing and evaluating machine learning methods, NVIDIA A100-SXM4-40GB GPU with CUDA version 12.1

3.4 Assessing the Impact of Variability on Classification

In this study, the models were evaluated on perturbed and non-perturbed data. The non-perturbed dataset orresponds to the original feature distribution, which represents the baseline EEG data, while the perturbed dataset captures the perturbed or augmented feature distribution, simulating real-world conditions by introducing variability [7]. These datasets are used to assess the impact of variability and uncertainty on EEG microstate features, with each feature modeled as a probabilistic representation using Gaussian Fuzzy Labels.

Resting-state EEG recordings were conducted using 61 scalp electrodes arranged according to the international 10-20 extended localization system, with FCz as the reference electrode. A vertical electrooculography (EOG) electrode was added to monitor right-eye activity. The recordings had a sampling frequency of 2500 Hz, which was later down-sampled to 250 Hz and then to 100 Hz, with an amplitude resolution of 0.1 μV, and an initial bandpass filter (0.0151 kHz), that was refined later for specific analyses. The ground electrode was located at the sternum and the electrode impedance was maintained below 5 kΩ. FOr the timespan of 16 minutes, alternate 60-second EO and EC blocks were recorded in an electrically shielded, sound-attenuated booth (8 blocks each, be-

ginning with EC). Participants in EO were told to stay awake and focus on a black cross on a white background. All the details are reported in detail by the experts who made the dataset available [4].

3.4.1 Microstates and their artificial perturbation

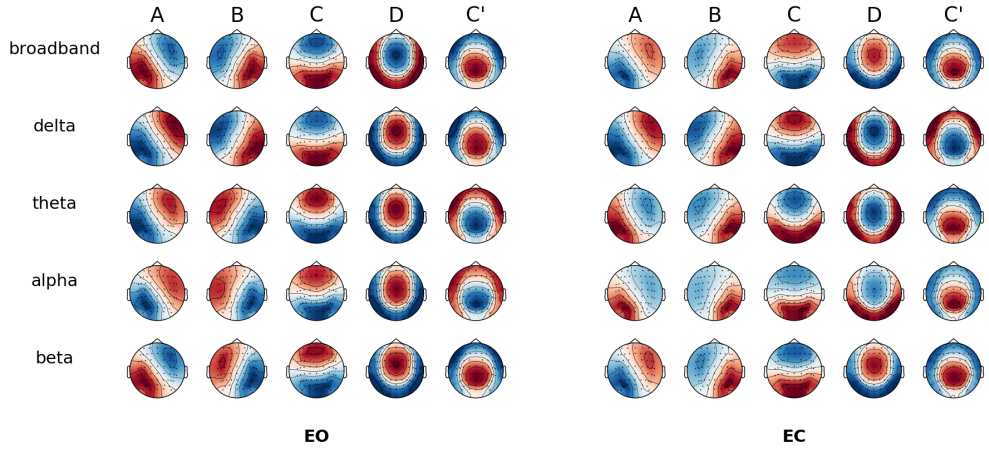


Figure 3.2: Spatial correlation between microstate (MS) topographies across behavioural conditions. Global cluster centroids of each frequency band for the eyes-open (EO) or eyes-closed (EC) condition [16].

To quantify the variability inherent in EEG microstate features, in this thesis work, it was implemented a systematic computational framework that calculates variability metrics across analytical, within-subject, and between-subject levels of the monitored individuals over multiple time points. This approach enables the evaluation of feature reliability and consistency, which is critical for understanding EEG microstate dynamics and improving the robustness of machine learning models [7] [16].

These metrics were calculated for the following EEG microstate features:

- Gev (Global Explained Variance): Reflects the explanatory power of a microstate.
- MeanDurs (Mean Durations): The average duration of microstate segments.
- TimeCov_corrected (Corrected Time Coverage): The fraction of time covered by a specific microstate [16].

IV perturbations of the original dataset were kept as close to reality as possible in order to simulate real life perturbations that might occur when training

machine learning model with clinical data. In this study perturbations were performed on the machine learning models in order to assess how robust where the models faced with variability. These methods where applied to the ML models that were designed based on literature to be more robust and take in consideration uncertainty and variability, and also to the simple ML model used for the microstates behavioral classification. Below is the description of the methods used in this thesis work to introduce uncertainty to the models, and later on to assess how this uncertainty influenced the performance and robustness of the model.

Gaussian Fuzzy Label and Random Label: These classes introduce variability and probabilistic reasoning into label representation, making them particularly useful for tasks involving imprecise or noisy data, such as EEG microstate analysis [7].

The Gaussian Fuzzy Label class represents a fuzzy label modeled as a Gaussian distribution, defined by a mean (μ) and standard deviation (σ). This allows labels to be associated with a range of plausible values rather than fixed deterministic values, which is essential for capturing uncertainty in measurements [7].

The Random Label class represents a general approach to probabilistic labels, supporting any probability distribution (e.g., Gaussian, uniform, or exponential). This flexibility allows for modeling variability in scenarios where data do not follow a Gaussian distribution. Both classes are integrated into the data augmentation pipeline to introduce realistic variability into EEG microstate features. This approach enhances the robustness of machine learning models by ensuring they are exposed to a range of plausible scenarios, ultimately improving generalization and reliability. These probabilistic labeling techniques provide a foundational step toward uncertainty-aware modeling, enabling more accurate interpretations of noisy, real-world data [7].

3.4.2 Evaluation Metrics

to evaluate the robustness of machine learning models when exposed to perturbations, simulating real-world variability. The performance of the models was assessed using these key metrics: Area Under the Curve (AUC), Accuracy (ACC), and F1 Score, under both baseline and perturbed conditions.

To assess the model performance under both baseline and perturbed condi-

tions, a strategy combining cross-validation and detailed performance metrics was used. This strategy ensured that variability was incorporated into the process, and the robustness of machine learning models was evaluated.

Metrics:

- AUC (Area Under the Curve): This metric measures the model's ability to distinguish between classes, providing an aggregate measure of performance across all classification thresholds. Accuracy: Represents the proportion of correctly predicted instances among the total instances evaluated, offering a direct measure of overall correctness.
- Accuracy: Reflects the proportion of correct predictions.
- F1 Score: Balances precision (the ratio of true positive predictions to all positive predictions) and recall (the ratio of true positive predictions to all actual positives). This metric is particularly important for imbalanced datasets where precision and recall trade-offs need to be balanced.

4

Results and Discussions

This chapter presents the results of this study, focusing on the performance of machine learning models applied to EEG microstate features under both baseline and perturbed conditions. The analysis emphasizes the effects of variability and uncertainty on the data distributions and machine learning models, highlighting the role of techniques such as data augmentation and variability-aware methods in improving model robustness. Key sections include variability analysis, the impact of data augmentation, classifier performance, and model robustness under perturbations. The results demonstrate that classifiers leveraging data augmentation and variability-aware methods significantly outperform traditional models under conditions of variability and uncertainty. The findings underscore the importance of robustness as a critical criterion for machine learning models applied to EEG microstate analysis and similar real-world datasets.

Figure 4.1 illustrates a snapshot of raw EEG signals recorded from multiple channels across the scalp, showcasing the complexity and variability inherent in these biosignals. Complementing the time-domain signals, Figure 3.2 [16] presents the topographic distribution of EEG microstates across five frequency bands (broadband, delta, theta, alpha, beta) for eyes-open (EO) and eyes-closed (EC) conditions. The topographic maps highlight the spatial patterns associated with each microstate (A, B, C, D, C') and their variations across different frequency bands and behavioral conditions. These visualizations provide insights into how brain activity is distributed and modulated by frequency-specific dynamics and external states. In addition, Figure 3.2 displays the power spectral density (PSD) of the EEG signals, which decomposes the EEG recordings into

their constituent frequency components. The PSD plot distinctly highlights the contributions of delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), and beta (13-30 Hz) frequency bands. Using this data this work tries to investigate how the inherent complexity of EEG signals forms the basis for exploring variability and developing robust models capable of handling real-world conditions.

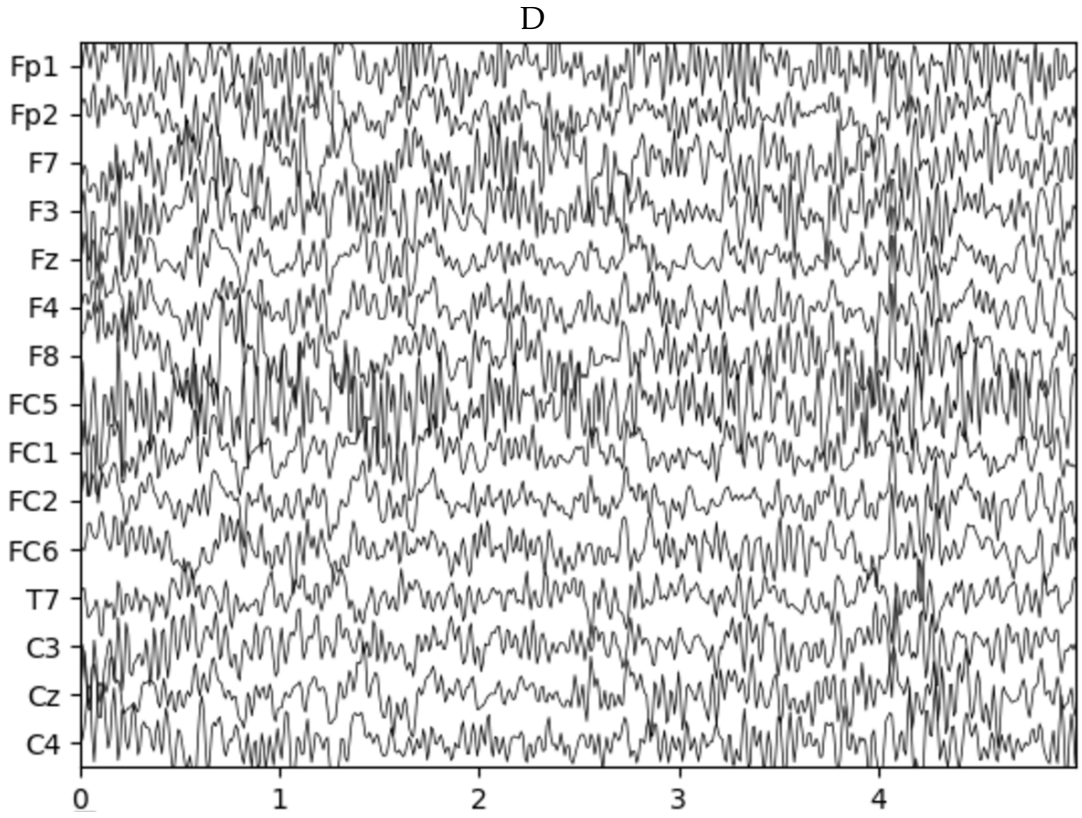


Figure 4.1: Raw EEG data

4.1 Variability Analysis

The results of the analysis, quantified using Hellinger and Mahalanobis distances, provide insights into the impact of perturbations on the data distributions and their potential implications for machine learning model performance.

Hellinger Distance: 0.909. A value of 0.909 indicates a high degree of dissimilarity between the baseline (original) and perturbed data distributions. This means that the perturbations introduced into the data caused changes to its distribution. *Mahalanobis Distance: 3.873*, the observed value of 3.873 indicates a substantial divergence in the multivariate structure of the perturbed data com-

pared to the baseline. This suggests that the perturbations affected not only individual features but also their interrelationships, such as correlations or covariance patterns. Both metrics, the Hellinger and Mahalanobis distances, reveal a comprehensive picture of how the perturbations influenced the data.

The high Hellinger distance suggests that models trained on baseline data might face difficulty when applied to perturbed data, as the shift in feature distributions could reduce classification accuracy. The high Mahalanobis distance further indicates that the altered feature relationships may distort decision boundaries, compounding the difficulty of maintaining robust performance under variability. In Table 4.3 are reported also the values for the variability factors discussed in methods section 3.2.3.

Map	**Metric**	**CVA**	**CVI(s)**	**CVI(m)**	**CVI(ma)**
Map1	Gev	0.4484	0.3129	0.3103	0.6287
Map2	Gev	0.4320	0.3017	0.3050	0.6088
Map3	Gev	0.3766	0.2676	0.2497	0.5251
Map4	Gev	0.6101	0.4966	0.3462	0.8595
Map5	Gev	0.6277	0.5024	0.3708	0.8853
Map1	MeanDurs	0.2948	0.2832	0.0989	0.4206
Map2	MeanDurs	0.3170	0.3037	0.1001	0.4503
Map3	MeanDurs	0.4691	0.3937	0.1627	0.6337
Map4	MeanDurs	0.2904	0.2680	0.1060	0.4091
Map5	MeanDurs	0.2994	0.2738	0.1098	0.4203
Map1	TimeCov_corrected	0.3512	0.2460	0.2436	0.4931
Map2	TimeCov_corrected	0.3355	0.2341	0.2373	0.4729
Map3	TimeCov_corrected	0.2911	0.1940	0.2083	0.4072
Map4	TimeCov_corrected	0.4372	0.3414	0.2714	0.6175
Map5	TimeCov_corrected	0.4369	0.3292	0.2865	0.6176

Table 4.1: Variability Factors for Gev, MeanDurs, and TimeCov_corrected Features.

Then we evaluate how data augmentation influenced the statistical properties of EEG microstate features, focusing on key metrics: mean, standard deviation (Std), skewness, and kurtosis, before and after augmentation. These metrics reveal how augmentation affects the data's central tendency, variability, symmetry, and the presence of outliers, providing valuable insights into the robustness and generalizability of augmented datasets for machine learning applications. All the results can be seen in Figures 4.6 - 4.8, alongside the respective statistical metrics in Table 4.1-4.2. The first two panels in each figure present

the histograms of specific features, highlighting the differences in data distribution before and after the augmentation process. In the "Before Augmentation" histograms, the distributions appear narrower, reflecting lower variability in the feature values. This representation captures the original dataset's characteristics, which remain unaltered by external perturbations. While, the "After Augmentation" histograms exhibit broader distributions, signifying an increase in variability. This observed is consistent with the intended effects of the augmentation process, which introduces controlled noise or variability to emulate real-world diversity in the data. It is also worth mentioning that the y-axis scales in these panels differ, as the augmented dataset typically includes a larger number of samples or greater spread compared to the original data.

The third panel in each figure displays overlapping distributions, comparing the probability density functions (PDFs) of the original and augmented datasets. These panels reveal a noticeable shift in the augmented dataset's distribution, indicating that variability has been successfully introduced while preserving the general shape and fundamental characteristics of the original data. For key features such as MeanDurs, TimeCov_corrected, and Gev, the augmentation process demonstrates its effectiveness by maintaining the essential properties of the original feature distributions while expanding their range. This balance ensures that the augmented data not only reflects the inherent variability present in real-world scenarios but also remains representative of the original dataset's core traits.

From the values reported in Tables 4.1-4.2 , we can see that: The tables illustrate the statistical properties of the EEG microstate features "MeanDurs," "TimeCov_corrected," and "Gev" before and after the augmentation process. The mean values for all features ("MeanDurs," "TimeCov_corrected," and "Gev") remained relatively consistent after augmentation. For example, the mean of "MeanDurs" shifted slightly from 0.1017 to 0.1035. This stability indicates that the augmentation process preserved the central tendency of the data, which is essential for ensuring that the original signal characteristics remain intact. Large deviations in mean values could distort the underlying patterns of the dataset, potentially compromising the interpretability and reliability of downstream machine learning models. A noticeable increase in the standard deviation is evident across all features post-augmentation. For instance, the standard deviation of "TimeCov_corrected" increased from 0.0585 to 0.0933, and "Gev" saw a rise from 0.0378 to 0.0604. This broadening reflects the successful introduction of vari-

ability into the dataset, which mimics real-world diversity. Such variability is particularly advantageous in EEG microstate studies, where biological and technical factors inherently introduce uncertainty. By broadening the spread of data points, the augmentation process enhances the generalizability of models to unseen scenarios. The skewness metric, which measures the asymmetry of feature distributions [23], exhibited mixed trends. For "MeanDurs," skewness decreased slightly from 1.2651 to 1.0454, suggesting a more symmetric distribution. However, "TimeCov_corrected" and "Gev" showed increases in skewness, moving from 0.3951 to 0.6731 and 0.6609 to 0.8892, respectively. Kurtosis, a measure of the "tailedness" or extremity of outlier values in a distribution, also demonstrated feature-dependent behavior. For "MeanDurs," kurtosis increased from 2.3217 to 2.7469, reflecting a more pronounced presence of outliers. In contrast, "TimeCov_corrected" exhibited a reduction in kurtosis from 0.8698 to 0.5952, indicating a smoothing of the distribution and fewer extreme values. For "Gev," kurtosis reduced from 1.8387 to 1.0569, similarly pointing to a decline in the prevalence of extreme values. These changes highlight the dual role of augmentation in both dampening and amplifying outlier behavior, contingent on the feature's original distribution characteristics [37]. The mean values of most features remained relatively stable after augmentation, indicating that the central tendency of the data was preserved. This consistency is important as drastic shifts in mean values could distort the underlying patterns within the data, potentially affecting the interpretability and effectiveness of machine learning models. A consistent increase in the standard deviation was observed across all features, highlighting that data augmentation successfully introduced higher variability. By broadening the spread of data points, the augmentation process mimics real-world diversity, enabling models to generalize better to unseen conditions. This is particularly beneficial in EEG microstate studies, where variability due to biological and technical factors is inherent [16]. The data augmentation process effectively introduced variability into the dataset while preserving the central tendencies (mean values) of most features [7] [25]. This balance ensures that augmented data remains representative of the original dataset while simulating real-world variations.

These adjustments are useful for training machine learning models to handle variability and uncertainty. The enhanced variability observed in the augmented features improves model robustness, equipping classifiers to generalize better across diverse conditions.

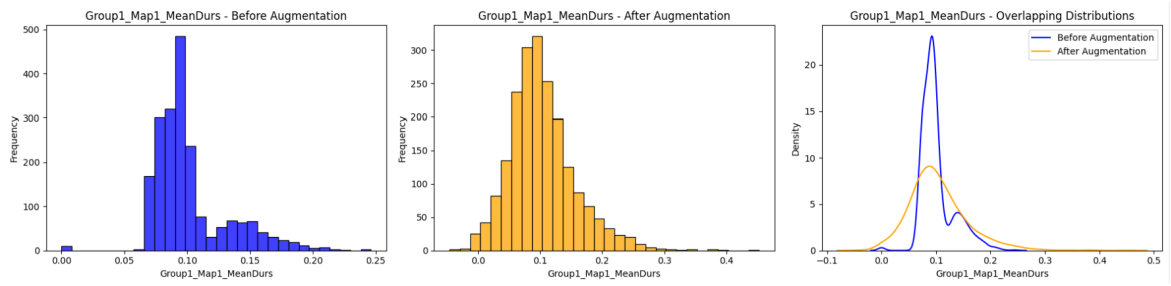


Figure 4.2: Mean Duration Feature Map 1 Before and After Augmentation

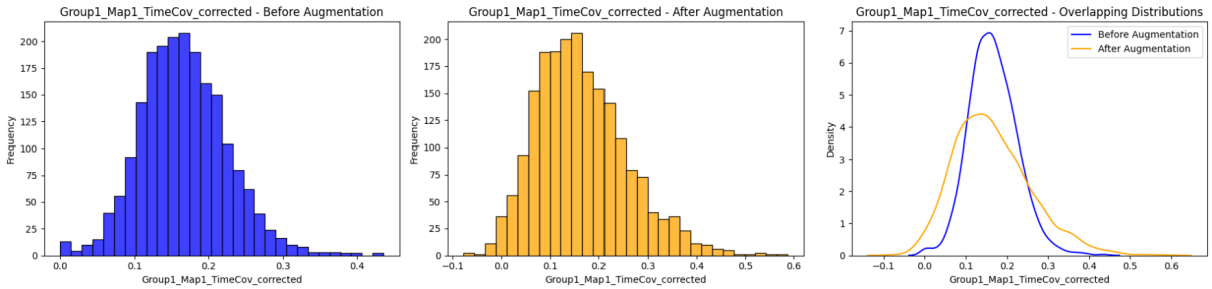


Figure 4.3: TimeCov_Corrected Feature Map 1 Before and After Augmentation

Statistic	MeanDurs	TimeCov_corrected	Gev
Mean	0.1017	0.1666	0.0843
Standard Deviation	0.0300	0.0585	0.0378
Skewness	1.2651	0.3951	0.6609
Kurtosis	2.3217	0.8698	1.8387

Table 4.2: Statistical Metrics for Map 1 - Before Augmentation

Statistic	MeanDurs	TimeCov_corrected	Gev
Mean	0.1035	0.1644	0.0848
Standard Deviation	0.0538	0.0933	0.0604
Skewness	1.0454	0.6731	0.8892
Kurtosis	2.7469	0.5952	1.0569

Table 4.3: Statistical Metrics for Map 1 - After Augmentation

4.2 The Impact of Variability on Classification

This section presents the evaluation of machine learning models applied to EEG microstate features under conditions of variability and uncertainty. Four classifiers : KND-f, ACS-a, ACG-a, and WSF-i were implemented and tested us-

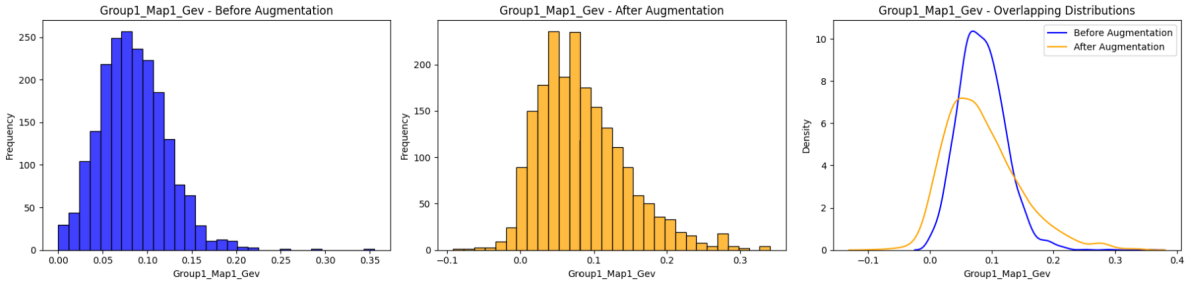


Figure 4.4: Gev Feature Map 1 - Before and After Augmentation

ing metrics like AUC (Area Under the Curve), Accuracy, and F1 Score. Stratified K-Fold cross-validation with three folds was used to ensure robust evaluation, with results expressed as mean ± standard deviation across splits [7].

4.2.1 Classification Performance with no Perturbations

Below are reported the results of the classifiers on the baseline datasets with no perturbations. The chance level is 20% and the classification task in the code is a multi-class classification. The goal is to predict the frequency band (broadband, delta, theta, alpha, beta) corresponding to EEG microstate data.

KND-f (ImpreciseKNeighborsClassifier)

- AUC: 0.9366 ± 0.0022
- Accuracy: 0.7916 ± 0.0140
- F1 Score: 0.7921 ± 0.0143

The KND-f classifier demonstrates moderate performance across all metrics, with relatively low variability. Its probabilistic approach effectively handles uncertainty but does not outperform other models in distinguishing between classes.

ACS-a (Augmented SVC)

- AUC: 0.9855 ± 0.0009
- Accuracy: 0.9054 ± 0.0032
- F1 Score: 0.9054 ± 0.0031

The ACS-a model achieves the highest performance across all metrics, with exceptional AUC, accuracy, and F1 score. Its minimal standard deviations highlight its robustness and consistency. The use of data augmentation in this model

likely enhances its ability to generalize across variable conditions, making it the most effective classifier in this study.

ACG-a (Augmented Gradient Boosting Classifier)

- AUC: 0.9596 ± 0.0039
- Accuracy: 0.8167 ± 0.0207
- F1 Score: 0.8159 ± 0.0219

ACG-a exhibits high AUC, indicating strong class separation ability. However, its accuracy and F1 score are slightly lower than those of ACS-a, with higher variability (standard deviations), suggesting sensitivity to data splits and a need for further optimization.

WSF-i (Weighted Resampling Forest)

- AUC: 0.9589 ± 0.0030
- Accuracy: 0.8222 ± 0.0099
- F1 Score: 0.8190 ± 0.0116

WSF-i shows good performance, with metrics close to those of ACG-a but slightly lower variability. This stability makes it a reliable alternative for scenarios requiring robust classification under uncertain conditions.

The ROC curve in Figure 4.5 evaluates the classification performance of the models applied to EEG microstate data, with the Area Under the Curve (AUC) serving as the primary metric to quantify discriminatory power. Among the models, the Weighted Sampling Forest (WSF) shows the most consistent and robust performance, achieving high AUC values across all classes (0.910.95). The Augmented Classifier Support Vector Machine (ACS) also performs well, with AUC values generally around 0.850.86, though it struggles with Class 4, which has a lower AUC of 0.77. The Augmented Gradient Boosting Classifier (ACG) and Imprecise KNeighbors Classifier (KND) exhibit moderate performance, with AUC values ranging from 0.76 to 0.89 and 0.79 to 0.90, respectively. Class 4 proves to be the most challenging to classify, likely due to overlapping feature distributions with other classes, while Classes 0, 1, and 2 are classified more effectively. All models outperform the random classifier baseline (AUC = 0.5), underscoring their ability to leverage the input features for classification.

The confusion matrices in Figures 4.64.9 provide a comparative analysis of the performance of the KND, ACS, ACG, and WSF models in classifying EEG microstate data into five classes (04). The KND model shows strong performance

for Class 1 but struggles with overlapping classes, evident in misclassifications like Class 2 being predicted as Class 0 or 3. The ACS model demonstrates improved precision for certain classes, particularly Class 4, while still exhibiting some misclassification issues, such as confusion between Class 0 and Class 1. The ACG model achieves the highest accuracy for Class 1 but underperforms for Class 4 and exhibits significant misclassifications for Class 2. The WSF model achieves a balanced performance, excelling in Class 1 and Class 3, but faces challenges with Class 0, which shows significant confusion with Classes 2 and 4. This comparison highlights the varying strengths and weaknesses of each model, with the ACG model excelling in handling separable features, the ACS model achieving consistent performance, and the WSF model showing promise but requiring improvements in handling overlapping classes.

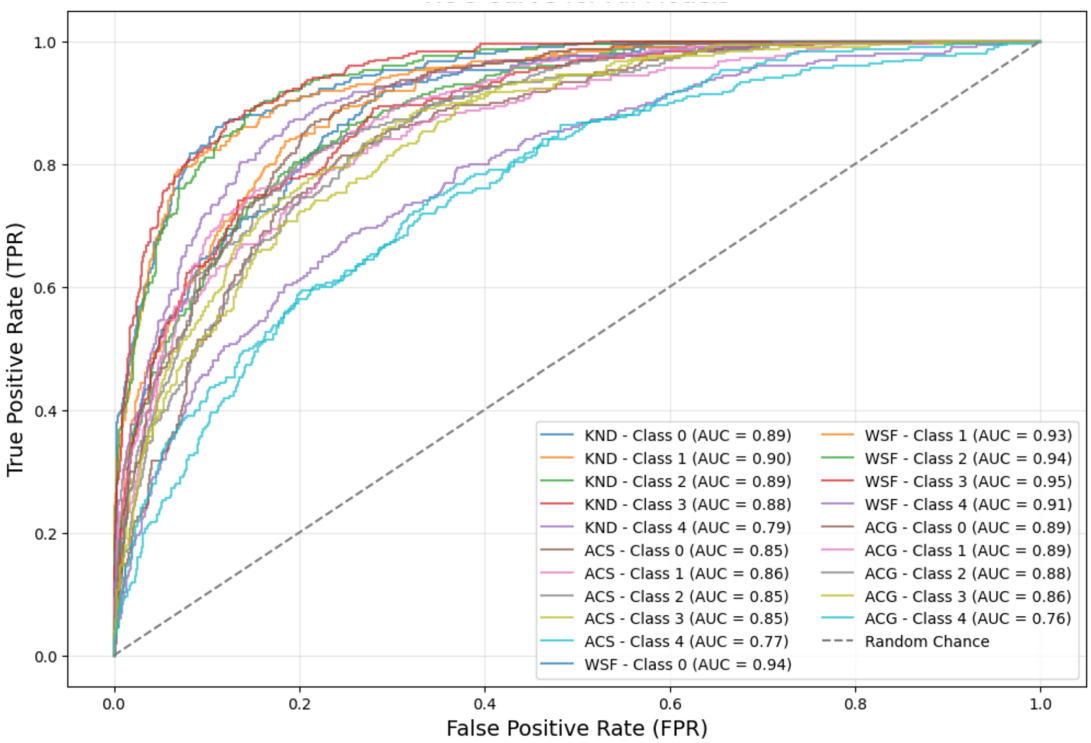


Figure 4.5: ROC Curve for all the models.

Based on the performance of the classifiers under baseline conditions, ACS outperforms all models, achieving near-perfect AUC (0.99) and the highest Accuracy and F1 Score (0.91 ± 0.00). This reflects the effectiveness of the augmented classifier in leveraging variability in training data to improve predictive power. ACG and WSF provide competitive results but fall short of ACS, particularly in standard deviations, which are slightly higher, indicating less consistent per-

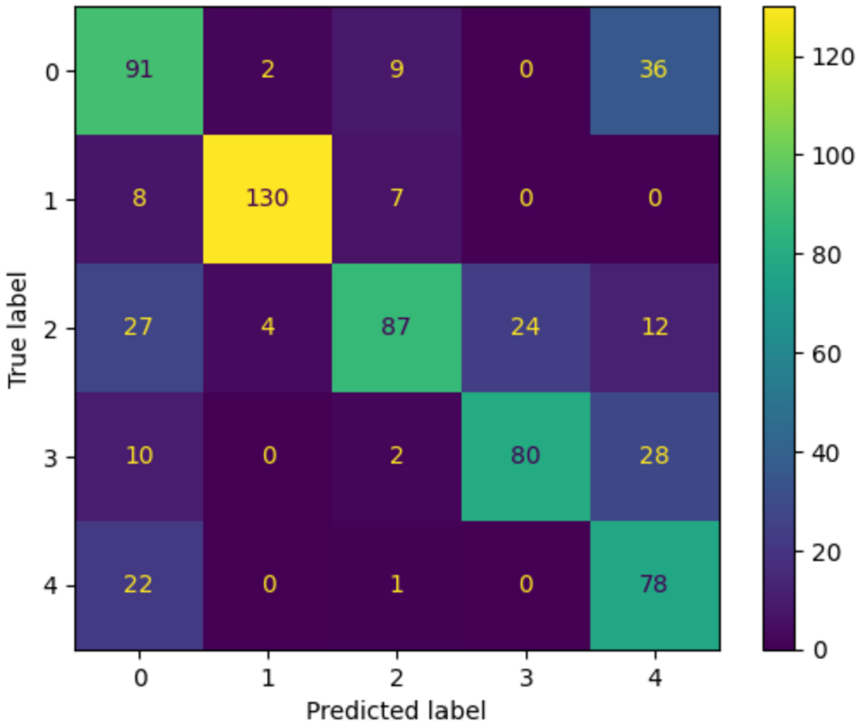


Figure 4.6: KND Model Confusion Matrix

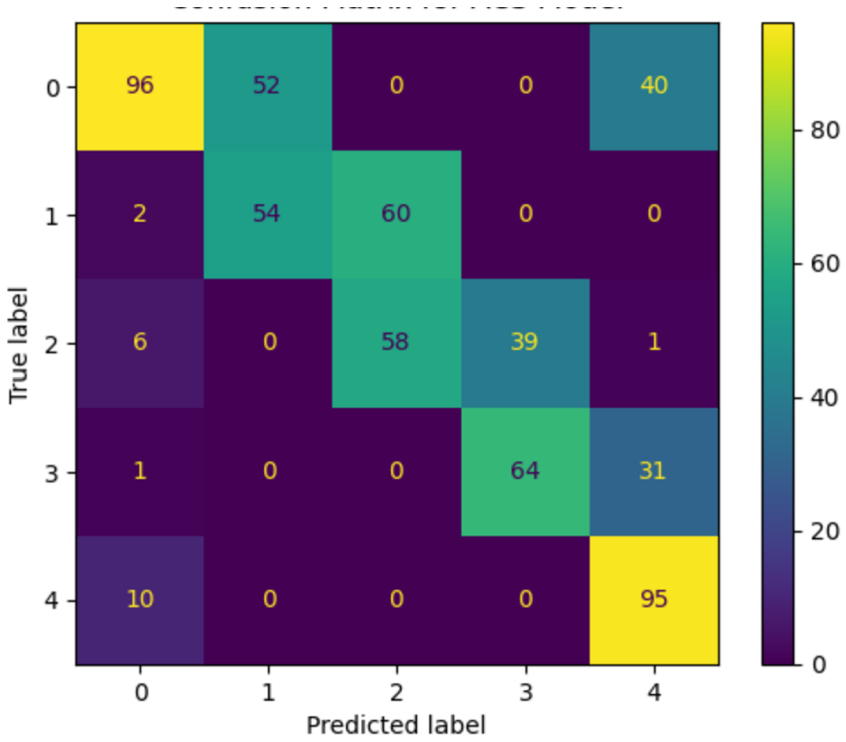


Figure 4.7: ACS Model Confusion Matrix

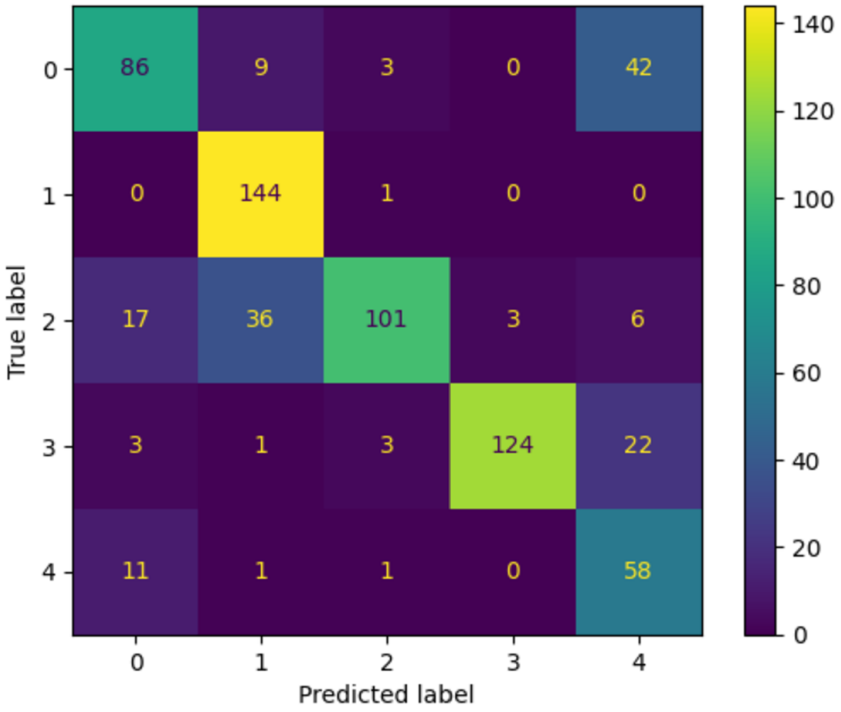


Figure 4.8: ACG Model Confusion Matrix

formance. KND lags behind the other models, showing moderate performance that is more adequate for controlled conditions.

4.3 Performance Degradation under Perturbations

The results provided below, demonstrate the performance of four machine learning models: KND, ACS, ACG, and WSF evaluated under both baseline (results discussed in the section above) and under perturbed conditions. The evaluation metrics used are AUC (Area Under the Curve), Accuracy, and F1 Score, which are indicators of classification performance. Perturbation is used to simulate real-world variability, introducing noise or uncertainty into the data. The results highlight that all models exhibit some level of performance degradation when moving from baseline to perturbed conditions. This reflects the models' varying degrees of robustness in handling uncertainty. However, their performance is quite good, and the results are still reliable compared to the performance of the ML that didn't have methods to account for variability. Below are the results obtained from the perturbation of the 4 ML models. In figure 4.16 can also be seen the visualization of these results and the comparison of the

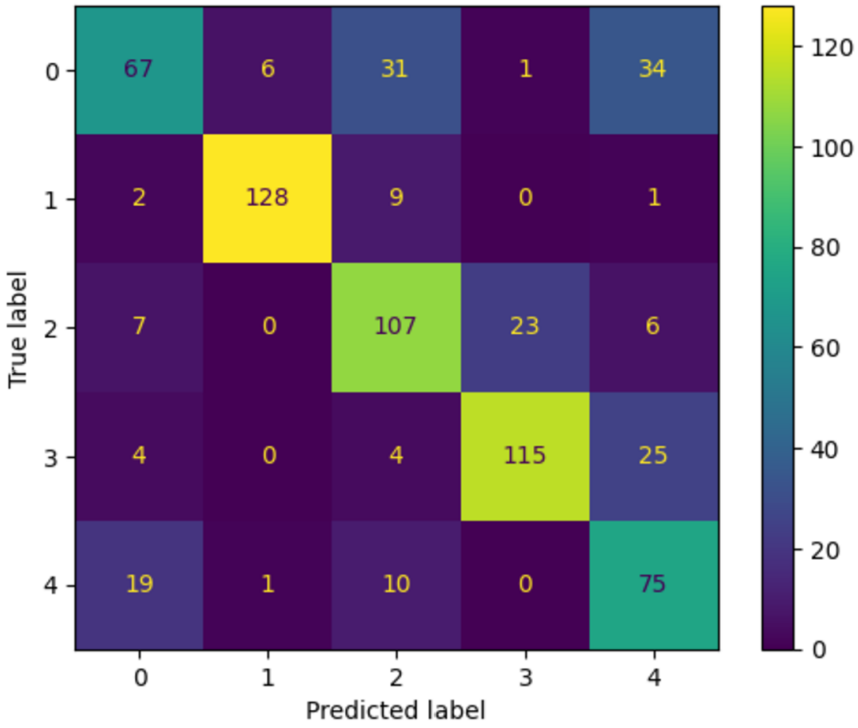


Figure 4.9: WSF Model Confusion Matrix

baseline vs pertubed performance. Since the results of the baseline performance are discussed in the previous section, below will be discussed the performance of the models with a perturbed dataset.

KND (Imprecise KNeighbors) - Perturbed Performance: It can be observed a performance decline, with AUC dropping to 0.84 and both Accuracy and F1 Score falling to 0.71 ± 0.02. This substantial drop makes this model the least robust model in the group.

ACS (Augmented Classifier with SVM) - Perturbed Performance: Despite perturbation, ACS maintains high performance relative to other models, with AUC dropping slightly to 0.89 and Accuracy and F1 Score reducing to 0.81 ± 0.00. The smaller performance drop demonstrates ACS's robustness to data variability. ACS emerges as the most reliable model, showcasing the effectiveness of data augmentation techniques in improving generalization and stability under real-world variability.

ACG (Augmented Gradient Boosting Classifier) - Perturbed Performance: Performance decreases to an AUC of 0.86 and Accuracy and F1 Score to 0.74 ± 0.03. The slightly higher standard deviations compared to ACS indicate that ACG is more sensitive to variability in the data. While ACG performs well under

baseline conditions, it is moderately impacted by perturbations, suggesting it benefits less from data augmentation than ACS.

WSF (Weighted Sampling Forest) - Perturbed Performance: Similar to ACG, WSF shows a decline in performance, with AUC at 0.86 and Accuracy and F1 Score at 0.74 ± 0.01. However, the smaller standard deviations compared to ACG indicate more stable behavior under perturbations. WSF demonstrates robust performance and stability, making it a valuable alternative for scenarios requiring both robustness and consistent predictions.

4.3.1 Comparative Analysis of Models

Performance Under Perturbed Conditions: ACS remains the most robust model, with a smaller relative drop in all metrics (AUC: -0.10, Accuracy: -0.10, F1 Score: -0.10). This indicates that its data augmentation techniques enable it to better handle real-world variability. WSF and ACG are closely matched in perturbed conditions, with nearly identical AUC (0.86) and slightly lower Accuracy and F1 Scores (0.74). However, WSF exhibits smaller variability, making it slightly more reliable. KND shows the largest performance decline out of the four, with a drop in AUC (-0.10) and Accuracy and F1 Score (-0.08). This shows it has a lower robustness to variability compared to the other models. On the other hand, we have the performance of the LinearSVC Model under baseline and perturbed conditions. The results highlight the sensitivity of the LinearSVC model to input variability and provide insights into its robustness in real-world scenarios. The figure 4.10 illustrates topographic maps of EEG microstates under baseline and perturbed conditions, showing the spatial distribution of brain activity. The baseline maps represent the original neural activity patterns, while the perturbed maps demonstrate changes introduced through data augmentation. Despite the added variability, the core spatial characteristics of the baseline maps are retained. The perturbations increase diversity in the data, simulating real-world variability such as biological and technical inconsistencies. This highlights the effectiveness of augmentation in expanding the range of training data, improving model robustness, and enabling better generalization to unseen scenarios. Each row corresponds to a specific frequency band, while columns within a condition show the maps for five microstates labeled A, B, C, D, and C'. Perturbed maps are labeled similarly, with a "(P)" suffix, to indicate their altered nature.

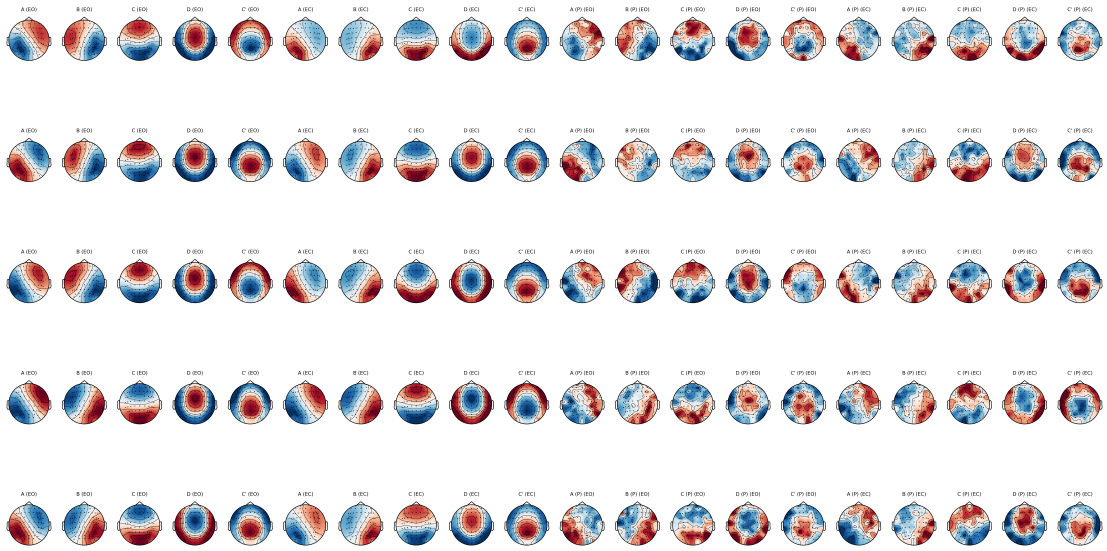
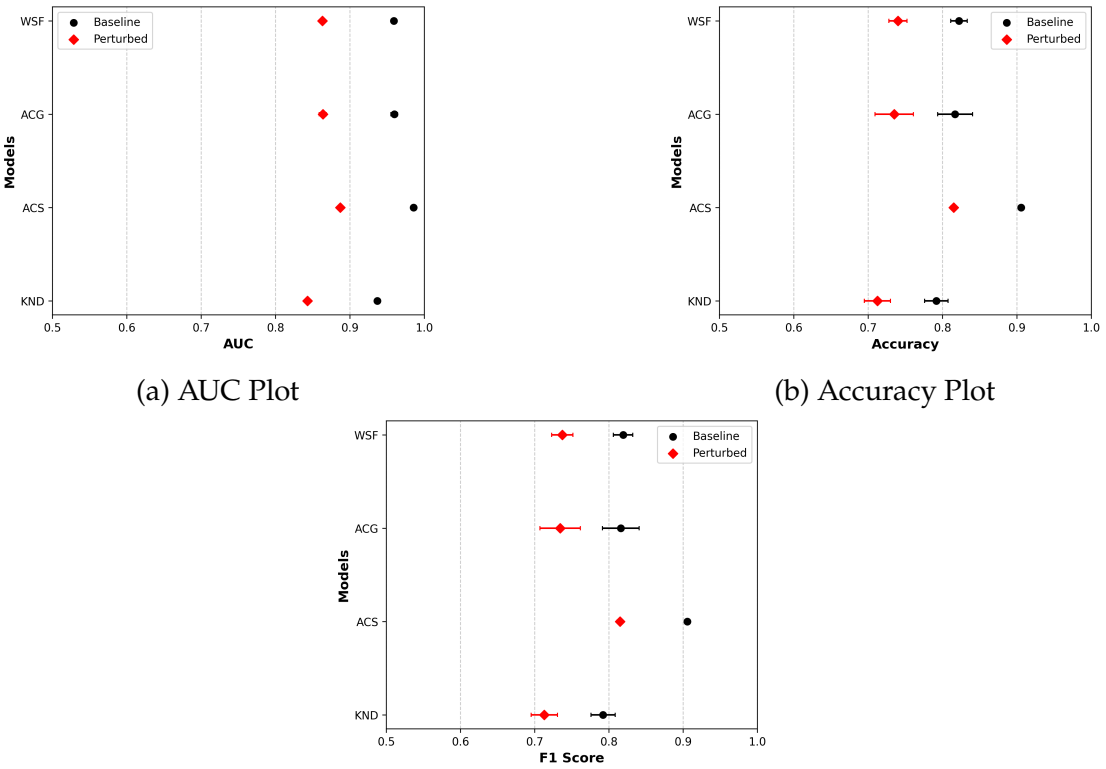


Figure 4.10: Topographic Maps of Baseline and Perturbed Conditions

For the baseline: Accuracy ranges from 0.834 to 0.884, indicating high performance under controlled conditions. Perturbed: Accuracy drops significantly to 0.3540.430, reflecting a substantial degradation in performance. This means that the model is highly sensitive to perturbations. The F1 scores are closely aligned with accuracy values, ranging from 0.833 to 0.884, reflecting balanced performance between precision and recall, while for the perturbed: F1 scores decline to 0.3410.419. The perturbed F1 scores suggest reduced reliability in predicting the positive class, further highlighting the model's vulnerability to input variability. The consistently high accuracy and F1 scores demonstrate that LinearSVC performs well in idealized conditions with clean and precise data. A significant drop across both accuracy and F1 scores underlines the model's inability to generalize when faced with variability or noise. The perturbed performance suggests a lack of inherent robustness, making the model unsuitable for applications where input variability is expected. This analysis demonstrates that while LinearSVC excels in idealized scenarios with clean data, its performance degrades significantly under perturbed conditions. Data augmentation and variability-aware methods, as implemented in ACS, WSF, and ACG, play a crucial role in improving model robustness to real-world conditions [7]. The visualization of the performance of the models under baseline and per-

turbed conditions is showed in Figure 4.11. The analysis provided several key



(a) AUC Plot

(b) Accuracy Plot

(c) F1 Score Plot

Figure 4.11: Baseline vs. Perturbed Performance Metrics for Different Models. Results for measuring the impact of IV on the performance of data augmentation-based and data imprecisiation-based models, for both baseline (that is, non-perturbed) and IV perturbed data.

insights, where perturbed metrics systematically exhibited lower mean values, demonstrating reduced performance under stressed conditions. Also exhibiting increased uncertainty. Perturbed metrics had higher standard deviations, resulting in wider confidence intervals and indicating greater variability. Confidence intervals for the differences in Accuracy between conditions. Visualizations comparing baseline and perturbed metrics for all models. This methodology provided a systematic framework to evaluate the robustness of machine learning models, particularly in handling variability and uncertainty.

5

Conclusions and Future Works

This thesis work explored the impact of instantial variability on the classification accuracy and robustness of machine learning models, focusing on EEG microstate data. The results demonstrated that advanced classifiers incorporating variability techniques, such as data augmentation and probabilistic labeling, performed better than traditional models in handling variability and uncertainty.

Augmentation effectively increased the variability of EEG microstate features without distorting their central tendencies, enhancing the generalization of machine learning models to unseen conditions. Regarding the performance of the models, among the tested classifiers: the Augmented Support Vector Classifier (ACS-a) consistently achieved the highest performance across all metrics (AUC: 0.9855 ± 0.0009, Accuracy: 0.9054 ± 0.0032, F1 Score: 0.9054 ± 0.0031) under baseline and perturbed conditions, demonstrating robustness. The Weighted Resampling Forest (WSF-i) and Augmented Gradient Boosting Classifier (ACG-a) showed competitive performance, with WSF exhibiting slightly higher stability under perturbations. The Imprecise KNeighbors Classifier (KND-f) displayed moderate performance but was less robust to perturbations. All models exhibited performance degradation under perturbations, with the ACS-a showing the least decline. This underscores the importance of incorporating uncertainty-aware methods for robust real-world applications. In summary, the research highlights the significance of integrating variability-aware techniques and uncertainty quantification to improve the robustness and reliability of machine learning models, particularly in applications involving biosignals like EEG mi-

crostates.

While this study provides promising insights into handling variability and uncertainty in EEG microstate classification, there are several areas that require further improvements. These include, incorporating larger, more diverse datasets to assess the generalizability of the proposed methods across varied populations and conditions. Also including data from clinical populations to evaluate the robustness of classifiers in pathological conditions.

Another aspect to be considered would be the application of deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), for EEG microstate analysis. An important focus could be incorporating Bayesian deep learning techniques to improve uncertainty quantification in EEG microstate analysis. Additionally, the simulation of variability can be refined by developing more advanced augmentation methods that better replicate real-world noise and variability. Exploring alternative probabilistic representations for labels, such as mixture models or adaptive uncertainty quantification approaches, could further enhance the modeling of uncertainty and variability in the data.

Future work could involve applying the proposed framework to real-time applications like brain-computer interfaces (BCIs) and clinical decision support systems. It could also be explored in personalized healthcare to tailor treatments based on patient-specific EEG patterns. Incorporating advanced evaluation metrics, such as calibration error and task-specific measures, could improve understanding of model reliability. Additionally, adapting the framework to other fields, such as speech recognition or sensor data analysis, would test its versatility in handling variability and uncertainty.

This study lays the groundwork for developing variability-aware, uncertainty-quantifying machine learning models, emphasizing their importance in robust EEG microstate analysis. Future efforts should focus on advancing the proposed methods to bridge the gap between experimental and real-world applications, ultimately improving the reliability of machine learning models in clinical and neuroscientific domains.

References

[1] Aasne K Aarsand et al. "The Biological Variation Data Critical Appraisal Checklist: A Standard for Evaluating Studies on Biological Variation". In: *Clinical Chemistry* 64.3 (Mar. 2018), pp. 501–514. ISSN: 0009-9147. DOI: 10.1373/clinchem.2017.281808. eprint: https://academic.oup.com/clinchem/article-pdf/64/3/501/32642222/clinchem0501.pdf. URL: https://doi.org/10.1373/clinchem.2017.281808.

[2] Moloud Abdar et al. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges". In: *Information Fusion* 76 (2021), pp. 243–297. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2021.05.008. URL: https://www.sciencedirect.com/science/article/pii/S1566253521001081.

[3] Elena Antonova et al. "EEG microstates: Functional significance and short-term test-retest reliability". In: *Neuroimage Reports* 2.2 (Mar. 2022), p. 100089. DOI: 10.1016/j.ynirp.2022.100089. URL: https://doi.org/10.1016/j.ynirp.2022.100089.

[4] Anahit Babayan et al. "A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults". In: *Scientific Data* 6.1 (Feb. 2019). DOI: 10.1038/sdata.2018.308. URL: https://doi.org/10.1038/sdata.2018.308.

[5] Marília Barandas et al. "Evaluation of uncertainty quantification methods in multi-label classification: A case study with automatic diagnosis of electrocardiogram". In: *Information Fusion* 101 (2024), p. 101978. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2023.101978. URL: https://www.sciencedirect.com/science/article/pii/S1566253523002944.

[6] Verena Brodbeck et al. "EEG microstates of wakefulness and NREM sleep". In: *NeuroImage* 62.3 (May 2012), pp. 2129–2139. DOI: 10.1016/j.neuroimage.2012.05.060. URL: https://doi.org/10.1016/j.neuroimage.2012.05.060.

[7] Andrea Campagner et al. "Everything is varied: The surprising impact of instantial variation on ML reliability". In: *Applied Soft Computing* 146 (2023), p. 110644. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2023.110644. URL: https://www.sciencedirect.com/science/article/pii/S1568494623006622.

[8] Pablo Casas, Christophe Mues, and Huan Yu. *A Distributionally Robust Optimisation Approach to Fair Credit Scoring*. 2024. arXiv: 2402.01811 [cs.LG]. URL: https://arxiv.org/abs/2402.01811.

[9] Probal Chaudhuri. "C R Rao and Mahalanobis distance". In: *Proceedings - Mathematical Sciences* 130.1 (July 2020). DOI: 10.1007/s12044-020-00586-4. URL: https://doi.org/10.1007/s12044-020-00586-4.

[10] Pierpaolo Croce et al. "EEG microstates associated with intra- and inter-subject alpha variability". In: *Scientific Reports* 10.1 (Feb. 2020). DOI: 10.1038/s41598-020-58787-w. URL: https://doi.org/10.1038/s41598-020-58787-w.

[11] Gabriel Chaves de Melo, Gabriela Castellano, and Arturo Forner-Cordero. "A procedure to minimize EEG variability for BCI applications". In: *Biomedical Signal Processing and Control* 89 (2024), p. 105745. ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2023.105745. URL: https://www.sciencedirect.com/science/article/pii/S1746809423011783.

[12] Tiehang Duan et al. *Distributionally robust cross subject EEG decoding*. Sept. 2023. DOI: 10.3233/faia230318. URL: https://doi.org/10.3233/faia230318.

[13] Tiehang Duan et al. "UNCER: A framework for uncertainty estimation and reduction in neural decoding of EEG signals". In: *Neurocomputing* 538 (2023), p. 126210. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2023.03.071. URL: https://www.sciencedirect.com/science/article/pii/S0925231223003156.

[14] Didier Dubois, Henri Prade, and Sandra Sandri. *On Possibility/Probability transformations*. Jan. 1993, pp. 103–112. DOI: 10.1007/978-94-011-2014-2\{_}10. URL: https://doi.org/10.1007/978-94-011-2014-2_10.

[15] Joakim Ekström. "Mahalanobis' Distance Beyond Normal Distributions". In: *UCLA* (Sept. 2011). URL: https://escholarship.org/content/qt24w7k7m1/qt24w7k7m1.pdf?t=mn3reo.

[16] Victor Férat et al. "Beyond broadband: Towards a spectral decomposition of electroencephalography microstates". In: *Human Brain Mapping* 43.10 (Mar. 2022), pp. 3047–3061. DOI: 10.1002/hbm.25834. URL: https://doi.org/10.1002/hbm.25834.

[17] Duarte Folgado et al. "Explainability meets uncertainty quantification: Insights from feature-based model fusion on multimodal time series". In: *Information Fusion* 100 (2023), p. 101955. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2023.101955. URL: https://www.sciencedirect.com/science/article/pii/S1566253523002713.

[18] Sarah J. Gascoigne et al. *Seizure Onset Zone Localisation Algorithms with Intracranial EEG: Evaluating Methodological Variations and Targeted Features*. 2024. arXiv: 2410.13466 [q-bio.NC]. URL: https://arxiv.org/abs/2410.13466.

[19] Allan Grønlund, Lior Kamma, and Kasper Green Larsen. "Margins are Insufficient for Explaining Gradient Boosting." In: *Neural Information Processing Systems* 33 (Jan. 2020), pp. 1902–1912. URL: https://proceedings.neurips.cc/paper/2020/file/146f7dd4c91bc9d80cf4458ad6d6cd1b-Paper.pdf.

[20] Mahboobeh Jafari et al. "Empowering precision medicine: AI-driven schizophrenia diagnosis via EEG signals: A comprehensive review from 20022023". In: *Applied Intelligence* 54.1 (Dec. 2023), pp. 35–79. DOI: 10.1007/s10489-023-05155-6. URL: https://doi.org/10.1007/s10489-023-05155-6.

[21] Maciej Jedynak et al. "Variability of Single Pulse Electrical Stimulation Responses Recorded with Intracranial Electroencephalography in Epileptic Patients". In: *Brain Topography* 36.1 (Dec. 2022), pp. 119–127. DOI: 10.1007/s10548-022-00928-7. URL: https://doi.org/10.1007/s10548-022-00928-7.

[22] Ivo Pascal de Jong, Andreea Ioana Sburlea, and Matias Valdenegro-Toro. *Uncertainty Quantification in Machine Learning for Biosignal Applications – A Review*. 2023. arXiv: 2312.09454 [eess.SP]. URL: https://arxiv.org/abs/2312.09454.

[23] M. Kanimozhi and R. Roselin. "Statistical Feature Extraction and Classification using Machine Learning Techniques in Brain-Computer Interface". In: *International Journal of Innovative Technology and Exploring Engineering* 9.3 (Jan. 2020), pp. 1354–1358. DOI: 10.35940/ijitee.k2343.019320. URL: https://doi.org/10.35940/ijitee.k2343.019320.

[24] Thomas Koenig et al. "EEG-Meta-Microstates: Towards a more objective use of Resting-State EEG Microstate findings across studies". In: *Brain Topography* (July 2023). DOI: 10.1007/s10548-023-00993-6. URL: https://doi.org/10.1007/s10548-023-00993-6.

[25] Mario Michael Krell, Anett Seeland, and Su Kyoung Kim. "Data augmentation for Brain-Computer Interfaces: Analysis on Event-Related Potentials data". In: *arXiv (Cornell University)* (Jan. 2018). DOI: 10.48550/arxiv.1801.02730. URL: https://arxiv.org/abs/1801.02730.

[26] Eric Liu, Cidnee Luu, and Lyndia C. Wu. "Resting State EEG variability and implications for interpreting clinical effect sizes". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 32 (Jan. 2024), pp. 587–596. DOI: 10.1109/tnsre.2024.3355956. URL: https://doi.org/10.1109/tnsre.2024.3355956.

[27] K.L. Lopez et al. "Stability, change, and reliable individual differences in electroencephalography measures: A lifespan perspective on progress and opportunities". In: *NeuroImage* 275 (2023), p. 120116. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2023.120116. URL: https://www.sciencedirect.com/science/article/pii/S1053811923002628.

[28] Constanza L Andaur Navarro et al. "Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review". In: *BMJ* (Oct. 2021), n2281. DOI: 10.1136/bmj.n2281. URL: https://www.bmj.com/content/375/bmj.n2281.

[29] J. Oosterhoff and W. R. Van Zwet. *A note on contiguity and hellinger distance*. Nov. 2011, pp. 63–72. DOI: 10.1007/978-1-4614-1314-1\{_}6. URL: https://doi.org/10.1007/978-1-4614-1314-1_6.

[30] Mario Plebani, Andrea Padoan, and Giuseppe Lippi. "Biological variation: back to basics". In: *Clinical Chemistry and Laboratory Medicine (CCLM)* 53.2 (Jan. 2015). DOI: 10.1515/cclm-2014-1182. URL: https://doi.org/10.1515/cclm-2014-1182.

[31] Benjamin A. Seitzman et al. "Cognitive manipulation of brain electric microstates". In: *NeuroImage* 146 (Oct. 2016), pp. 533–543. DOI: 10.1016/j.neuroimage.2016.10.002. URL: https://doi.org/10.1016/j.neuroimage.2016.10.002.

[32] Andrea Seveso et al. "Ordinal labels in machine learning: a user-centered approach to improve data validity in medical settings". In: *BMC Medical Informatics and Decision Making* 20.S5 (Aug. 2020). DOI: 10.1186/s12911-020-01152-8. URL: https://doi.org/10.1186/s12911-020-01152-8.

[33] Thanh-Tung Trinh et al. "Identifying Individuals With Mild Cognitive Impairment Using Working Memory-Induced Intra-Subject Variability of Resting-State EEGs". In: *Frontiers in Computational Neuroscience* 15 (2021). ISSN: 1662-5188. DOI: 10.3389/fncom.2021.700467. URL: https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2021.700467.

[34] Hristos Tyralis and Georgia Papacharalampous. "A review of predictive uncertainty estimation with machine learning". In: *Artificial Intelligence Review* 57.4 (Mar. 2024). DOI: 10.1007/s10462-023-10698-8. URL: https://doi.org/10.1007/s10462-023-10698-8.

[35] Neeraj Wagh et al. *Evaluating Latent Space Robustness and Uncertainty of EEG-ML Models under Realistic Distribution Shifts*. 2022. arXiv: 2209.11233 [eess.SP]. URL: https://arxiv.org/abs/2209.11233.

[36] Haoran Wang et al. "Cross-Subject Assistance: Inter- and Intra-Subject Maximal Correlation for Enhancing the Performance of SSVEP-Based BCIs". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (Jan. 2021), pp. 517–526. DOI: 10.1109/tnsre.2021.3057938. URL: https://doi.org/10.1109/tnsre.2021.3057938.

[37] Jing Xiang et al. "Kurtosis and skewness of high-frequency brain signals are altered in paediatric epilepsy". In: *Brain Communications* 2.1 (Jan. 2020). DOI: 10.1093/braincomms/fcaa036. URL: https://doi.org/10.1093/braincomms/fcaa036.

[38] Bihui Yu et al. *Human-computer Interaction for Brain-inspired Computing Based on Machine Learning And Deep Learning: A Review*. 2024. arXiv: 2312.07213 [cs.AI]. URL: https://arxiv.org/abs/2312.07213.

[39] Kunkun Zhao et al. "Intra-Subject and Inter-Subject Movement Variability Quantified with Muscle Synergies in Upper-Limb Reaching Movements". In: *Biomimetics* 6.4 (2021). ISSN: 2313-7673. DOI: 10.3390/biomimetics6040063. URL: https://www.mdpi.com/2313-7673/6/4/63.

[40] Kai Zheng, Pui Cheong Fung, and Xiaofang Zhou. "K-nearest neighbor search for fuzzy objects". In: *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2010* (June 2010), pp. 699–710. DOI: 10.1145/1807167.1807243. URL: https://doi.org/10.1145/1807167.1807243.

Acknowledgments

I would like to sincerely thank my supervisor, Prof.Giulia Cisotto, for the guidance and support throughout this journey. Her expertise, patience and thoughtful feedback have been invaluable during this thesis work.

Declaration

I hereby declare that all contents and analyses presented in this thesis are my original work. To enhance the clarity and readability of the text, I utilized OpenAIs ChatGPT for assistance with text refinement and proofreading. This support was strictly limited to linguistic improvements and did not affect the originality, integrity, technical and scientific content of this thesis.