Welcome everyone. Today, we'll discuss our research on improving ADHD diagnosis using machine learning and polysomnography data.

Before we dive deeper, here's a brief overview of what we'll cover in this presentation. We'll start with related works in the field, followed by our methodology, then move on to our experimental results. After that, we'll discuss the implications of our findings, and finally, we'll conclude with our conclusions and future work.

**Frame 1**

ADHD is a common neurodevelopmental disorder that affects individuals across all age groups. It manifests through symptoms such as inattention, hyperactivity, and impulsivity.

Currently, ADHD diagnosis primarily relies on subjective assessments such as clinical interviews and behavioral checklists. This approach introduces significant variability between clinicians, increasing the risk of misdiagnosis, especially since ADHD symptoms overlap with other disorders like anxiety and depression. Consequently, this can lead to delayed diagnosis and inadequate management of the condition.

To address these challenges, there's a critical need for objective biomarkers—biological indicators that can reliably diagnose ADHD. Objective biomarkers offer several advantages, including increased consistency across different clinicians, enhanced diagnostic accuracy, and the possibility of early detection, which is essential for timely intervention. Potential biomarkers for ADHD include EEG patterns, specific characteristics of sleep stages, and various physiological signals obtained from polysomnography data."

**Frame 2**

Previous researches has advanced ADHD diagnosis mainly by looking at how graph theory has been used to understand brain connectivity in ADHD, the role of multimodal biomarkers in enhancing diagnosis, and the application of machine learning techniques in this field.

Our study builds on these works by integrating multiple physiological signals—EEG, EOG, EMG, and ECG—providing a more comprehensive analysis. Additionally, we focus on sleep stage transitions, which have been largely overlooked in prior research. By applying machine learning models, we aim to create an objective, reliable diagnostic tool.

**Frame 3**

Our data was collected from 48 children at the University of Mississippi Medical Center's pediatric sleep center, including 25 ADHD patients and 23 controls. Each participant underwent an overnight polysomnography, which provided approximately 8 hours of sleep data across multiple physiological channels, including EEG, EOG, EMG, and ECG. In total, we recorded over 20,000 epochs for ADHD patients and nearly 20,000 for non-ADHD participants.

This data was segmented into five sleep stages—Wake, Stage 1, Stage 2, Stage 3-4, and REM. These stages are critical for understanding the sleep patterns and potential biomarkers of ADHD.

**Frame 4**

Data preprocessing is essential because the raw physiological signals we work with, such as EEG, EOG, EMG, and ECG, are inherently noisy. These signals often contain unwanted artifacts like eye blinks, muscle movements, and environmental noise, which can distort the data and lead to inaccurate results.

Data preprocessing is a critical step in ensuring the accuracy of our analysis. First, for the EEG channels, we applied the Reference Electrode Standardization Technique, which minimizes variability between electrode placements. This was followed by a band-pass filter to remove noise outside the relevant frequency range, and a notch filter at 60 Hz to eliminate interference from power lines.

In addition to EEG, we preprocessed signals from EOG, EMG, and ECG channels, ensuring they were filtered and synchronized with EEG for consistent analysis across all channels.

**Frame 5**

To extract meaningful features from our multimodal data, we first compute **correlation matrices** for each 30-second epoch. These matrices capture the pairwise relationships between signals from all 17 channels—including EEG, EOG, EMG, and ECG—allowing us to understand how different physiological systems interact during sleep.

We then model these interactions using **graph theory**. By treating the correlation matrices as **adjacency matrices**, we construct **weighted, undirected graphs** where:

* **Nodes** represent the various physiological sensors.
* **Edges** are weighted by the absolute value of the correlation coefficients, indicating the strength of interaction between signals.

This graph-based approach enables us to apply graph metrics to quantify the efficiency of information transfer within and between physiological systems.

One key metric we use is the **average shortest path length**, which measures the average minimum number of steps needed to connect any two nodes in the network. This metric quantifies the efficiency of information transfer in brain networks and other physiological systems:

* **Shorter average shortest path lengths** suggest more efficient connectivity and communication.
* **Longer path lengths** may indicate impairments or less efficient interactions, which could be associated with ADHD.

For each patient, we calculate this metric across all five sleep stages, resulting in a **five-dimensional feature vector**. This vector reflects the efficiency of physiological network interactions during different stages of sleep and serves as input for our machine learning models to classify ADHD and non-ADHD individuals.

**Frame 6**

After extracting features, we applied a **Random Forest classifier** to classify ADHD and non-ADHD patients. Random Forest was chosen because of its ability to handle high-dimensional data and avoid overfitting by averaging predictions from multiple decision trees.

The input to the model was a **five-dimensional feature vector** for each patient, representing physiological system interactions during different sleep stages.

To ensure the model’s reliability, we used **nested cross-validation**. In the outer loop, we split the data into five folds, and in each iteration, we trained the model on four folds and tested it on the fifth. Within each training set, we performed another 5-fold cross-validation to tune hyperparameters, such as the number of trees and the maximum tree depth, using **GridSearchCV**.

Once we identified the best hyperparameters, the model was evaluated on the independent test set.

**Frame 7**

Our Random Forest model achieved solid performance in classifying ADHD patients. The model’s average accuracy was 72%, with a precision of 71% and a recall of 85%, resulting in an F1-score of 76%. This balanced performance highlights the model’s ability to accurately identify ADHD cases while minimizing false positives.

On the left, you can see the table summarizing the performance metrics of our model across 5 folds. The key metrics are **Accuracy**, **Precision**, **Recall**, and **F1-Score**:

* The model achieved an average **accuracy of 72%**, with a **precision of 71%** and a **recall of 85%**.
* The **F1-score** averaged **76%**, indicating a good balance between precision and recall

(Precision: Out of all the cases that the model predicted as positive (ADHD in your case), how many were actually correct? A high precision means the model makes few false positive errors.

Recall: Out of all actual ADHD cases, how many did the model correctly identify? A high recall means the model makes few false negative errors.)

**Correlation Coefficients Between Sleep Stages:**

On the right, we have a **correlation matrix** that compares the sleep stage transitions between ADHD and non-ADHD individuals.

* The **non-ADHD correlations** are shown with underlines, while significant differences are highlighted with bold numbers.

Now let’s focus on two specific sleep transitions:

**Sleep Stage 1 to Sleep Stage 3-4 Transitions:**

* For ADHD individuals, the transition between **Sleep Stage 1 and Sleep Stage 3-4** is disrupted, as shown by a **negative correlation (-0.09)**. This negative value indicates that these two stages are not aligned, which may reflect difficulty in progressing from light to deep sleep.
* In contrast, **non-ADHD individuals** show a **positive correlation (0.63)** between these stages, suggesting smoother transitions into deep sleep, which is important for restorative sleep.
* A **high correlation** means when the person moves from one sleep stage to the next, the transition is smooth and consistent. In contrast, a **lower correlation** means the transition between the two stages may be disrupted or irregular.

**Frame 8**

For our Random Forest model, Stage 1 and Stage 3-4 emerged as the most important features. These stages had the highest contribution to the model's accuracy.

Permutation analysis was applied to evaluate how important **Stage 1** and **Stage 3-4** are for the model's performance. By shuffling these features—meaning we randomly mixed up the values for these sleep stages—we broke their meaningful relationships with the ADHD label.

**Frame 10**

In conclusion, our study demonstrated that machine learning can effectively predict ADHD using sleep stage data, achieving an average accuracy of 72%. **Stage 1** and **Stage 3-4** were identified as the most critical sleep stages for distinguishing between ADHD and non-ADHD individuals. **Stage 3-4**, in particular, emerged as the most important feature, suggesting that disruptions in deep sleep are key indicators of ADHD.

These findings provide valuable insights into the connection between sleep disturbances and ADHD, and could have important clinical applications for ADHD diagnosis and treatment monitoring.