Software Engineering Department
Braude College

Capstone Project Phase A

Order Pattern Matching for ADHD analysis using EEG waves

25-1-R-1

Lecturers Guide:
Dr. Anat Dahan
Dr. Idrees Ghazawi Samah

Students Submit:
Tamir Nahari, id: 318189941
Yaakov Shitrit, id:304959422

Contents

1. **Abstract**...3

2. **Introduction** 3

3. **The objective of the project** 3

4. **Background and related work** 4

4.1 What is EEG and How Does it Work? 4

4.2 What Are EEG Frequencies
and What Do They Represent? 4

4.3 Processing and Cleaning EEG Data 6
4.3.1 Low-Pass filter...6
4.3.2 High-Pass filter 7
4.3.3 ICA 7
4.3.4 FFT 7

4.4 Pattern Matching..8

4.5 Order Preserving Matching...9

5. **Advantages of OPM**...10

6. **Dataset Description** 11

7. **Literature review** 12

8. **Our proposed solution**..15

9. **Process flow chart**...19

10. **Verification plan** 20

11. **References** 21

1. Abstract

This study examines whether there are unique and recurring patterns of brain electrical activity among children aged 7-12 diagnosed with ADHD (Attention Deficit Hyperactivity Disorder) using EEG (Electroencephalography) technology. The research focuses on identifying distinctive electrical activity patterns that differentiate children with ADHD from those without the disorder. Pattern matching techniques, specifically the Order Preserving Pattern Matching algorithm, are applied to achieve this goal. The study aims to determine the existence of these unique patterns, which could serve as a foundation for developing future tools for faster and more accurate ADHD diagnosis.

2. Introduction

In this project, we analyze EEG data collected from 121 children aged 7-12, comprising 61 children diagnosed with ADHD and 60 children in a control group without the disorder. An experienced psychiatrist conducted ADHD diagnoses according to DSM-IV criteria, and all participants in the ADHD group had been treated with Ritalin within six months before the study [5,6]. The control group consisted of children with no history of psychiatric disorders, epilepsy, or high-risk behaviors [6].

The data were gathered during a visual attention task designed to assess one of the primary cognitive deficits in children with ADHD—visual attention [5]. In this task, the children viewed a sequence of cartoon images and were instructed to count the number of characters in each image. The number of characters per image ranged from 5 to 16, and the images were presented in rapid succession without pausing after the child's response. The recording duration was tailored to the response speed of each child [5,6].

EEG recordings were conducted using 19 channels adhering to the 10-20 system, with a sampling rate of 128 Hz [4,7](see Figure 1 for 10-20 system). The A1 and A2 electrodes on the earlobes were reference electrodes [4].

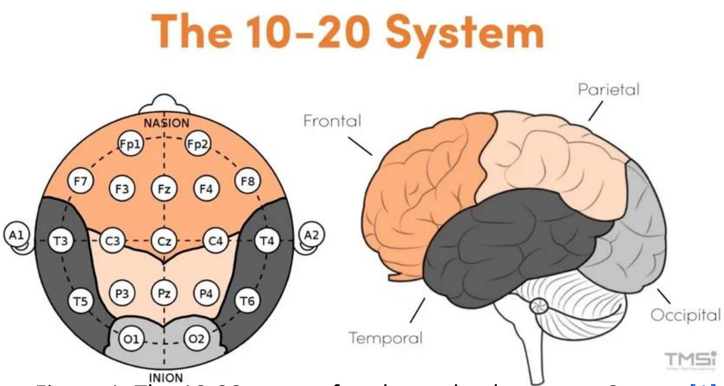


Figure 1: The 10-20 system for electrode placement. Source:[1]

3. The objective of the project

This project aims to analyze the collected data, identify unique wave patterns associated with children diagnosed with ADHD, and examine the differences between the two groups. To achieve this, we propose classifying the waves into sequences of letters representing different wave types while addressing several challenges. One challenge involves the removal of electrical noise, as EEG signals often include external interference such as muscle movements and eye blinks, which must be eliminated to enhance the quality of the analysis. Another challenge is the identification of oscillations as waves, requiring a clear definition of how specific oscillations, such as alpha, beta, or gamma waves, are classified based on their frequencies and physiological characteristics. Additionally, pattern detection will involve using the Order Preserving Matching (OPM) algorithm to locate recurring patterns in the recorded signals. The OPM algorithm will be applied to detect and compare sequential patterns across different time windows, facilitating a more accurate classification of ADHD-related EEG patterns. A further step involves determining how these identified patterns can be utilized to create models or derive practical insights. Finally, the project will examine whether specific patterns can serve as markers for ADHD, enabling the classification of children with the disorder in comparison to those in the control group. This research aims to offer new insights into the relationship between EEG patterns and attention-deficit/hyperactivity disorder, leveraging advanced computational tools to tackle these challenges.

4. Background and related work

To explore how unique brain activity patterns in children with ADHD can be identified, we first delve into the principles of EEG and its potential to detect changes in electrical brain patterns.

4.1 What is EEG and how does it work?

Electroencephalography (EEG) measures the brain's electrical activity using electrodes attached to the scalp (see Figure 2 for electrode placement). It captures voltage changes caused by neuronal activity, primarily in the cerebral cortex. This method provides high temporal resolution, making it ideal for studying rapid brain processes, but its spatial resolution is limited due to overlapping signals from different brain regions [4].

EEG is non-invasive, safe, and widely used for medical and research purposes. It records various brain waves, classified by frequency, including delta, theta, alpha, beta, and gamma waves, each associated with distinct mental states and activities [6,7]. Applications of EEG range from diagnosing neurological conditions such as epilepsy, sleep disorders, and ADHD to advancing cognitive research and developing brain-computer interfaces [4,5]. While EEG is cost-effective and accessible, it is sensitive to external noise and less effective at detecting activity in deep brain structures, yet it remains a cornerstone in understanding brain function [4].

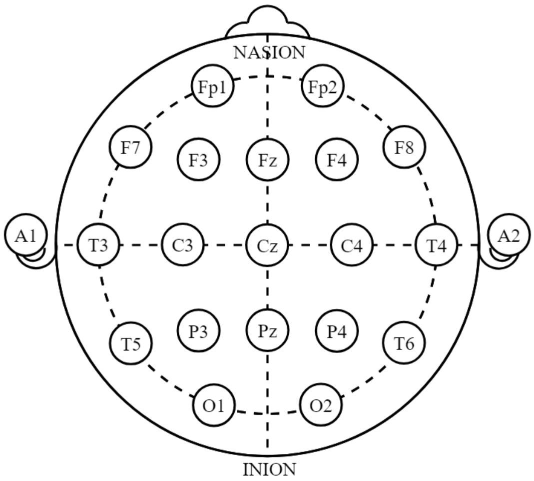


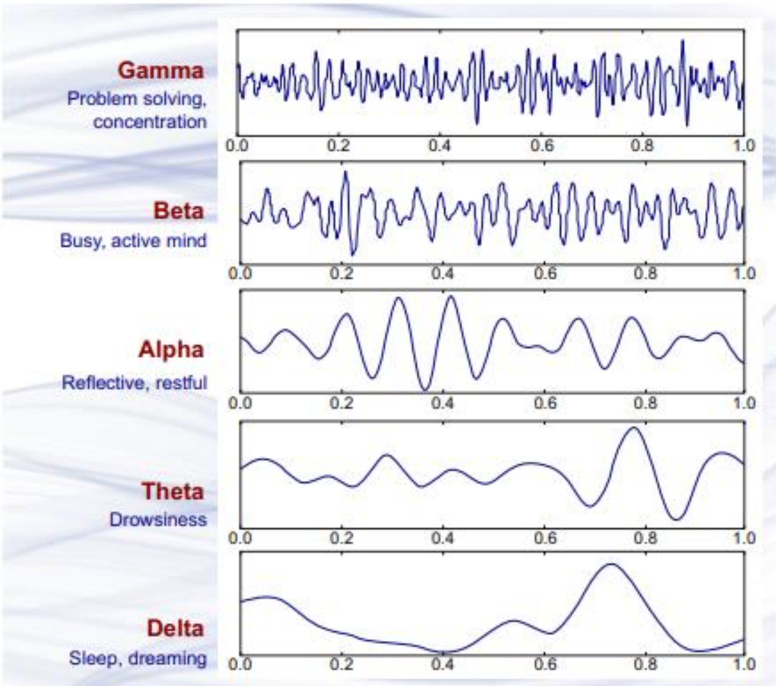
Figure 2: electrodes of International 10-20 system for EEG [2[

The functionality of EEG relies heavily on accurately identifying different brainwave frequencies, which represent a wide range of mental states.
Next, we examine the key characteristics of these frequencies.

4.2 What are EEG frequencies and what do they represent?

EEG captures electrical brain waves, classified into distinct frequency bands that reflect various brain states and activities. These frequencies provide critical insights into both clinical conditions and cognitive processes [5,6](see Figure 3 for The different frequencies). Delta waves, with low frequencies between 0.5 and 4 Hz, are typically associated with deep sleep, while theta waves, ranging from 4 to 8 Hz, are linked to states of relaxation or meditation [5,7]. Alpha waves, in the 8–12 Hz range, are observed during calm but alert states, and beta waves, with frequencies of 12–30 Hz, are connected to concentration and stress. Gamma waves, above 30 Hz, are associated with complex cognitive functions such as information processing [5,6].

figure 3: source[14] Brain wave samples with dominant frequencies of beta, alpha, theta, and delta bands and gamma waves.



These frequency bands play a central role in medical diagnoses, including epilepsy, sleep disorders, and ADHD, where EEG helps identify abnormal brain activity patterns [4,6]. In scientific research, EEG is instrumental in exploring cognitive functions like memory, attention, and decision-making [5,7]. By analyzing these frequencies, EEG provides a window into the dynamic activity of the brain and its underlying processes [5,6].

Understanding brain wave frequencies is just one part of the process; to achieve meaningful and accurate analysis, it is essential to address background noise and other interfering factors, requiring advanced signal cleaning and processing techniques.

4.3 Processing and cleaning EEG data

The processing and cleaning of EEG data is a critical phase in the research process, as these signals are highly sensitive to external noise and interference that can distort the essential information related to brain activity. Such noise includes eye blinks, muscle movements, electrical disturbances, and uncontrolled environmental factors [4,5]. Consequently, meticulous cleaning and processing techniques are required to enhance signal quality and ensure the most accurate data possible [6]. One of the primary methods for cleaning EEG signals is the application of digital filters, such as low-pass filters for removing high-frequency noise unrelated to brain activity and high-pass filters to eliminate low-frequency components like head movements or general electrical oscillations [7]. Another essential process involves the use of Independent Component Analysis (ICA), an advanced method for separating signal components, enabling the isolation and removal of noise sources such as eye blinks or muscle artifacts from the original signal [4,5]. Additionally, to ensure comparability between data collected from different subjects, normalization processes are applied, including subtracting the mean and dividing by the standard deviation of the data. This standardization allows for a more consistent comparison of measurements across subjects [6].

Beyond cleaning, signal processing involves decomposing EEG data into frequency components, representing distinct brain waves such as delta, theta, alpha, beta, and gamma. This decomposition is achieved using mathematical techniques such as Fast Fourier Transform (FFT) or Wavelet Transform, which enable the identification of frequency components within the signal and the analysis of their relative power [7].

4.3.1 Low-Pass Filter:

A low-pass filter allows low-frequency components of the EEG signal to pass through while attenuating (reducing) high-frequency components that may introduce unwanted noise. This is particularly useful for eliminating high-frequency artifacts such as electrical interference from equipment and muscle activity. By applying a low-pass filter, we can focus on the brain's natural wave frequencies, such as delta, theta, alpha, beta, and gamma, without interference from high-frequency noise.

4.3.2 High-Pass Filter:

A high-pass filter works in the opposite way, allowing high-frequency signals to pass while removing low-frequency components such as baseline drift caused by slow movements of the head or body. This filtering technique is beneficial for eliminating slow oscillations and ensuring that critical neural signals within the target frequency bands remain intact.

4.3.3 Independent Component Analysis (ICA):

ICA is an advanced statistical method used to separate mixed signals into independent components, allowing us to identify and remove artifacts from EEG data. For example, ICA can isolate and eliminate noise sources such as eye blinks, muscle activity, and heartbeats from the EEG recordings. This method works by assuming that the recorded EEG signals are a mixture of brain activity and noise, and it attempts to mathematically separate these components into independent sources based on their statistical properties.

4.3.4 Fast Fourier Transform (FFT):

FFT is a powerful mathematical technique used to decompose EEG signals into their constituent frequency components. This allows us to analyze the power and presence of different brain wave frequencies, such as alpha (8–12 Hz) and beta (12–30 Hz). By transforming the EEG signal from the time domain to the frequency domain, FFT helps identify patterns associated with cognitive and neurological states, such as attention levels in ADHD patients.

After performing FFT, each time window of the EEG signal is broken down into its frequency components, with each frequency (e.g., alpha, beta, gamma) assigned a numerical value representing its power (Power Spectral Density - PSD). These extracted values create a unique sequence for each frequency across different time windows. This data can be directly utilized in the second proposed approach, where we analyze frequency-based sequences over time windows of 8 seconds with a 1-second overlap.

In this approach, after applying FFT, we generate lists of frequency values for each participant and each EEG channel, maintaining their temporal dynamics. The Order Preserving Matching (OPM) algorithm can then be applied to the sequences of frequency values to detect recurring patterns that distinguish between the ADHD group and the control group.

This method allows us to focus on the unique characteristics of EEG frequency bands over time while reducing the influence of noise and individual differences among participants.

These processes provide the foundation for in-depth EEG data analysis and facilitate the identification of recurring or unique patterns that can yield significant insights into the brain activity of subjects with ADHD compared to the control group [5,7].

After processing the data, the next step involves identifying recurring patterns within the cleaned signals, which may reveal differences between children with ADHD and those in the control group.

4.4 Pattern matching

Pattern matching is a computational technique used to identify specific sequences or structures within a dataset(see Figure 4 for structure example). EEG signal analysis involves detecting recurring patterns in the brain's electrical activity that may correspond to specific cognitive or neurological states [4,7]. These patterns can be useful for classifying subjects, identifying biomarkers for disorders, or understanding the underlying brain mechanisms[5].

The process typically involves several steps:

1. **Data Representation**: EEG signals are converted into a representation suitable for pattern matching, such as sequences of numerical values, binary states, or symbolic labels [4,6].
2. **Pattern Specification**: A predefined pattern or a set of patterns is identified based on prior knowledge or learned through machine learning techniques [4,7].
3. **Search and Comparison**: The algorithm scans the EEG data to locate segments that match the specified pattern. Depending on the signal variability, the comparison may involve exact or approximate matching [7].
4. **Evaluation**: Matched patterns are analyzed to assess their significance, frequency, or association with specific events or conditions [5,6].

Pattern matching is widely used in EEG analysis for tasks like event detection, classification of neurological disorders, and identification of abnormal brain activity [4,7].

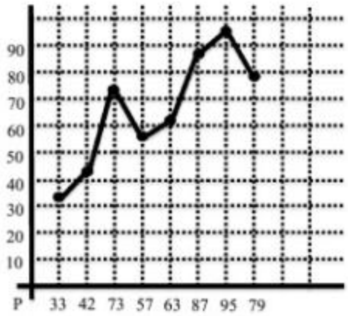


figure 4: The x-axis describes the indices of the pattern P.

The y axis describes the value at each index [8]

Traditional approaches to pattern detection in EEG signals often focus on absolute values. However, meaningful information frequently lies in the relative order between different points on the signal graph. For instance, consecutive rises and falls in amplitude can reflect specific brain activity, even when absolute values vary between subjects or recordings.

While pattern recognition is a useful technique, applying Order Preserving Matching (OPM) allows for a more precise focus on the relative structures within EEG signals, making it particularly suitable for analyzing complex and diverse data.

4.5 Order preserving matching

Order Preserving Matching (OPM) is a specialized form of pattern matching that focuses on the relative order of elements within a sequence rather than their exact values. This approach is particularly useful in applications where the order of changes carries more meaning than the specific magnitudes, such as in EEG signals [8,12].

How OPM Works:

1. **Sequence representation**: A numerical sequence, such as an EEG signal, is transformed into a representation that captures the relative ordering of its values. For example, in a sequence (3, 1, 4, 2), the order can be represented as (2, 4, 1, 3), where each number corresponds to its rank in the original sequence [10,12].
2. **Pattern definition**: The pattern to be matched is defined based on its relative order. For example, the pattern (1, 3, 2) represents an order where the first element is the smallest, the third is intermediate, and the second is the largest [10].
3. **Matching process**: The algorithm scans the data and checks if the relative order of elements in any subsequence matches the defined pattern. Variations in the magnitude of values are ignored as long as the order is preserved [11,12].
4. **Applications**: In EEG analysis, OPM can be used to detect specific patterns of brain wave activity that repeat over time(see Figure 5 for The detection of specific

pattern) or across subjects, regardless of variations in signal amplitude. This makes it robust to individual differences and noise [5,7,11].

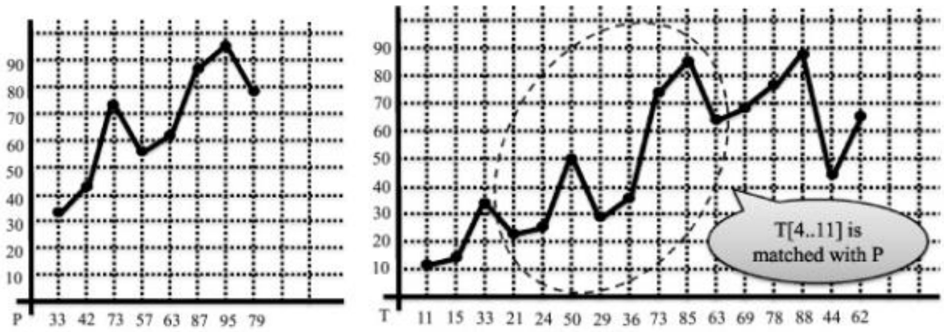


figure 5: The pattern P, on the left, matches the subpattern T[4,11] of the text T, on the right. [8]

5. Advantages of OPM

- **Robustness to noise**: Since OPM focuses on order rather than exact values, it is less sensitive to amplitude variations and baseline shifts [12,13].
- **Suitability for EEG data**: EEG signals often exhibit variability across subjects or conditions, making OPM an ideal choice for identifying consistent patterns in noisy data [4,7].

- **Personalized Analysis of EEG Data:**
  Handling Differences Between and Within Individuals:
  One of the main challenges in analyzing EEG signals is the high variability between different people and tasks. This research suggests a method that reduces the effect of these differences by focusing on patterns that repeat themselves, regardless of the signal's strength. This approach helps to consistently identify biological markers (biomarkers) that could be used to diagnose ADHD [16,17].

- **Development of Rapid Diagnostic Tools:**
  The research utilizes tools such as **Wavelet Transform** [16] and **FFT** [19] to decompose the signals into distinct frequency bands. This decomposition enables the identification of relevant waves (such as Theta and Beta), which are critical for ADHD detection [18]. By recognizing distinctive patterns, the study has the potential to lead to the development of clinical diagnostic tools based on intelligent algorithms, offering fast processing times and high accuracy.

- **Focus on Theta/Beta Waves:**
  The research reinforces findings on the **Theta/Beta Ratio (TBR)** [5] as a clinical marker and introduces new methods to assess it accurately, even in the presence of external noise. Unlike the use of FFT and Wavelet Transform alone, the integration of pattern recognition, such as OPM, overcomes the limitations of external noise. This is achieved by allowing pattern detection to remain effective despite noise interference.

By leveraging the OPM technique, researchers can uncover subtle but meaningful patterns in EEG signals that might otherwise be obscured by noise or individual variability, providing insights into neurological conditions such as ADHD [6,7].

6. Dataset description

In this project, we will use EEG data to analyze brain activity patterns associated with ADHD. The objective is to identify unique markers that differentiate children diagnosed with ADHD from typically developing children. To achieve this, we will utilize a dataset containing EEG recordings from 121 children. Below is a detailed description of the dataset.

The dataset consists of EEG recordings from 121 children, divided into two groups:

- **ADHD group**: 61 children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD) according to DSM-IV criteria. These children were under Ritalin treatment for up to 6 months before the recording.
- **Control group**: 60 typically developing children with no history of psychiatric disorders, epilepsy, or high-risk behaviors.

Recording Details:

- **Channels**: EEG signals were recorded from 19 electrodes placed according to the International 10-20 system. The electrode locations are: Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, and O2. Reference electrodes (A1 and A2) were placed on the earlobes.
- **Sampling frequency**: 128 Hz.
- **Duration**: EEG duration varied depending on the child’s response time during the tasks.
- **Filtering**: Signals were preprocessed to remove noise and artifacts.

Experimental Protocol:

- Participants completed a visual attention task. Images of cartoon characters were presented, and the children were instructed to count the number of characters in each image.
- The number of characters in the images ranged randomly from 5 to 16, ensuring that the task was engaging and visually accessible to the participants.
- EEG data collection was performed during the task, with each image displayed following the participant's response, creating a dynamic and interactive recording session.

Purpose:

The dataset is designed to study brain connectivity, attention mechanisms, and the non-linear dynamics of EEG signals. It provides a valuable resource for investigating diagnostic methods and computational approaches for ADHD [15].

7. Literature review

The following section presents a review of articles that provide insights and methodologies relevant to the dataset used in this study. Each paper is summarized to highlight its contribution, relevance to EEG analysis, and potential applications in ADHD diagnostics or pattern matching in biological data. This review aims to provide a foundation for understanding the theoretical and practical approaches employed in the dataset's analysis.

1. Detecting ADHD children using attention continuity as a nonlinear feature of EEG

This study investigates how to identify children with ADHD by analyzing their brain activity using EEG, which records electrical signals from the brain. The researchers focused on patterns in these signals that are not easy to see with regular methods. They used a type of advanced analysis, called nonlinear features, to measure the complexity and changes in brain activity. By comparing the brain signals of children with ADHD to those without, they found clear differences, especially in the front part of the brain, which is responsible for attention and decision-making. Using a computer program to sort the data, they were able to accurately identify ADHD in 96.7% of cases. This research helps show how brainwave patterns can be used to diagnose ADHD more precisely. The main difference between this paper and our approach lies in the method used to analyze EEG data. while this study relies on nonlinear features, our study relies on OPM [5].

2. Investigating the Discrimination of Linear and Nonlinear Effective Connectivity Patterns of EEG Signals

This study examined the differences in brain connectivity patterns between children with ADHD and typically developing children using EEG (electroencephalography). The researchers used a new method called nCREANN, which identifies complex relationships between brain regions, alongside a simpler method, dDTF, that analyzes direct connections between regions. They found that children with ADHD have fewer connections in specific brain areas compared to those without the disorder, with these connections mostly appearing on the left side of the brain. Using this data, the researchers achieved highly accurate classification (99%) of ADHD cases through computer algorithms.

This research relates to our project as both analyze EEG data to identify patterns associated with ADHD. While their study focuses on brain connectivity, our project highlights recurring signal patterns. These complementary approaches can provide valuable insights into understanding and diagnosing ADHD [7].

3. **Direction of Information Flow Between Brain Regions in ADHD and Healthy Children Based on EEG**

This study explores how information flows between different parts of the brain in children with ADHD compared to children without the disorder. Using EEG and a method called Directed Phase Transfer Entropy (dPTE), the researchers found that children with ADHD have disrupted information flow in specific brain waves. For example, in healthy children, information flows more smoothly from the back to the front of the brain, especially in certain wave frequencies, but this pattern was weaker or reversed in children with ADHD. This connects to our research, as both studies focus on identifying unique brain patterns in ADHD using EEG [6].

4. **EEG Classification of ADHD and Normal Children Using Non-linear Features and Neural Network**

The study investigates how electrical brain activity, recorded using EEG, can help notice children diagnosed with ADHD from those without the disorder. The researchers analyzed EEG signals from 30 children with ADHD and 30 healthy children by applying advanced mathematical methods to capture the complexity and chaos in the brain's activity. Specifically, they extracted non-linear features, such as fractal dimension and approximate entropy, which describe the irregular and unpredictable nature of the brain's electrical patterns. These features were then used to train a type of artificial intelligence called a neural network, which achieved a high accuracy (over 93%) in classifying ADHD and non-ADHD children.

This research relates to our project because both studies aim to identify unique patterns in EEG signals to differentiate between ADHD and non-ADHD individuals. While their study uses non-linear features and neural networks for classification, our research focuses on identifying recurring patterns in the EEG data using the Order Preserving Matching (OPM) algorithm [4].

Key Difference in Research Approaches:
The first four studies mentioned focus on advanced analysis of EEG databases, utilizing techniques such as non-linear features, neural networks, or analysis of information flow between brain regions. Their primary goal is to uncover and explain complex connections within the data to gain insights into the brain activity of children with ADHD compared to a control group.

In contrast, our project adopts a fundamentally different approach. Rather than decoding or interpreting the database itself, we focus on identifying **recurring patterns** within the EEG data using the Order Preserving Matching (OPM) algorithm.

In the following articles, we focus on string matching techniques, particularly order-preserving pattern matching, and their potential applications in the analysis of biological datasets, including EEG signals.

5. **Alternative Algorithms for Order-Preserving Matching**

This conference presented a collection of studies focusing on algorithmic techniques for pattern recognition, particularly order-preserving pattern matching. Although not directly applied to EEG data, the theoretical foundations discussed in the conference offer valuable insights for adapting such techniques to biological data analysis. Filtering algorithms and efficient pattern detection methods described in the conference could significantly enhance the processing and analysis of noisy EEG signals [13].

6. **An Encoding for Order-Preserving Matching**

The research focuses on developing an efficient method to find patterns in data sequences by matching the relative order of their elements, rather than their exact values. This technique is useful in scenarios where the pattern's structure is more important than specific numerical values, such as in time-series data. This aligns with our project, as we aim to identify recurring EEG signal patterns that differentiate children with ADHD from the control group. [8].

7. **A Filtration Method for Order-Preserving Matching**

This study introduces an efficient approach to finding patterns in numerical sequences where the relative order of elements matters. This method converts data into a simplified binary format that captures increases and decreases between numbers, making it faster to locate patterns using string-matching algorithms. By focusing on relative order rather than exact values, the approach is effective for analyzing time-series data, like financial or temperature records.

This connects to our ADHD project because both studies use order-preserving matching to identify patterns within complex datasets. Our project applies these principles to EEG data to detect unique brainwave patterns in children with ADHD [10].

8. **Order-Preserving Pattern Matching with Scaling**

This study introduces a new approach to pattern matching called scaled order-isomorphism, which identifies patterns in sequences based on their relative order and scaling (adjusting lengths between elements), rather than focusing on exact values. The paper presents an efficient algorithm to solve this problem with a time complexity comparable to the best-known exact OPPM algorithms. This method is particularly useful for analyzing time-series data, such as EEG signals, where natural variations in signal length and amplitude are common. [12].

We want to use string-matching techniques, like order-preserving pattern matching, to analyze our EEG data. These methods help find repeating patterns in the signals and handle noise well. By using them, we hope to discover important patterns that could help identify ADHD.

8. Our proposed solution

After reviewing the existing methods and addressing the challenges in identifying unique brain activity patterns in children with ADHD, we propose a solution based on the Order Preserving Matching (OPM) algorithm. This approach enables the detection of recurring patterns in brain waves, focusing on the relative order between signal values, which helps address the natural variations in EEG signals. We will now elaborate on the principles of this solution and how it can be implemented to tackle the challenges outlined effectively.

Proposed Solution:

The dataset includes **60 EEG recordings from children diagnosed with ADHD** and **61 EEG recordings from children without ADHD,** each recording consists of **19 channels**.

The dataset noises will be cleaned by Filtering and ICA methods and after that, we will normalize the dataset.

We propose using the **Order Preserving Pattern Matching (OPM)** algorithm to identify recurring patterns in the EEG signals. We will try to find patterns that distinguish the group of children with ADHD.
We also aim to identify patterns that are unique to children without ADHD. If such a pattern is found and confirmed to appear consistently in the non-ADHD group, its absence in a subject's EEG data could serve as a potential indicator of ADHD. This approach allows us to draw meaningful conclusions and enhances the accuracy of ADHD diagnosis by focusing on the presence or absence of specific distinguishing patterns.

Input:

- Dataset-EEG Signals of all the 121 recordings.

Output:

- ADHD-specific patterns
- Non-ADHD-specific patterns*

We will consider two approaches:

1)

For every record and every channel do:
Divide each channel into different time units t(0.5s,1s,2s..)

Each time unit will be represented as a pattern P for the OPM algorithm.
Each P pattern will be compared to all other 121*t parts in the same channel for the 121 student's recordings using OPM.

For any P pattern which has a match:
Sum all the matches in ADHD group defined by P(ADHD)
Sum all the matches in Non-ADHD group defined by P(Non-ADHD)
divide each group by total sum and check which one greater than 0.9.

$$IOU = \frac{P(ADHD)}{P(ADHD)+P(Non-ADHD)} > 0.9 \text{ OR } \frac{P(Non-ADHD)}{P(ADHD)+P(Non-ADHD)} > 0.9$$

The IOU metric compares the prediction region (bounding box or mask) to the ground truth region by calculating the ratio of the intersection area to the union area of the two regions. A minimum IOU threshold (e.g., 0.5 or 0.7) can be set to determine whether a particular prediction is considered a "success". In medical tasks, it is customary to measure with a higher threshold between 0.8-0.9, so we will use the stricter approach of a threshold of 0.9.

2)

In this approach, we will try to emphasize examining patterns at separate frequencies.

For every record and every channel do:
Divide each channel into time units of windows $t_i's$ of 8 seconds with 1 sec overlap

Divide each channel into different frequencies units (α,β,γ...) using FFT.
Construct a list of values of all windows $t_i's$ in a specific frequency (α,β,γ...) using Fourier series values, for all frequencies in every channel.

Each sequence length (3,4,5,...):

Will be represented as a pattern P for the OPM algorithm.

Each P pattern will be compared to all other t_i parts in the same frequency in all 19 channels for each 121 recordings using OPM.

For any P pattern which has a match define:

The sum of all matches in ADHD group is defined by P(ADHD)

The sum of all matches in Non-ADHD group is defined by P(Non-ADHD)

divide each group by total sum and check which one greater then IOU>0.9.

$$\frac{P(ADHD)}{P(ADHD)+P(Non-ADHD)} > 0.9 \text{ OR } \frac{P(Non-ADHD)}{P(ADHD)+P(Non-ADHD)} > 0.9$$

Further Analysis:

Important questions to answer:

1. Can we see based on the sets what channels are significant for ADHD?
2. What channels are significant for non-ADHD recordings?

Definition of Research Success:

The success of our research will be determined by identifying two distinct groups of recurring EEG patterns and a success rate of 90 percent or more based on the IOU metric we mentioned above:

1. **ADHD-specific patterns:** Patterns that occur in children with ADHD **at least 90% of the time**, relative to their total occurrences across both groups.
2. **Non-ADHD-specific patterns:** Patterns that occur in children without ADHD **at least 90% of the time**, relative to their total occurrences across both groups.

Once these patterns are identified, they will undergo further validation to ensure their significance within the respective groups. Specifically, a pattern will be considered successful if it meets the following criteria:

- The pattern must appear in **at least 90% of the recordings** within the identified group.

Example Calculation:

If a pattern P is found 58 times across all recordings:

- 56 occurrences are found in ADHD group recordings.
- 2 occurrences are found in non-ADHD group recordings.

To check if the pattern represents ADHD, we calculate the proportion:

$$\frac{56}{58} = 0.965 \quad \text{—>} \quad 96\% > 90\%$$

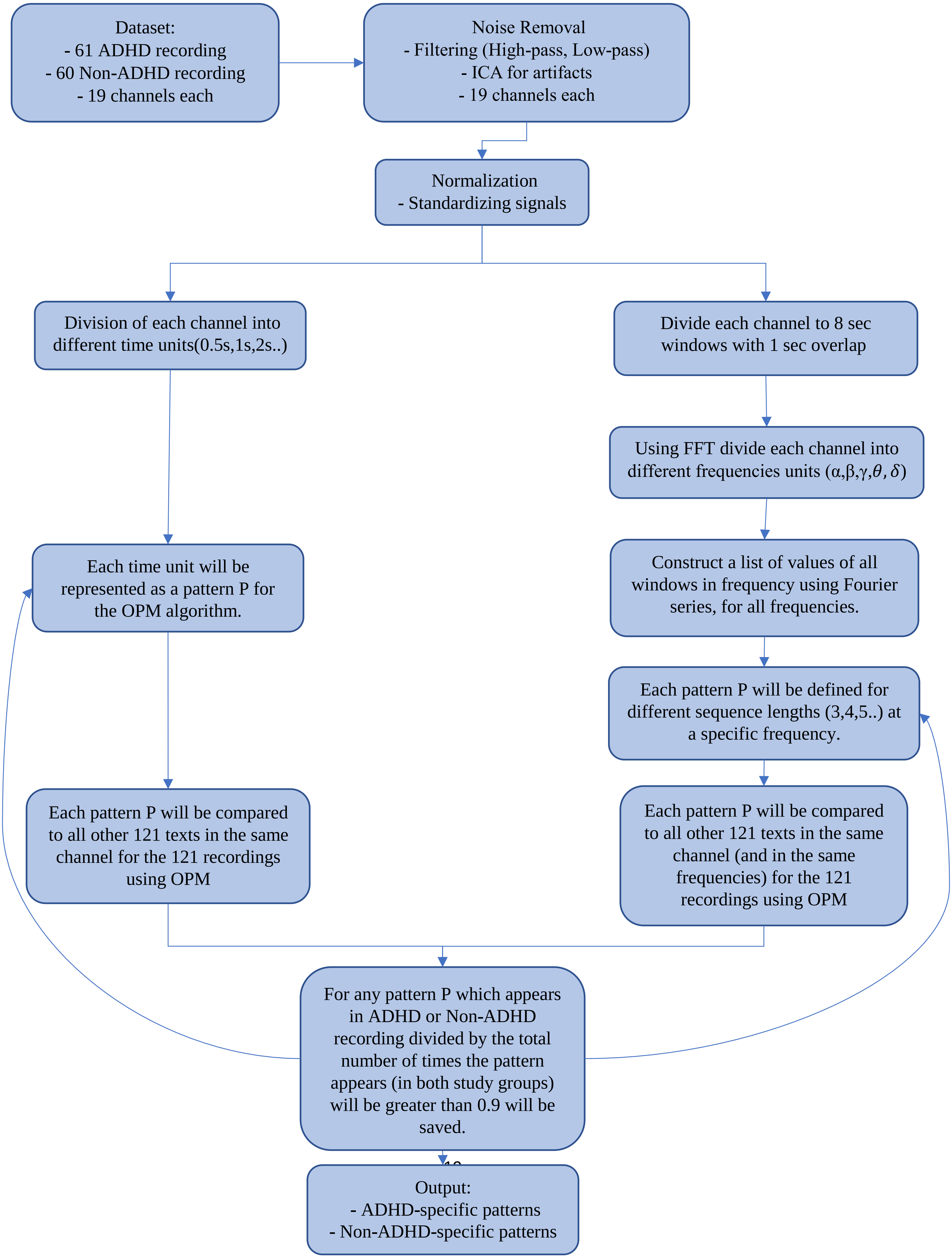
Since the pattern appears more than 90% of the time in ADHD recordings, we proceed to check its presence within the ADHD group as a whole:

$$\frac{56}{61} = 0.91 \quad \text{—>} \quad 91\% > 90\%$$

Since the pattern appears in more than 90% of ADHD recordings, it will be considered **a successful ADHD-specific pattern**.

Identifying such patterns is crucial as they provide valuable insights into the timing, frequency, and occurrence of ADHD-related brain activity. These patterns can contribute to developing more effective diagnostic tools for ADHD.

9. Process flow chart



10. Verification plan

case no.	Step	Action Description	Expected Result	Success Criteria
1	Noise removal and filtering	Apply high-pass and low-pass filters and remove artifacts using ICA for all recordings and channels.	Signals are filtered, and noise are minimized	No noise remain, and signals retain their integrity for further processing.
2	Normalization	Standardize all signals across channels and recordings to prepare them for analysis.	All signals are normalized to the same scale [0, 1]	Signals are comparable across channels and recordings with no scaling inconsistencies.
3	Frequency decomposition using FFT	Apply FFT to divide each window into specific frequency bands (alpha, beta, gamma, theta, delta).	Each window is represented by its frequency components across all specified bands.	Accurate frequency decomposition for all windows in all channels.
4	OPM pattern comparison	Compare each pattern (P) across all other texts using OPM.	Patterns are consistently compared to identify significant matches across recordings.	OPM accurately matches patterns and preserves order across recordings.
5	A filtered wave enters as input to the normalization algorithm.	Between every 2 consecutive values in the array, we verify that the ratio between them is equal to their normalized ratio.	All values in the array are normalized.	

11. References

[1] Source: TMSi, *The 10-20 System for EEG*, Retrieved from https://info.tmsi.com/blog/the-10-20-system-for-eeg.
[2] Nowicka, I. (n.d.). 21 electrodes of International 10-20 system for EEG [Image]. Wikimedia Commons. Retrieved from https://commons.wikimedia.org/wiki/File:21_electrodes_of_International_10-20_system_for_EEG.svg

[3]Arns, M., Conners, C. K., & Kraemer, H. C. (2013). A decade of EEG theta/beta ratio research in ADHD: A meta-analysis. *Neuroscience Letters, 560*, 96–100. https://www.sciencedirect.com/science/article/pii/S0304397513007585

[4]Mohammad Reza Mohammadi, et al. "EEG Classification of ADHD and Normal Children Using Non-linear Features and Neural Network." *Biomed Eng Lett*, 2016.

[5]Armin Allahverdy, et al. "Detecting ADHD Children Using the Attention Continuity as Nonlinear Feature of EEG." *Frontiers in Biomedical Technologies*, 2016.

[6]Ali Ekhlasi, et al. "Direction of Information Flow Between Brain Regions in ADHD and Healthy Children Based on EEG Using Directed Phase Transfer Entropy." Cognitive Neurodynamics, 2021.

[7]Nasibeh Talebi, et al. "Investigating the Discrimination of Linear and Nonlinear Effective Connectivity Patterns of EEG Signals in Children with ADHD." Computers in Biology and Medicine, 2022.

[8] Travis Gagie, Giovanni Manzini, and Rossano Venturini. "An Encoding for Order-Preserving Matching." Leibniz International Proceedings in Informatics, 2016.

[9]Armin Allahverdy, et al. "Order-Preserving Matching in EEG Data Analysis." Frontiers in Biomedical Technologies, 2016.

[10]Nasibeh Talebi, et al. "A Filtration Method for Order-Preserving Matching." Computers in Biology and Medicine, 2022.

[11]Kubica, Marcin, et al. "Order-Preserving Pattern Matching." Journal of Discrete Algorithms, 2013.

[12]Kim, Y., Kang, M., Na, J. C., & Sim, J. S. (2023). Order-preserving pattern matching with scaling. Information Processing Letters, 180, 106333.

[13]Chhabra, T., Külekci, M. O., & Tarhio, J. (2015). Alternative Algorithms for Order Preserving Matching. Prague Stringology Conference Proceedings, 36-45.

[14]Abhang, P. A., Gawali, B. W., Mehrotra, S. C. (2016). Chapter 2 - Technological Basics of EEG Recording and Operation of Apparatus. Introduction to EEG- and Speech-Based Emotion Recognition, 19–50.
https://www.sciencedirect.com/science/article/pii/B9780128044902000026?via%3Dihub
https://www.biorender.com/template/eeg-brain-wave-frequencies

[15]Motie Nasrabadi, A., Allahverdy, A., Samavati, M., & Mohammadi, M. R. (2023). EEG data for ADHD / Control children. IEEE DataPort.

[16] Travis Gagie, Giovanni Manzini, Rossano Venturini. Conference Paper, 2016. "An Encoding for Order-Preserving Matching".

[17] Nasibeh Talebi, Ali Motie Nasrabadi. Computers in Biology and Medicine, 2022. "Investigating the Discrimination of Linear and Nonlinear Effective Connectivity Patterns of EEG Signals in Children with Attention-Deficit/Hyperactivity Disorder and Typically Developing Children".

[18] Mohammad Reza Mohammadi, Ali Khaleghi, Ali Moti Nasrabadi, Safa Rafieivand, Moslem Begol, Hadi Zarafshan. Biomed Eng Lett, 2016. "EEG Classification of ADHD and Normal Children Using Non-linear Features and Neural Network"

[19] Guillermo Sarasa, Ana Granados, Francisco B. Rodriguez. *Computer Methods and Programs in Biomedicine*, 176 (2019), pp. 225–235. https://www.sciencedirect.com/science/article/abs/pii/S0169260718316535?via%3Dihub "Algorithmic Clustering Based on String Compression to Extract P300 Structure in EEG Signals".