A

REPORT

ON

EEG-based ADHD Classification using Machine Learning

By

Name (s) of the student (s)	Enrolment/Registration No.
Garv Malik	200C2030144

Prepared in the partial fulfilment of the

Practice School III Course

AT

BML Munjal University

A Practice School III Station of

BML MUNJAL UNIVERSITY

(July 2024)

Certificate of Authenticity

CERTIFICATE

This is to certify that Practice School Project of ___Garv Malik__ titled EEG-based ADHD Classification using Machine Learning____ is an original work and that this work has not been submitted anywhere in any form. Indebtedness to other works/publications has been duly acknowledged at relevant places. The project work was carried during __4 March 2024 to ___21 July 2024___ in ___BML Munjal University.

Signature of PS-II Faculty Advisor	**Signature of industry Mentor/Supervisor**
Name: Dr. Devanjali Relan	**Name:**
Designation: Associate Professor	**Designation:**
Date 22/07/2024	***(Seal of organization with Date)***

Acknowledgment

I am highly grateful to **Dr. Devanjali Relan, Associate Professor**, BML Munjal University, Gurugram, for providing supervision to carry out the **PS3** study from **February - July 2024**. **Dr. Devanjali Relan** has provided great help in carrying out my work and is acknowledged with reverential thanks. Without wise counsel and able guidance, it would have been impossible to complete the training in this manner.

I would like to express thanks profusely to thank **Dr. Devanjali Relan**, for stimulating me from time to time. I would also like to thank the entire team at BML Munjal University. I would also thank my friends who devoted their valuable time and helped me in all possible ways toward successful completion.

Garv Malik

Abstract

Attention Deficit Hyperactivity Disorder (ADHD) is a common childhood developmental disorder. Early detection is very important for effective intervention. Electroencephalography (EEG) signals offer a great way to objectively classify ADHD based on characteristic brain wave patterns. This study showcases a comprehensive EEG machine-learning pipeline set to automate ADHD identification using modern techniques. We used a diverse dataset comprising EEG recordings from 121 children to ensure robustness and cut down on measurement biases. We detailed the signal preprocessing and data preparation steps, highlighting the need for clean data handling & feature extraction. Our approach uses advanced machine learning methodologies. Firstly, a neural network trained via 10-fold cross-validation on 19-channel EEG data achieved an accuracy of 95%. This shows the neural network's effectiveness in processing complex neurophysiological data for diagnostic purposes. Additionally, gradient boosting algorithms were applied, focusing on 13 selected time-domain features, achieving even higher accuracy of 98%. This paper goes in-depth into the methods used, covering preprocessing strategies, feature selection criteria, (and model optimization techniques). The implications these findings have for clinical applications are analyzed too. They highlight the potential of EEG-based approaches to significantly boost ADHD diagnostic accuracy and inform tailored treatment strategies. By blending advancements in EEG signal analysis with strong machine learning frameworks, this study aids ongoing efforts to leverage neuroimaging technologies for early ADHD detection & intervention.

Table of Contents

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a prevalent neurological disorder. Primarily, it impacts children, but its effects can continue into the adulthood. This disorder is described by inattention, hyperactivity and impulsivity. ADHD changes an individual's performance in educational settings, workplaces, & social interactions and many others. Therefore, early & accurate diagnosis is crucial for effective treatment and care. One method gaining traction is Electroencephalography (EEG). EEG offers a non-invasive way to record brain activity. It has been investigated as a diagnostic tool for attention-deficit/hyperactivity disorder, as it strikingly captures brain activity correlated with the condition.

In the past years, several publications have used machine learning and deep learning techniques in the analysis of EEG data for ADHD classification. The results have been very encouraging. For example, Maniruzzaman et al. suggested a machine learning framework based on using a publicly available EEG dataset comprising 121 patients. They extracted morphological and time-domain features while adopting t-test and LASSO methods for filter feature selection. Evaluation was conducted for different classifiers such as SVM, k-NN, MLP, and LR. Their LASSO-based SVM model was the most successful individually, at an accuracy rate of 94%. Another deep learning model, by Moghaddari et al. (2020), incorporates a high number of convolutional neural network layers before classification is done. They achieved an accuracy of 97.91% by turning EEG frequency bands into RGB images first and then using a model with 13 convolutional layers.

Other studies have been concentrated on improving ADHD classification with different approaches in EEG analysis. Alim et al. (2023) proposed an EEG machine learning pipeline focusing on linear features using SVM with 10-fold cross-validation and achieving accuracy of 93.2%. Another study by Maniruzzaman et al. (2023) explored a hybrid feature selection method. Another study combined Support Vector Machines & Independent T-Tests and attained an accuracy as high as 97%, using GPC too. Similarly, in their research, Esas et al. applied robust local mode decomposition and variational mode decomposition as recently as 2023. Their methods yielded more than 87% accuracy in classification using a deep learning technique. Building on these developments, this paper presents an approach for the improvement of ADHD classification based on EEG data and a CNN model. Nineteen-channel EEG data were acquired from 121 children at a sampling rate of 128 Hz. Afterward, an extensive preprocessing pipeline was applied, which included segmentation, feature extraction, and normalization. This was followed by a detailed study of various machine learning models, into which we could integrate our CNN design.

The CNN model is also very good at capturing complex characteristics in raw data because it has the ability for automatic pattern recognition, both in time and space, within EEG signals. In this paper, a comparative analysis between traditional machine learning models and the CNN approach for classifying ADHD cases is discussed. Results indicate that the CNN model far outperforms traditional techniques by producing an accuracy peak of 98.53% after 3000 training epochs. This wonderful performance further underlines the role of CNNs in picking out intricate patterns within EEG, thus establishing the latter as a very valuable tool for diagnosis in cases of ADHD. We outline, subsequently, how our methodology and experimental setup were undertaken, and what results were returned in these sections while discussing these findings that contribute to ongoing efforts toward developing reliable & non-invasive diagnostic tools for ADHD.

Main

Dataset

The dataset features EEG signal recordings from a sample of 121 children, albeit with some signal noise. The duration of these EEG recordings differs among participants. The study divided participants into two categories: 61 children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD & a control group of 60 children without ADHD. Gender representation was balanced, including both boys and girls aged between 7 and 12 years. Specifically, there are 98 boys (48 with ADHD & 50 controls) and 23 girls (13 with ADHD & 10 controls)

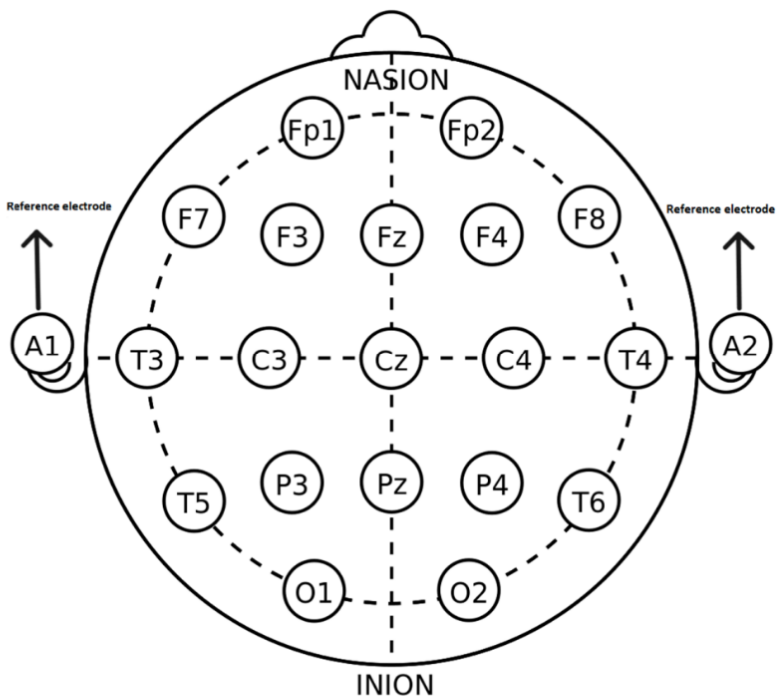

Fig.1. The 10–20 system for electrode position with A1 and A2 reference electrodes

EEG recordings were gathered using 19 channels (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1 & O2), with a frequency distribution of 128 Hz. ADHD diagnoses were made by psychiatrists adhering to DSM-5 criteria. Given that visual attention deficits are primary in children with ADHD, the recording focused on visual attention tasks.

For these tasks, Children viewed images of cartoon characters and counted them; the number ranged between 5 and 16. Upon each child's response the image was immediately substituted by another (ensuring constant stimulation). The duration of the EEG recording for this cognitive visual task was calibrated according to each child's performance.

Preprocessing

In our study, we take special care in preparing the EEG data. This meticulous process ensures the data is ready for in-depth analysis. First, we load EEG files from two main folders: one for ADHD subjects & another for control subjects. Each file is read carefully, and unnecessary columns are removed. The data then gets with specific channel names that match the standard EEG electrode placements, and we set the sampling frequency to 128 Hz. With help from the MNE library, we created an info object, which includes channel names & types. This info object becomes vital as we construct a RawArray object that holds the EEG data.

To conduct a detailed analysis, we divide the continuous EEG recordings into six-second epochs. There's a one-second overlap between each segment. We use functions in MNE to handle this segmentation, resulting in a dataset where every epoch captures a part of the original recording. These segmented data arrays are saved into lists, divided between ADHD & control groups.

Next, we assign labels to each segment. Control segments get a label of 0, while ADHD segments are labeled as 1This contributes to the generation of lists with segment labels corresponding to each patient's data. After segmentation of a given dataset, we merge the segmented data with their corresponding labels from all patients and generate aggregate data. Furthermore, we also generate a grouping index to keep track of the source of each segment. It is quite an important step in keeping intact the integrity of the individual subject data in further analyses.

The last step will reshape the data arrays. We first stack all of our data arrays vertically to obtain a single dataset. Likewise, we combine the labels and group indices. At last, we are going to transpose the data array into the reordering of axes so that it becomes feasible for our required axes for further machine learning or statistical analysis. We have completed this thorough preprocessing pipeline, guaranteeing the consistency of our EEG data and that it is well prepared enough for in-depth analyses in our ADHD research study.

Feature Extraction

Statistic	Explanation
Mean	The average of all the values in the dataset x.
Standard Deviation	The standard deviation, measuring the amount of variation or dispersion of the values in x.
Variance	The variance, representing the average of the squared differences from the mean of x.
Peak-to-peak	The range (peak-to-peak), which is the difference between the maximum and minimum values in x.
Minimum	The minimum value in the dataset x.
Maximum	The maximum value in the dataset x
Arg Minimum	The index of the first occurrence of the minimum value in x.
Arg Maximum	The index of the first occurrence of the maximum value in x.
Root mean square	The root mean square, which is the square root of the mean of the squared values in x.
Absolute Difference	The sum of absolute differences between consecutive elements in x, indicating the signal's variability.
Skewness	The measure of the asymmetry of the probability distribution of x.
Kurtosis	The measure of the "tailedness" of the probability distribution of x, indicating the presence of outliers.

Table 1. Summary of EEG features employed in this study.

Computing these statistics over the last axis of the data array corresponds to its time dimension. As will be explained, the EEG segments thus translate into a number of features that provide very interesting and informative views on their characteristics. We stack these different vectors of features into one single array per segment while keeping the order of features for coherence. After extraction, all features are then combined into one large dataset where now every EEG segment is represented by a 1D array of features; this mark signifies its transition from the original format to a structured feature matrix.

Classifier Selection

In our study, we adopted a comprehensive approach for evaluating different machine learning techniques against the EEG features we extracted. the purpose behind this is to distinguish between the ADHD subject and those belonging to the control group. First, we did a split for training and testing data in an 80:20 ratio. After that, we created models and estimated their effectiveness using k-fold cross-validation with 10 folds. It is a technique that cuts this real estate training dataset into 10 pieces. Each of the segments is used for validation while the rest are used for training. Since we repeated this process 10 times, it simply means each segment was tested at least once. As such, this approach will give a more accurate measure of the model's performance and at the same time alleviate overfitting.

We tried out several classifiers: Extra Trees, Random Forest, XGBoost, LightGBM, Gradient Boosting, AdaBoost, & Logistic Regression. We also used different preprocessing techniques aimed at standardizing & normalizing the data. This step is important for making sure machine learning algorithms work great. For preprocessing, we had a few methods: MinMaxScaler (which scales features to a fixed range like [0, 1]), StandardScaler (to standardize features by subtracting the mean), MaxAbsScaler (which scales by maximum absolute value but keeps sparsity), PowerTransformer (applies a power transformation to stabilize variance), QuantileTransformer (transforms features to resemble uniform or normal distributions), and Normalizer (scales samples to have a unit norm). Every model went through these transformations to find out which preprocessing method worked best.

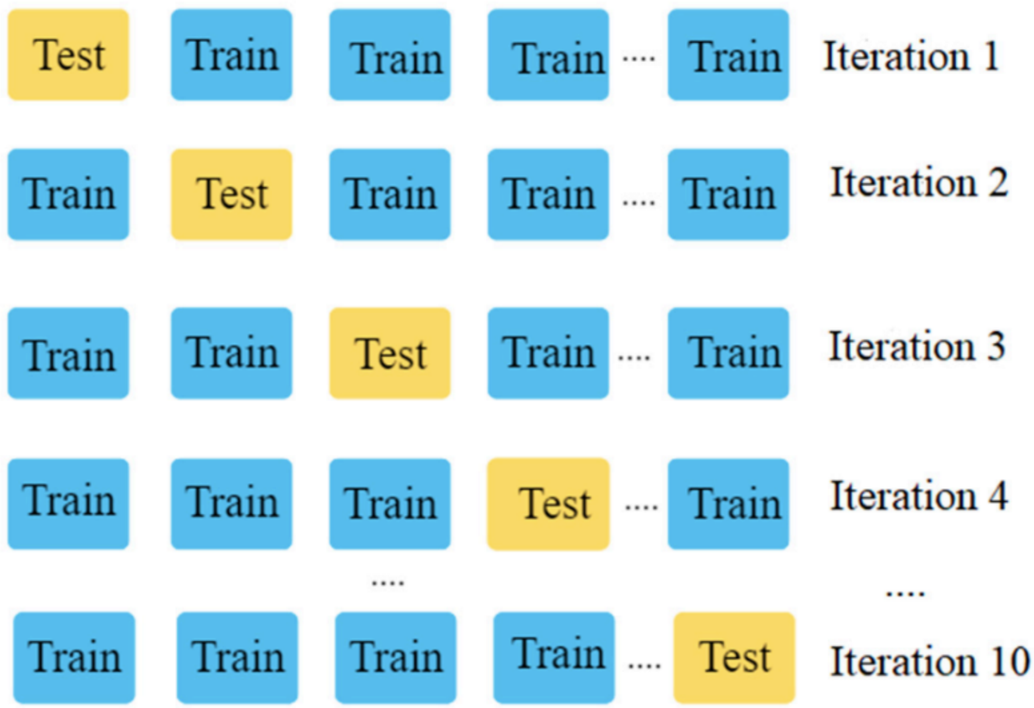

Fig. 2. K-fold (10-fold) cross-validation process

In our tests of different classifiers, the Extra Trees Classifier came in at an accuracy of 75%. It did okay on complex datasets but not best overall. Random Forest Classifier did better, hitting an accuracy of 81%. It has an ensemble of decision trees, which combine their predictions. It didn't hit top accuracy levels, though. At last but not least, there is the XGBoost Classifier—the latest in gradient boosting capabilities. It recorded a superb accuracy of 97.8%. This means that it was able to identify complex patterns within the EEG data. The LightGBM Classifier went a notch higher to post an accuracy of 98.5%. This sterling performance is a pointer that it is characterized by smooth operations due to its fast and efficient handling of large datasets.

In comparison, Gradient Boosting Classifier returned an accuracy of 95%. While this score was pretty respectable, it somehow still remained a bit lower than those of XGBoost & LightGBM. The AdaBoost Classifier managed to obtain an accuracy of 90% through its integration of weak learners but did not keep pace with gradient boosting methodologies. Then there's the Logistic Regression model, a simpler cousin that hit an accuracy of 91%. Still, this figure does not hold a candle to the more sophisticated gradient boosting models.

CNN Model

We decided to use a Convolutional Neural Network as our secondary model. The reason for implementing this model is that using a CNN to analyze EEG data gives very huge advantages against traditional machine learning models like Extra Trees, Random Forest, and methods of gradient boosting. There is one major strength of CNN: they can automatically extract hierarchical features from raw data. This allows for complex patterns in the time and space domains of the EEG signals to be captured. Accordingly, this automated feature extraction ability of this skill reduces the need for manual preprocessing and domain-specific feature engineering.

Furthermore, CNNs demonstrate remarkable effectiveness when managing large, high-dimensional datasets. The accuracy provided by our results firmly supports this assertion. These capabilities render CNNs especially suitable for tasks like classifying ADHD based on EEG data, where identifying subtle and complex patterns is essential for accurate diagnosis.

In our approach to classify EEG data for ADHD detection, we utilized a Convolutional Neural Network (CNN), coupled with k-fold cross-validation—similar to what we applied in our previous methodology. Using this consistent k-fold cross-validation approach across both methodologies serves to ensure both reliability & rigor in our analysis.

CNN Architecture

The CNN model is built with great care to capture the complex time-related patterns in EEG data. It has many layers, specifically convolutional ones, followed by normalization, activation, pooling, & dropout layers. You can see this in fig 1.1. First, the model uses a 1-dimensional convolutional layer that has 5 filters. The kernel size is 3, and it has a stride of 1. This layer helps in pulling out local temporal features from the input EEG signals.

Then comes a BatchNormalization layer. This layer standardizes the outputs which speeds up training & ensures stability. Next, a non-linear LeakyReLU activation function is applied. This step adds non-linearity—very important for learning complicated patterns. After that, there’s a MaxPooling1D layer with a pool size of 2 & a stride of 2. It reduces the size of the feature maps and helps control overfitting by dropping non-maximal values.

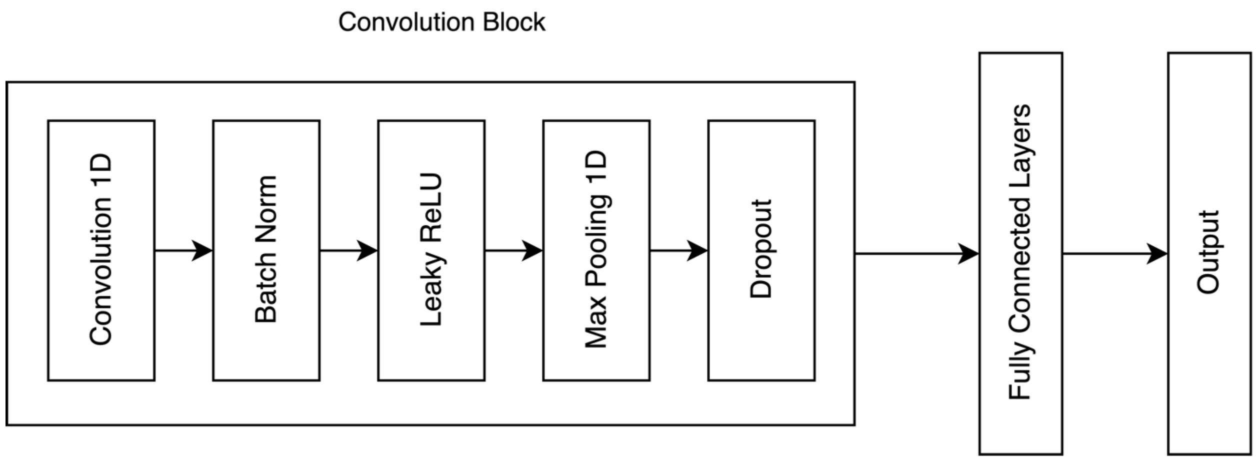

Fig.3. CNN Architecture, Convolution block has been repeated 5 times.

To tackle overfitting further, a Dropout layer is included with a dropout rate of 0.4. This means that during each training update, some input units are randomly set to zero. That's important! The pattern of convolution followed by normalization, activation, pooling, and dropout happens four more times which steadily improves the feature extraction process.

Once all convolutions are done, we have a Flatten layer that turns the output from 3D into a 1D vector to prepare for the fully connected layers. There are two dense layers next—one with 48 units & another with 32 units—both using ReLU for activation. To aid regularization, Dropout layers are added after each dense layer.

Finally, there's one last dense layer which has just one unit and uses a sigmoid activation function. This outputs the probability of whether the input signals are classified as ADHD or not.

We also use the Adam optimizer with a learning rate of 0.0003 for model compilation. Binary cross-entropy is our loss function, while accuracy is our evaluation metric. we prefer this because it is efficient and very excellent in binary classification tasks. As we increase the number of epochs, the accuracy of the CNN model increases drastically in our experiments. The initial trials of 50 epochs resulted in an accuracy of 78.76%. This surged to as high as 98.53% when the number of epochs reached 3000. Clearly seen from this trend is the ability of the model to learn and generalize better from EEG data if it is trained for a longer time. The highest averaged accuracy achieved was 98.53%, underlining the outstanding capability of this CNN in identifying sophisticated temporal and spatial patterns in any EEG signal. Thus, it becomes one of the excellent choices to perform ADHD classification tasks.

Results

In this full-detailed analysis, classification of ADHD using EEG data was done with two different methodologies in order to find out which one is the best. One path includes traditional machine learning models combined with a variety of preprocessing techniques; another one is the application of Convolutional Neural Network (CNN) model with K-fold cross-validation methodology for better results. The performance evaluation of both strategies was done on rigorous terms and the factors impacting them were understood.

Deep exploration of many machine learning algorithms improved with preprocessing methods oriented towards standardizing and normalizing data was the first method. We have examined models containing Extra Trees, Random Forest, XGBoost, LightGBM, Gradient Boosting, AdaBoost, and Logistic Regression. Each model was measured with six different scalers: MinMaxScaler, StandardScaler, MaxAbsScaler, PowerTransformer, QuantileTransformer, and Normalizer. Such extensive comparison showed that gradient boosting models — specifically LightGBM and XGBoost — significantly outperform others.

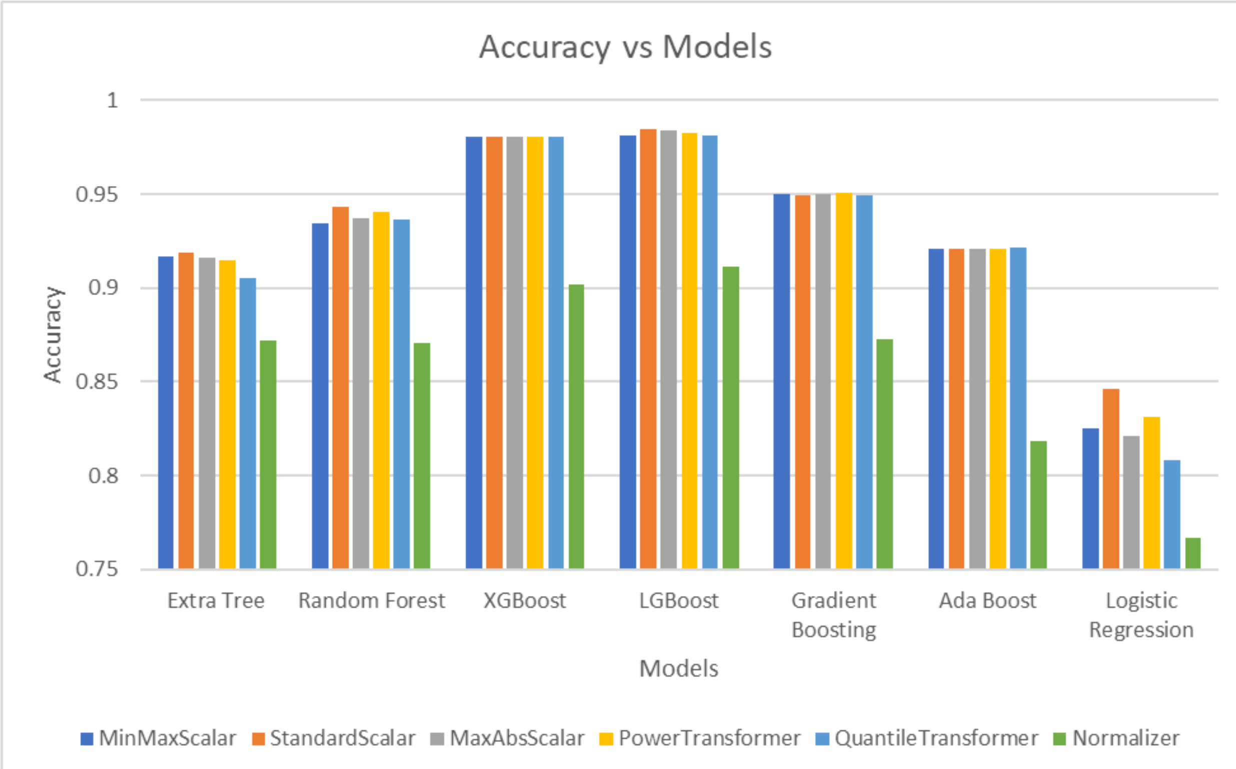

Fig.4. Models vs Accuracy

Much to my surprise, LightGBM with StandardScaler gave an accuracy of 98.40%, thus depicting LightGBM's excellent capability in handling complex data patterns and large datasets very effectively. Other than that, XGBoost also performed quite well, reaching about 98% with most of the scalers. Simpler models like logistic regression and AdaBoost returned respectable results but did not perform as compared to the advanced gradient boosting techniques. This may be related to the ability of gradient boosting models in capturing subtle relationships in EEG data, due to their strong ensemble learning coupled with efficient gradient descent optimization.

In the second approach, the architecture used was a CNN, where the powerful feature extraction properties in CNN were pitted against the effective analysis of EEG data. The CNN model had several convolutional layers followed by batch normalization, Leaky ReLU functions, max-pooling layers, and dropout layers, which reduced risks associated with fitting. This design for the architecture wanted to effectively understand spatial and temporal characteristics within EEG signals. We have performed k-fold cross-validation to ensure the robustness of our results while checking model performance over different epochs: 50, 100, 200, 500, 1000, 1500, 2000 & 3000. The CNN results were excellent since it continuously improved with greater numbers of epochs, reaching as high as 98.53% at 3000 epochs, thereby beating all gradient boosting models considered in our previous analysis. This performance highlights the remarkable ability of CNNs in learning complex features directly from raw EEG data, wherein hierarchically extracted features cover low- as well as high-level patterns.

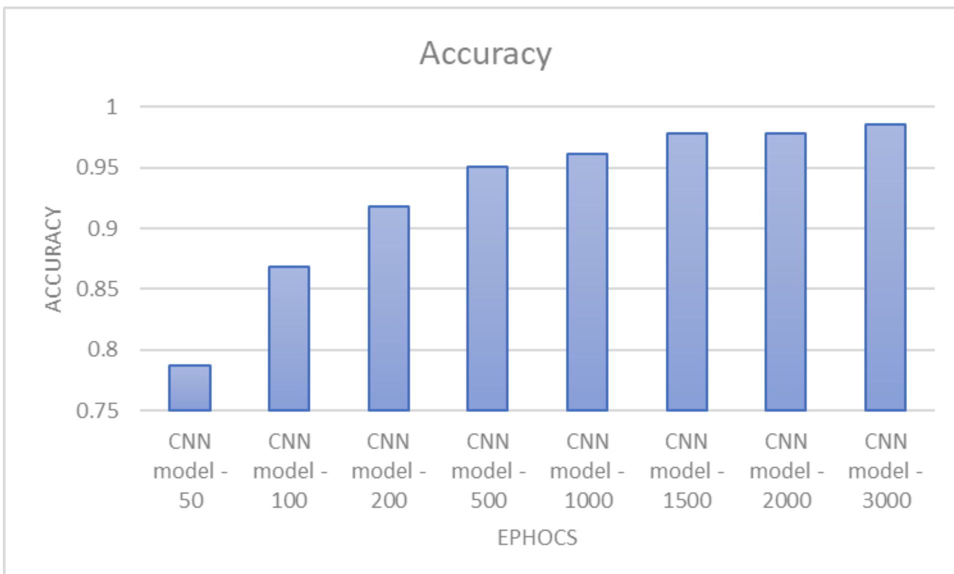

Fig.5. Model accuracy vs Epoch.

Several factors set the CNN apart in performance. First, since convolutions in a CNN efficiently bring down the dimension while retaining the essential features, it is good at handling high-dimensional data such as EEG signals. Secondly—the reason lying in their hierarchical structure—CNNs express intricate patterns at multiple levels of abstraction, and this makes them very appropriate to be used for analyzing complex datasets like EEG. Furthermore, the introduction of dropout layers in our CNN design reduced the chances of overfitting and made our model generalize well on new unseen data. Traditional machine learning approaches rely on feature extraction and other preprocessing steps that are manual and might miss some key patterns in a dataset, hence leading to marginally inferior performance outcomes overall.

Conclusion

This study offers intriguing insights. There are some very exciting insights in this study. A detailed analysis about ADHD classification based on EEG data is presented here. We have also made comparisons between traditional machine learning models and a Convolutional Neural Network (CNN). Our results are very promising, showing that higher-order machine learning techniques perform better, especially gradient boosting models. Notably, XGBoost returned an accuracy of 98.40%. These models enjoyed the benefits of strong ensemble learning coupled with efficient gradient descent optimization strategies, such that they had no problem returning complex patterns from within large EEG datasets.

Now, let us discuss the CNN model. This model is also constructed with many layers; added to the convolutional layers are normalization, activation, pooling, and dropout layers. Shockingly, this model turned out to have higher performances with an accuracy rate of 98.53% after 3000 epochs. This rise in performance could be significant since CNN would identify essential features from the raw EEG data itself. This is ability to capture fine-grain patterns supersedes what has been traditionally handled, often at a high degree of manual adjustment.

Additionally, the dropout layers in the CNN architecture were an expedient factor that helped significantly in the avoidance of overfitting. This feature improves its flexibility while fitting to new data. These results will further reinforce the important role of deep learning models in the diagnosis of neurodevelopmental disorders. They represent considerable promise for developing automated & non-invasive tools for ADHD diagnosis.

This review has been called to the attention of all that the diagnostic accuracy of CNNs is quite high. This research study allows for future prospects by incorporating more types of data and deep learning techniques into this research to make the process of diagnosis more accurate. The potential of the CNN model in identifying and classifying ADHD from EEG data, therefore leading to improved medical outcomes and interventions for those in the most need, was also brought out by its successful implementation in this model. Indeed, this is an exciting time in terms of developments in ADHD diagnosis.

References

[1] Md. Maniruzzaman, Jungpil Shin, Md. Al Mehedi Hasan, and Akira Yasumura. 2022. Efficient Feature Selection and Machine Learning Based ADHD Detection Using EEG Signal. *Computers, Materials & Continua/Computers, Materials & Continua (Print)* 72, 3: 5179–5195. https://doi.org/10.32604/cmc.2022.028339

[2] Tor, Hui Tian, Chui Ping Ooi, Nikki Sj Lim-Ashworth, Joel Koh En Wei, V Jahmunah, Shu Lih Oh, U Rajendra Acharya, and Daniel Shuen Sheng Fung. “Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals.” *Computer Methods and Programs in Biomedicine* 200 (March 1, 2021): 105941. https://doi.org/10.1016/j.cmpb.2021.105941.

[3] TaghiBeyglou, Behrad, Navid Hasanzadeh, Fatemeh Bagheri, and Mehran Jahed. “ADHD diagnosis in children using Common spatial pattern and Nonlinear Analysis of Filter banked EEG,” August 4, 2020.

[4] Alim, Anika, and Masudul H. Imtiaz. “Automatic Identification of Children with ADHD from EEG Brain Waves.” *Signals* 4, no. 1 (February 21, 2023): 193–205. https://doi.org/10.3390/signals4010010.

[5] Maniruzzaman, Md., Md. Al Mehedi Hasan, Nobuyoshi Asai, and Jungpil Shin. “Optimal Channels and Features Selection Based ADHD Detection From EEG Signal Using Statistical and Machine Learning Techniques.” *IEEE Access* 11 (January 1, 2023): 33570–83. https://doi.org/10.1109/access.2023.3264266.

[6] Moghaddari, Majid, Mina Zolfy Lighvan, and Sebelan Danishvar. “Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG.” *Computer*

Methods and Programs in Biomedicine 197 (December 1, 2020): 105738. https://doi.org/10.1016/j.cmpb.2020.105738.

[7] Esas, Mustafa Yasin, and Fatma Latifoğlu. "Detection of ADHD from EEG signals using new hybrid decomposition and deep learning techniques." *Journal of Neural Engineering* 20, no. 3 (June 1, 2023): 036028. https://doi.org/10.1088/1741-2552/acc902.

[8] Einizade, Aref, Mohsen Mozafari, Mohammadreza Rezaei-Dastjerdehei, Elnaz Aghdaei, Amir Mohammad Mijani, and Sepideh Hajipour Sardouie. "Detecting ADHD children based on EEG signals using Graph Signal Processing techniques," November 26, 2020. https://doi.org/10.1109/icbme51989.2020.9319456.

[10] Joy, R. Catherine, S. Thomas George, A. Albert Rajan, and M.S.P. Subathra. "Detection of ADHD From EEG Signals Using Different Entropy Measures and ANN." *Clinical EEG and Neuroscience* 53, no. 1 (August 23, 2021): 12–23. https://doi.org/10.1177/15500594211036788.

[11] Alchalabi, Alaa Eddin, Shervin Shirmohammadi, Amer Nour Eddin, and Mohamed Elsharnouby. "FOCUS: Detecting ADHD Patients by an EEG-Based Serious Game." *IEEE Transactions on Instrumentation and Measurement* 67, no. 7 (July 1, 2018): 1512–20.

[12] Ekhlasi, A., Nasrabadi, A. M., & Mohammadi, M. (2021). Classification of the children with ADHD and healthy children based on the directed phase transfer entropy of EEG signals. *Frontiers in Biomedical Technologies*.

[13] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794. 2016.

[14] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30

[15] Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. Journal-Japanese Society For Artificial Intelligence, 14(771-780), 1612.

[16] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32.

[17] LaValley, Michael P. "Logistic regression." *Circulation* 117, no. 18 (2008): 2395-2399.