**Predicting ADHD in Children Using EEG Signal Features and Machine Learning**

Student Name: Tushya Rastogi

Student ID:14586363

Supervisor Name: Dr.Omid Chatrabgoun

**Abstract**

The symptoms of Attention Deficit Hyperactivity condition (ADHD), a common neurodevelopmental condition impacting children globally, include impulsivity, hyperactivity, and inattention. A massive portion of traditional diagnostic techniques rely on behavioural assessments, which can be laborious and subjective. This work uses machine learning techniques and electroencephalogram (EEG) signals to provide an objective, data-driven method of diagnosing ADHD. We concentrate on features that are generated from EEG, such as wavelet decomposition metrics, non-linear measures, theta, alpha, and beta power, and the theta-beta ratio, which is a known biomarker for ADHD. Predict the presence of ADHD, a dataset comprising these EEG variables from kids with and without the disorder is examined using a variety of machine learning algorithms.

We use 400 EEG-based features (theta, alpha, and beta power across multiple scalp locations) in a dataset of 5,540 samples to represent brainwave activity. We then apply classification algorithms like Random Forest, Support Vector Machines (SVM), and Neural Networks to find patterns suggestive of ADHD. To find the best model for ADHD prediction, we measure model performance with criteria like as accuracy, precision, and recall.

This research aims to assess the predictive power of different EEG features and provide insights into the role of brain wave abnormalities in ADHD diagnosis. The study's findings could potentially pave the way for faster, more accurate, and objective ADHD detection, assisting clinicians in early intervention and personalized treatment plans.

**Introduction**

With an estimated 5-7% prevalence worldwide, attention deficit hyperactivity disorder (ADHD) is one of the most prevalent neurodevelopmental diseases in children. It shows up as a recurring pattern of hyperactivity, impulsivity, and inattention that has a considerable influence on day-to-day functioning and development. Traditionally, subjective behavioural tests and questionnaires have been used to diagnose ADHD. These methods are biased and may cause a delay in receiving prompt treatment. As ADHD is becoming more common, there is a growing need for objective diagnostic instruments to support clinical assessments.

A promising non-invasive method for assessing brain activity in people with ADHD is electroencephalography (EEG). Studies indicate that children diagnosed with ADHD display unique EEG patterns, especially with the strength of certain frequency bands such as theta (4-8 Hz) and beta (12-30 Hz). Notably, greater ratios generally indicate ADHD. The theta-beta ratio (TBR) has been regularly reported as a possible biomarker. Because EEG-based features capture neurophysiological abnormalities that are frequently associated with the illness, they can provide a more objective approach to diagnosis.

This study uses advanced machine learning algorithms in conjunction with EEG information to predict ADHD in youngsters. We make use of an extensive dataset that contains theta-beta ratios for distinct brain areas, wavelet decomposition metrics, power values in different frequency bands, and non-linear properties (e.g., fractal dimension and entropy). To improve diagnostic accuracy and offer an objective complement to conventional approaches, we aim to identify children as either having ADHD or not by using machine learning algorithms to these features.

**Research Question, Problem Statement, or Topic for Investigation**

**Research Question:**

How accurately can ADHD in children be predicted using EEG signal features and machine learning models, and which specific EEG features contribute most significantly to the prediction?

**Problem Statement:**

Millions of youngsters worldwide struggle with attention deficit hyperactivity disorder (ADHD), which impacts their ability to learn, interact with others, and live a generally better life. To provide prompt interventions that can lessen these negative impacts, early and precise diagnosis is essential. Unfortunately, most of the time-consuming, subjective, and biased diagnostic techniques used today mostly rely on behavioural observations and questionnaires. Inconsistent treatment suggestions and delays in diagnosis may result from this.

Objective, data-driven techniques that can increase the precision and efficacy of ADHD diagnosis are becoming more and more necessary considering the shortcomings of traditional diagnostic procedures. Because it can assess brain activity in real time and could identify neurophysiological irregularities linked to ADHD, electroencephalography (EEG) has shown promise as a diagnostic tool. Theta-beta ratio (TBR), which is a ratio of brain wave frequencies, has been specifically suggested as a biomarker for ADHD. Its efficacy in conjunction with machine learning methods, which can identify intricate patterns in huge datasets, will need to be further investigated.

**Topic for Investigation:**

This project seeks to investigate whether machine learning models can accurately predict ADHD in children based on features extracted from EEG signals. The primary focus will be on evaluating the predictive power of various EEG-derived features, such as:

* Power in different frequency bands (theta, alpha, beta),
* Theta-beta ratio (TBR),
* Wavelet decomposition metrics (e.g., D4, D3, and D2 energy),
* Non-linear features (e.g., fractal dimension, Hurst exponent, entropy).

By applying machine learning algorithms to this dataset, the project will assess how different EEG features contribute to ADHD classification and whether the inclusion of advanced EEG signal analysis techniques improves diagnostic accuracy.

**Evidence of the Problem:**

ADHD is acknowledged as a significant neurodevelopmental condition that is becoming more commonplace worldwide. Research indicates that attention deficit hyperactivity disorder (ADHD) is linked to disruptions in brain activity, namely in the frontal lobe, which manages attention, impulse control, and behaviour. EEG is a widely available, non-invasive technique for tracking brain activity, which makes it a perfect tool for looking into these anomalies.

Studies have shown that children diagnosed with ADHD typically have a higher theta-beta ratio (TBR), which suggests that slower brain frequencies are involved in tasks involving attention and control. Despite this, a diagnosis cannot be made using the TBR alone. However, because machine learning models can examine multiple EEG parameters at once, they have the potential to reveal patterns and interactions between various brainwave frequencies that would not be readily apparent through conventional analysis.

**Intended User or Group of Users and Their Requirements**

* **Clinicians and Psychologists:** Professionals involved in diagnosing and treating ADHD can benefit from more objective diagnostic tools. An accurate machine learning model could assist clinicians by providing supplementary evidence to support traditional diagnostic methods.
* **Researchers:** Neuroscientists and data scientists interested in EEG patterns and machine learning applications in health diagnostics can use the findings to further investigate brainwave-based diagnostics.
* **Parents and Guardians:** While not directly using the model, parents and guardians of children undergoing ADHD assessments could benefit from quicker, more reliable diagnoses.

**Need for the Project:** ADHD is typically diagnosed using behavioural checklists, which can lead to inconsistent or delayed diagnoses. There is a growing need for objective, quantifiable methods to support traditional diagnostic criteria. EEG-based ADHD diagnosis has shown promise in research, but the translation of these findings into clinical tools is still in its infancy. This project bridges this gap by exploring the feasibility of using EEG data and machine learning models to predict ADHD in a clinical or research setting.

**Systems Requirements, Project Deliverables, and Final Project Outcome**

**System Requirements:**

* **Data pre-processing tools:** Tools for cleaning, normalizing, and reducing the dimensionality of the dataset.
* **Machine learning framework:** A robust platform for model development, such as Python with libraries like Scikit-learn, TensorFlow, or PyTorch.
* **Evaluation tools:** Tools for assessing model performance using confusion matrices, ROC curves, and other classification metrics.
* **Computational resources:** Adequate computing power to handle a dataset with 400 features and 5,540 samples, including the possibility of cloud computing resources if necessary.

**Project Deliverables:**

* **Dataset preprocessing pipeline:** A reproducible workflow for cleaning and preparing EEG data for machine learning, including normalization and feature selection techniques.
* **Trained machine learning models:** A set of trained models (Random Forest, SVM, Neural Network) with performance evaluations based on accuracy, precision, recall, and F1-score.
* **Evaluation report:** A comprehensive report comparing the performance of the different models, including justifications for the best-performing model.
* **Documentation and code:** Detailed documentation and open-source code for reproducibility and further research.

**Final Project Outcome:** The project aims to deliver a machine learning model that accurately predicts ADHD based on EEG brainwave features. The outcome will contribute to the growing body of knowledge around objective diagnostic methods for ADHD and could serve as the foundation for future clinical applications or research into brainwave-based diagnostics.

**Primary Research Plan**

**Plan Overview:** The research plan involves multiple stages, including data preprocessing, model development, evaluation, and validation. Each stage is designed to ensure the model's performance is maximized while maintaining clinical relevance.

**Research Method:**

* **Data Preprocessing:**
  + Clean and normalize the EEG dataset to ensure consistency across all samples.
  + Apply dimensionality reduction techniques like Principal Component Analysis (PCA) or feature selection methods to reduce the number of features while retaining the most informative ones.
  + Split the data into training (80%) and testing sets (20%) to evaluate model performance.
* **Model Development:**
  + Train several machines learning models, including Random Forest, Support Vector Machines (SVM), and Neural Networks, using the training dataset.
  + Perform hyperparameter tuning (e.g., grid search or random search) to optimize model performance.
  + Use cross-validation to assess model generalization and avoid overfitting.
* **Model Evaluation:**
  + Evaluate model performance on the testing set using metrics such as accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC-ROC).
  + Analyze the importance of different EEG features to understand which brainwave characteristics are most predictive of ADHD.
* **Validation and Testing:**
  + Validate the results using a hold-out set or cross-validation to ensure the model generalizes well to new data.

**Initial/Mini Literature Review**

**Paper 1:** **EEG classification of ADHD and normal children using non-linear features and neural networks**

This study examines how children with attention deficit hyperactivity disorder (ADHD) are classified using electroencephalography (EEG) data analysis. Using non-linear data taken from EEG signals, it uses sophisticated neural network architectures to distinguish between children with ADHD and typical kids. The research highlights the significance of using non-linear metrics, such as entropy and complexity indices, which can capture the intricate dynamics of brain activity often overlooked by traditional linear methods [1].

The methodology involves preprocessing EEG data, extracting relevant non-linear features, and training neural network models to achieve accurate classification. Results demonstrate that the proposed approach significantly improves classification performance compared to conventional techniques, suggesting that neural networks can effectively model the complexities of ADHD-related brain patterns [1].

This project focuses on classifying ADHD in children using EEG data, which is an area that has seen growing interest in recent years. Similar studies have used EEG signals to differentiate between children with ADHD and typically developing children. The key focus of this project on non-linear features and neural networks adds a modern approach to the existing body of literature. Previous studies may have primarily utilized linear features or classical machine learning techniques. By leveraging non-linear features, this research aims to capture complex relationships in the data that linear methods might miss.

Paper 2: **Automatic minimization of eye blink artifacts using fractal dimension of independent components of multichannel EEG**

This study addresses the challenge of eye blink artifacts in multichannel Electroencephalography (EEG) recordings, which can interfere with accurate signal analysis. The authors propose an automated method to minimize these artifacts by utilizing the fractal dimension of independent components derived from EEG data [2].

The approach begins with preprocessing the multichannel EEG signals to isolate independent components using Independent Component Analysis (ICA). The fractal dimension—a measure of complexity in a signal—is then calculated for each independent component. Components identified as containing eye blink artifacts are selectively filtered out based on their fractal dimension values [2].

Results demonstrate that the proposed method effectively reduces eye blink artifacts while preserving the underlying brain signals. The automatic nature of the approach enhances its applicability in real-time EEG processing, making it a valuable tool for researchers and clinicians in neurophysiological studies [2].

The study highlights the utility of fractal dimension as a robust metric for artifact detection and removal, paving the way for improved accuracy in EEG signal interpretation and analysis [2].

While both papers contribute to the field of EEG research, this paper focuses on the classification of ADHD through innovative machine learning techniques, whereas the compared paper emphasizes the technical challenge of removing artifacts to ensure data integrity. Both approaches highlight the importance of advanced analysis methods in improving EEG signal interpretation, though they target different aspects of EEG study.

Paper 3: **Electroencephalogram (EEG) based prediction of attention deficit hyperactivity disorder (ADHD) using machine learning**

This study investigates the use of Electroencephalogram (EEG) signals to predict attention deficit hyperactivity disorder (ADHD) in children through machine learning techniques. The authors aim to develop a reliable predictive model that leverages EEG data to assist in the diagnosis of ADHD, providing an objective measurement to complement traditional assessment methods [3].

The methodology involves the collection of EEG recordings from children diagnosed with ADHD and a control group of neurotypical children. The EEG signals are pre-processed to remove noise and artifacts, followed by feature extraction to identify relevant characteristics of brain activity associated with ADHD. Various machine learning algorithms, including support vector machines, decision trees, and neural networks, are then employed to classify the data [3].

Results indicate that the machine learning models can effectively distinguish between ADHD and control groups based on EEG features, achieving high accuracy rates. The study highlights the potential of using EEG-based measures in clinical settings, paving the way for more objective and data-driven diagnostic approaches [3].

Both papers contribute to the growing body of research on using EEG signals to understand and diagnose ADHD, but they approach the problem from different angles. This paper focuses on the classification of ADHD using machine learning and neural network techniques and non-linear features, while the compared paper emphasizes a broader prediction framework using various machine learning algorithms. Together, they highlight the importance of integrating machine learning with EEG analysis to improve ADHD diagnosis and treatment strategies.

**References**

1. Mohammadi, M.R., Khaleghi, A., Nasrabadi, A.M. *et al.* EEG classification of ADHD and normal children using non-linear features and neural network. *Biomed. Eng. Lett.* **6**, 66–73 (2016). <https://doi.org/10.1007/s13534-016-0218-2>
2. M. Samavati, A. Motie Nasrabadi and M. R. Mohammadi, "Automatic minimization of eye blink artifacts using fractal dimension of independent components of multichannel EEG," *20th Iranian Conference on Electrical Engineering (ICEE2012)*, Tehran, Iran, 2012, pp. 1576-1578, doi: 10.1109/IranianCEE.2012.6292611.
3. Ahire, N., Awale, R. N., & Wagh, A. (2023). Electroencephalogram (EEG) based prediction of attention deficit hyperactivity disorder (ADHD) using machine learning. *Applied Neuropsychology: Adult*, 1–12. <https://doi.org/10.1080/23279095.2023.2247702>
4. S. Saini, R. Rani and N. Kalra, "Prediction of Attention Deficit Hyperactivity Disorder (ADHD) using machine learning Techniques based on classification of EEG signal," *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2022, pp. 782-786, doi: 10.1109/ICACCS54159.2022.9785356.
5. H. Chen, W. Chen, Y. Song, L. Sun and X. Li, “EEG characteristics of children with attention-deficit/hyperactivity disorder,” Neuroscience, vol. 406, no. 318, pp. 444–456, 2019.
6. S. Khoshnoud, M. A. Nazari and M. Shamsi, “Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals,” Journal of Integrative Neuroscience, vol. 17, no. 1, pp. 17– 30, 2018.