Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

University of South Bohemia

Faculty of Science

Artificial Intelligence and Data Science, M.Sc.

ADHD Detection with EEG Signals

Submitted for the course **Information Theory and Feature Engineering**

Professor: Doc. Ing. Ivo Bukovský, Ph.D.

Submitted by:	María Isabel Sánchez-O'Mullony Martínez
Student ID:	B24763
Date:	November 11, 2025

Supervisor: Doc. Ing. Ivo Bukovský, Ph.D.

Declaration

I declare that I have written this report by myself and have only used the sources and aids mentioned, and that I have marked direct and indirect citations as such. This report has not been submitted prior for any other examination.

I agree that the results of this study work / report may be used free of charge for research and lecturing purposes.

Contents

List of Abbreviations

ADHD	Attention-Deficit/Hyperactivity Disorder
EEG	Electroencephalography
DC	Direct Current

1 Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a prevalent neurodevelopmental condition that affects individuals across the lifespan. According to Furman [1], ADHD is not a single disease entity but rather a constellation of symptoms representing a final common behavioral pathway for a range of emotional, psychological, and learning difficulties.

Although the behavioral manifestations of ADHD are well-documented, its neurophysiological underpinnings remain an active area of research. Electroencephalography (EEG) has been widely used to study ADHD, as it provides a non-invasive measure of brain electrical activity. Previous studies have identified characteristic EEG patterns in individuals with ADHD, particularly abnormalities in the frontal and central regions, often reflected as altered spectral power or atypical signal complexity.

This study aims to investigate differences in EEG activity between healthy individuals and those diagnosed with ADHD, using the dataset provided by Sadeghi Bajestani *et al.* [2]. Specifically, the project focuses on (1) comparing the statistical and informational properties of EEG signals across both groups, (2) engineering features inspired by information theory—such as entropy and mutual information—to quantify signal complexity and connectivity, and (3) building a predictive model capable of distinguishing ADHD from healthy controls based on these features. This work bridges the domains of signal processing, feature engineering, and information theory to improve understanding and potential classification of ADHD-related brain activity.

2 Data Description

The dataset [2] comprises EEG recordings from a total of 79 adult participants, including 42 healthy controls and 37 individuals diagnosed with ADHD (age range: 20-68 years; male/female ratio: 56/23). EEG signals were recorded under four experimental conditions: (1) resting state with eyes open, (2) resting state with eyes closed, (3) a cognitive challenge task, and (4) an auditory task involving listening to an "omni harmonic" sound stimulus. Recordings were obtained from five scalp locations—O1, F3, F4, Cz, and Fz—at a sampling rate of 256 Hz. These regions encompass occipital and frontal areas known to play key roles in attentional control and executive function.

File organization and structure. The data are provided as four MATLAB `.mat` files, corresponding to the experimental groups:

- `FC.mat` - female control group
- `MC.mat` - male control group
- `FADHD.mat` - female ADHD group
- `MADHD.mat` - male ADHD group

Each file contains a 1×11 cell array, where each cell represents one experimental task or condition. Within each cell, the data are stored as a three-dimensional matrix with dimensions `[subjects×samples×channels]`. For instance, a typical entry of size 13×7680×2 indicates 13 participants, 7680 time samples (corresponding to 30 seconds of EEG data at 256 Hz), and two recorded channels.

The specific configuration of each cell (i.e., channel pair and duration) is summarized as follows:

Cell	Condition	Channels	Duration (s)
1	Eyes open baseline	Cz, F4	30
2	Eyes closed	Cz, F4	20
3	Eyes open	Cz, F4	20
4	Cognitive challenge	Cz, F4	45
5	Pre-Omni harmonic baseline	Cz, F4	15
6	Omni harmonic assessment	Cz, F4	30
7	Eyes open baseline	O1, F4	30
8	Eyes closed	O1, F4	30
9	Eyes open	O1, F4	30
10	Eyes closed	F3, F4	45
11	Eyes closed	Fz, F4	45

Table 1: Summary of EEG tasks, channel pairs, and recording durations.

One corrupted signal (subject 7 of the female ADHD group) was identified and excluded from further analysis.

This hierarchical data organization allows for flexible analysis across multiple dimensions: by gender, diagnosis, condition, and channel pair. For subsequent preprocessing, each task will be reshaped into individual subject-channel recordings, ensuring consistent sampling durations across conditions.

3 Data Visualization

3.1 Data Loading and Structure

To analyze EEG signals from the ADHD dataset [2], the recordings were imported from MATLAB `.mat` files using the `scipy.io.loadmat` library in Python. Four files were available, corresponding to each experimental group: `FC.mat` (female control), `MC.mat` (male control), `FADHD.mat` (female ADHD), and `MADHD.mat` (male ADHD). Each file contained a 1×11 cell array, where each cell represented a specific experimental task and stored EEG data as a three-dimensional matrix with dimensions `[subjects x samples x channels]`.

A custom loading function (`load_group`) was implemented to iterate through each cell, extract signal arrays, and store them in a structured `pandas DataFrame`. Each row in the resulting dataset corresponds to a single subject-task pair and includes the following metadata: participant group (Control/ADHD), gender, task number, subject identifier, and a two-channel EEG signal array. This organization facilitates subsequent preprocessing, visualization, and feature extraction. The corrupted EEG file corresponding to subject 7 of the female ADHD group was identified and excluded from analysis.

3.2 Signal Visualization by Group and Task

To inspect data quality and explore inter-subject variability, raw EEG signals were visualized for selected participants using `matplotlib`. Figure 1 presents the recordings from the first five female control subjects during Task 1 (eyes-open baseline). Each subplot represents one subject, showing the two recorded channels.

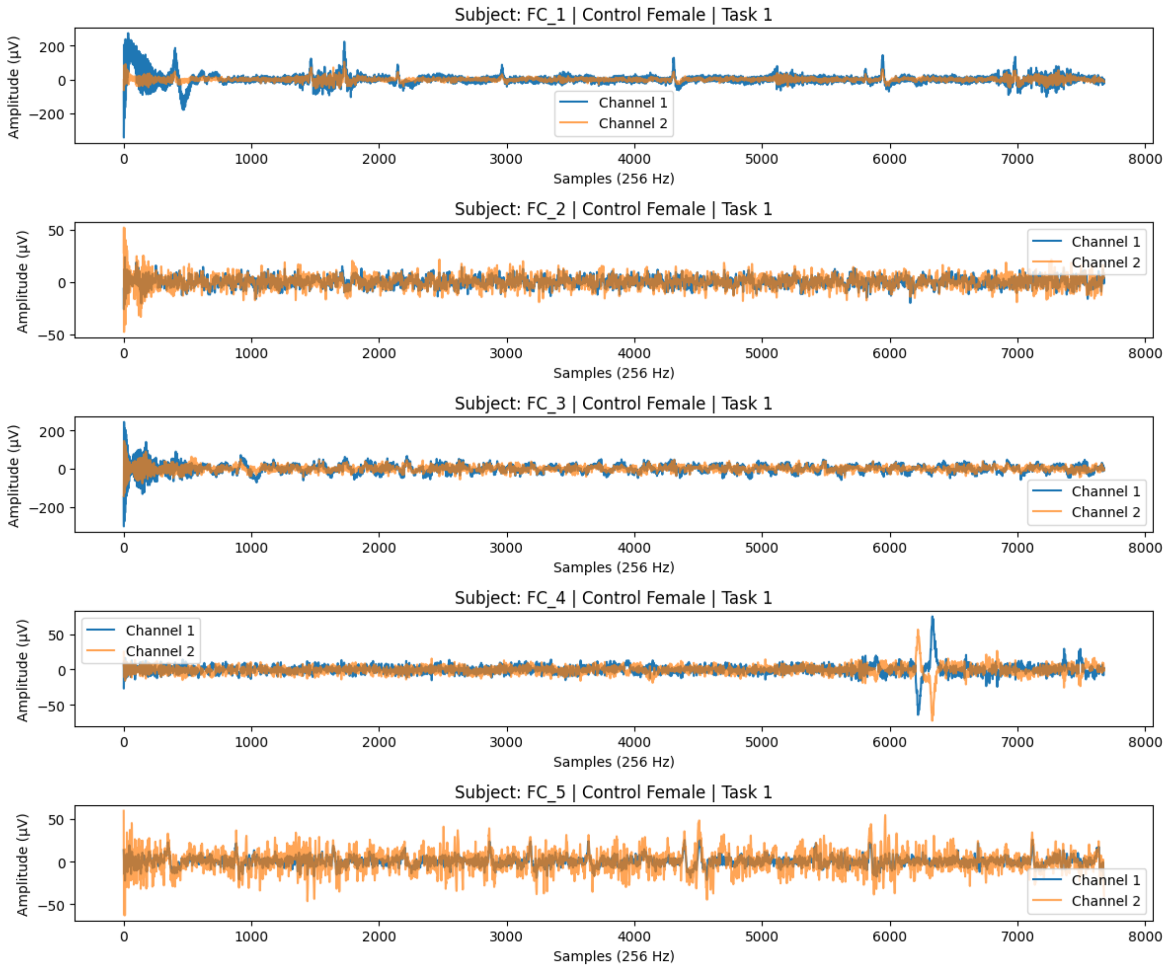

Figure 1: EEG signals from five female control participants during Task 1 (eyes-open baseline). Each subplot represents two simultaneously recorded channels.

As illustrated in Figure 1, the EEG traces display characteristic oscillatory patterns within an amplitude range of approximately $\pm 200\ \mu$V. Variations across participants are evident, with some showing strong low-frequency drifts or transient spikes, likely associated with ocular or muscular artifacts. Despite these variations, both channels exhibit similar temporal trends, suggesting functional coupling between the recorded electrode sites. These observations emphasize the need for further preprocessing steps, such as band-pass filtering (0.5-45 Hz), baseline correction, and normalization, to ensure consistent signal quality across participants.

3.3 Comparison Between ADHD and Control Groups

A complementary visualization was conducted to qualitatively compare EEG signals between ADHD and control subjects under the same task conditions. Figure 2 shows five pairs of male participants (Control vs. ADHD) performing Task 1, for Channel 0.

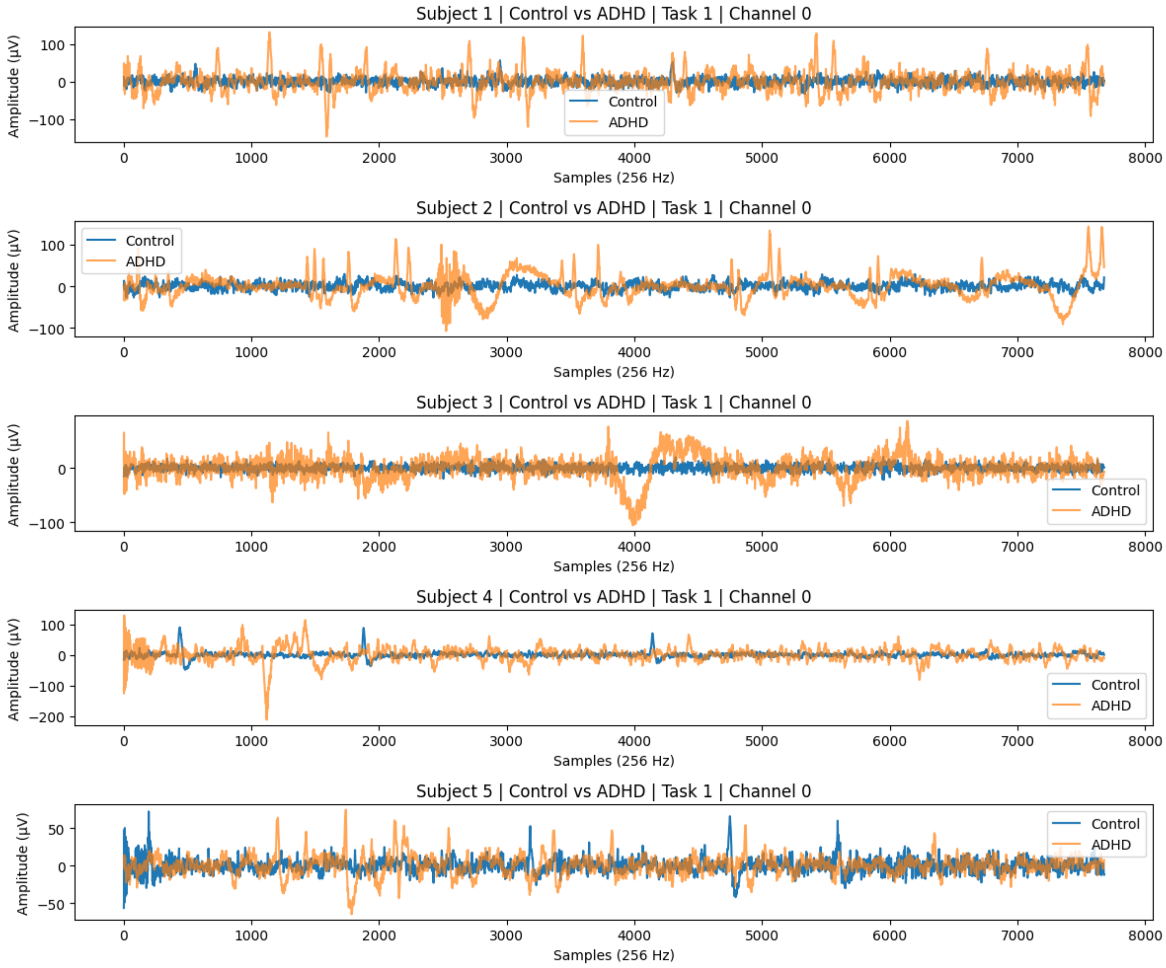

Figure 2: Comparison of EEG signals between control and ADHD male participants during Task 1 for Channel 0. Each subplot shows a pair of subjects (Control vs. ADHD).

In general, ADHD recordings exhibit greater amplitude fluctuations and reduced rhythmic stability compared to control subjects, who display smoother oscillatory activity. This increased irregularity suggests higher signal entropy and reduced neural synchronization in ADHD participants. Such qualitative patterns are consistent with prior research reporting atypical cortical activity and impaired oscillatory regulation in ADHD populations [3]. These visual differences motivate the extraction of quantitative features— particularly information-theoretic measures such as entropy and mutual information—to characterize signal complexity and connectivity.

3.4 Summary of Observations

The exploratory visualization phase revealed several important insights:

- EEG signals contain observable low-frequency drifts and high-amplitude artifacts, supporting the need for filtering and normalization.
- There is substantial inter-subject variability in signal amplitude and frequency content, even within the same condition.
- ADHD participants tend to exhibit higher signal variability and irregularity, indicating potential discriminative features for subsequent classification.

These findings informed the design of the preprocessing and feature extraction pipeline, where statistical, spectral, and information-theoretic features were computed to enable quantitative group comparison and machine learning-based classification.

4 Preprocessing

EEG data are highly sensitive to non-neural artifacts arising from muscle activity, eye blinks, and small head or body movements. Because the dataset used in this study was recorded non-invasively from scalp electrodes, such artifacts are inevitable and must be carefully mitigated to preserve the validity of the neural information [4]. The objective of the preprocessing stage was therefore to reduce non-cerebral noise and standardize the signals prior to feature extraction.

4.1 Detrending

EEG recordings were detrended by subtracting the mean amplitude of each channel to remove Direct Current (DC) offsets and slow baseline drifts [5]. EEG recordings often contain a small DC offset — a constant voltage bias caused by electrode-skin impedance and amplifier drift. This offset shifts the entire signal above or below zero microvolts without carrying any neural information. This step prevents artificial low-frequency power and ensures that subsequent analyses reflect genuine neural oscillations. Although the visual appearance of the waveforms remains largely unchanged, detrending is essential to prevent bias in later spectral and entropy-based feature extraction.

To verify the effectiveness of the detrending procedure, the mean amplitude of each EEG channel was computed before and after baseline correction. As shown below in 2, the mean values were reduced to approximately zero following detrending, confirming successful removal of the DC offset while preserving the oscillatory structure of the signal.

	Channel 1	**Channel 2**
Mean before detrending (μV)	1.5269	0.0936
Mean after detrending (μV)	-1.74×10^{-15}	1.18×10^{-16}

Table 2: Verification of mean amplitude before and after detrending.

The results confirm that the detrending operation correctly centered the EEG signals around zero microvolts, thereby eliminating slow baseline drifts without altering the underlying temporal dynamics. This ensures that subsequent frequency and entropy analyses are not biased by non-neural voltage offsets.

References

[1] L. Furman, "What is attention-deficit hyperactivity disorder (adhd)?" *Journal of child neurology*, vol. 20, no. 12, pp. 994–1002, 2005.

[2] G. Sadeghi Bajestani, S. Abedian, F. Makhloughi, M. Raoufitabar, and H. Saeedi, "A dataset of eeg signals from adults with adhd and healthy controls: Resting state, cognitive function, and sound listening paradigm," 2023. [Online]. Available: https://doi.org/10.17632/6k4g25fhzg.1

[3] A. Lenartowicz and S. K. Loo, "Use of eeg to diagnose adhd," *Current psychiatry reports*, vol. 16, no. 11, p. 498, 2014.

[4] A. Chaddad, Y. Wu, R. Kateb, and A. Bouridane, "Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques," *Sensors*, vol. 23, no. 14, p. 6434, 2023.

[5] A. De Cheveigné and D. Arzounian, "Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data," *NeuroImage*, vol. 172, pp. 903–912, 2018.