ISSN: 1672 - 6553

JOURNAL OF DYNAMICS AND CONTROL

VOLUME 10 ISSUE 01: P65-78

AN ARTIFICIAL INTELLIGENCE TECHNIQUE FOR ATTENTION DEFICIT HYPERACTIVITY DISORDER (ADHD) CLASSIFICATION BASED ON ELECTROPHYSIOLOGY (EEG) ON TWO BENCHMARK DATASETS

Lamiaa A. Amar[1], Marwa Y. Mohamed[5], Sara Gaballa[2],Mariam Hossam[2], Ahmed.M.Otifi[3], Noha Khattab[4]

[1]Department of Networks and Distributed Systems, Informatics Research Institute, City of Scientific Research and Technological Applications, Srta-City, Alexandria, 21934, Egypt.
[2]Faculty of Engineering, Alexandria University, Alexandria, 21544, Egypt
[3]Department of data science, Faculty of Computers and Data Science, Alexandria University, Alexandria, 21544, Egypt
[4]School of Science and Technology, University of Canberra, Australia.
[5]Department of Multimedia and Computer Graphics, Informatics Research Institute, City of Scientific Research and Technological Applications, Srta-City, Alexandria, 21934, Egypt

https://doi.org/10.71058/jodac.v10i01004

AN ARTIFICIAL INTELLIGENCE TECHNIQUE FOR ATTENTION DEFICIT HYPERACTIVITY DISORDER (ADHD) CLASSIFICATION BASED ON ELECTROPHYSIOLOGY (EEG) ON TWO BENCHMARK DATASETS

Lamiaa A. Amar[1*], Marwa Y. Mohamed[5], Sara Gaballa[2],Mariam Hossam[2], Ahmed.M.Otifi[3], Noha Khattab[4]

[1]Department of Networks and Distributed Systems, Informatics Research Institute, City of Scientific Research and Technological Applications, Srta-City, Alexandria, 21934, Egypt.
[2]Faculty of Engineering, Alexandria University, Alexandria, 21544, Egypt
[3]Department of data science, Faculty of Computers and Data Science, Alexandria University, Alexandria, 21544, Egypt
[4]School of Science and Technology, University of Canberra, Australia.
[5]Department of Multimedia and Computer Graphics, Informatics Research Institute, City of Scientific Research and Technological Applications, Srta-City, Alexandria, 21934, Egypt.
*Corresponding author(s): lamy2003@hotmail.com, lamar@srtacity.sci.eg;
Contributing authors: mdeeb@srtacity.sci.eg; saragaballa2002@gmail.com; m.hossam2551@gmail.com; ahmedotifi11112000@gmail.com; noha.khattab@canberra.edu.au

Abstract: The diagnosis of attention deficit hyperactivity disorder (ADHD) is characterized by persistent patterns of hyperactivity, inattention, and impulsivity, which significantly impact academic performance, social interaction, and occupational functioning. Treatment is often inconsistent and delayed due to traditional diagnostic methods that rely on subjective clinical assessments. Electroencephalography (EEG) is a non-invasive neurophysiology technique that measures the brain's electrical activity, providing a deeper understanding of the neurobiology of ADHD. Machine learning to analyze EEG data, identifying subtle, complex electrophysiological patterns in high-dimensional EEG data that are imperceptible to human observation. This combination leads to more accurate and automated ADHD classification, potentially enabling an earlier diagnosis and personalized intervention. This research aims to develop a more generic screening tool objective for ADHD using EEG signals. Two benchmark datasets were utilized with many features and extracted 24 features across multiple domains, including time, frequency, morphological, and non-linear characteristics. The LASSO-LR model was applied to refine feature selection, identifying the most relevant features for machine learning and deep learning algorithms. Furthermore, the fusion of unhealthy groups within dataset2 results in a strong and high classification accuracy. Among several models tested, the K-Nearest Neighbours (KNN) algorithm achieved the highest accuracy in distinguishing ADHD patients from healthy individuals across both datasets. This study contributes to the growing body of research by offering a novel approach, demonstrating the feasibility of machine learning-based methods as potential diagnostic tools for ADHD diagnosis. Due to their efficiency and adaptability, these models may prove especially advantageous for small-scale edge computing applications in the foreseeable future.

Keywords: ADHD, classification, machine learning, deep learning, feature extraction.

1 - Introduction

Attention-deficit/hyperactivity disorder (ADHD) ranks among the most prevalent neurodevelopmental disorders in children, impacting approximately 11 percent of school-aged children in the United States [1]. This condition is characterized by a persistent pattern of inattention, impulsivity, and/or hyperactivity that disrupts daily functioning and development [2]. ADHD was first diagnosed in children as young as three years old and can continue into adulthood. If not recognized and treated appropriately, ADHD can lead to significant negative outcomes, such as academic difficulties, family tension and upheaval, depression, relationship issues, and job-related challenges. Therefore, early detection and intervention are crucial [1].

https://doi.org/10.71058/jodac.v10i01004

Jasper et al . [3] were pioneers in the application of electroencephalograms (EEG) for the study of ADHD over 75 years ago . There technique has proven to be an effective instrument for diagnosing various brain disorders. Subsequently, in 1991, J. Lubar conducted the first study examining EEG signal abnormalities in individuals with ADHD [4]. He found an increase in theta band activity (4–8 Hz) accompanied by a significant decrease in beta band power (13–30 Hz). A characteristic feature of the brain wave patterns in ADHD patients is the prevalence of slow brain waves (delta or theta) and a deficiency of fast brain waves (beta). Consequently, these individuals exhibit a high theta-to-beta ratio, which may serve as a basis for the automatic detection of ADHD through brain wave analysis [4]. In recent years, researchers have also employed magnetic resonance imaging (MRI) for ADHD diagnosis. However, EEG remains more commonly utilized due to its affordability, non-invasive nature, portability, and ability to provide clear insights into brain function. The data generated by EEG encompasses a vast and intricate array of information, making it challenging for humans to manually identify abnormalities. This is where machine learning becomes advantageous [5].

Machine learning is an evolving technology that enables computers to automatically learn from data and apply this knowledge to current tasks. Numerous studies have been conducted to propose an automated system for the early detection of patients with ADHD, which are summarized in Table 1. In recent research, Alim and Imtiaz [6] extracted linear features from EEG data to develop a model based on Gaussian Support Vector Machines (SVM). They minimized computational demands by filtering out signals above 30 Hz and utilized only the first four sub-bands of EEG, achieving an average accuracy of 94%. Meanwhile, Maniruzzaman et al. [7] concentrated on the optimal selection of channels and features, employing a combination of two distinct methods, support vector machine and t-test to identify the most effective channels.

Table 1. Summary of the related works.

Reference	Data Size	Feature Extraction	Feature Selection	Model	ACC
[5]	50	Non-linear	Exhaustive	ML(KNN)	96%
[8]	60	Non-linear	DISR mRMR	ML(MLP)	93.65% 92.28%
[9]	144	---	---	DL(EEGNET)	83%
[10]	103	Statistical	PCA	ML(DT, RF, LB, SVM, NB)	98.43%
[6]	120	Linear	ANOVA	ML(SVM)	94%
[7]	121	Time Domain Morphological Non-linear	LASSO	ML(GPC, RF, KNN, MLP, DT, LR)	97.53%
[11]	---	---	---	DL	87%

Subsequently, a LASSO logistic regression model was employed to identify significant features. Six ML classifiers were utilized for ADHD detection, achieving an accuracy of 97.53%. In the study conducted by Esas and Latifoglu [11], EEG signals were analysed using robust local mode decomposition and variational mode decomposition techniques to create sub-bands. They developed a deep learning algorithm that achieved a classification accuracy exceeding 87%.

 https://doi.org/10.71058/jodac.v10i01004

In current research, two benchmark datasets with varying channel counts (19 and 56) are involved. By training different machine learning and deep learning models on the classification of children with ADHD and healthy controls for each dataset, we were able to assess our methodology for both datasets. Then, the results were compared from each dataset, focusing on how factors such as dataset size, data validity, and channel count could influence the identification of children with ADHD. Finally, we compared our findings from each dataset with previous published similar studies. Additionally, we investigated a novel method by merging the ADD and ADHD groups. Our findings revealed a significant enhancement in both classification accuracy and predictive performance when contrasted with the ternary classification model. This implies that integrating ADD and ADHD data could provide a more reliable and efficient strategy for analysing dataset 2. The contributions of this paper could be summarized as follow:

- Significant combination features as time domain, frequency domain, morphological and non-linear features were extracted from EEG signals.
- LASSO-LR method was employed for feature selection and reduction.
- Seven traditional ML classifiers as LR, RF, KNN, GB, DT, SVM, and MLP, in addition to four DL models as LSTM, BI-LSTM, GRU and CNN were selected and trained on both datasets to identify ADHD patients from healthy controls.
- A 5-foldcross validation were performed to avoid bias and overfitting. Furthermore, five performance metrics were used to evaluate our approach.

The remainder of the paper is organized as follows: Section 2 provides an explanation of the methodology, which includes data description, data pre-processing, feature extraction, feature selection, classification models, and evaluation metrics. Section 3 presents the results along with a discussion. Finally, Section 4 offers the conclusion.

2 - Methodology

2.1 Data Description

Dataset 1: a published dataset to investigate effective connectivity patterns in children with Attention-Deficit/Hyperactivity Disorder (ADHD) compared to Typically Developing (TD) children [12]. This dataset included a total of 121 children, comprising 61 diagnosed with ADHD and 60 healthy TD subjects. Both boys and girls aged between 7 to 12 years were part of the sample. The ADHD children were diagnosed according to the DSM-IV criteria by an experienced psychiatrist, ensuring that the diagnosis was reliable and standardized. The EEG recording was conducted using the 10-20 system with 19 channels (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2) at a sampling frequency of 128 Hz. The reference electrodes (A1 and A2) were placed on the earlobes as can be seen in Figure 1. Given that children with ADHD often have visual attention deficits, the EEG recording protocol involved a visual attention task. In this task, children were shown images of cartoon characters and asked to count the number of characters present. The number of characters in each image varied randomly between 5 and 16. The images were designed to be large enough for easy visibility and counting by the children. To maintain a continuous stimulus during EEG recording, each image was displayed immediately and without interruption following the child's response. The duration of the EEG recording was thus dependent on how quickly the child responded during the task. None of the children in the TD

https://doi.org/10.71058/jodac.v10i01004

group had a history of psychiatric disorders, epilepsy, drug abuse, head injury, or any report of high-risk behaviours. This exclusion criterion was crucial to ensure that the TD group was healthy and did not have confounding factors that could affect the results. Details of the dataset can be found in Table 2.

Table 2. Dataset1 participants' details.

Type	Boys	Girls	Age	Mean age
Healthy	50	10	7-12	9.85 ± 1.77
ADHD	48	13	7-12	9.62 ± 1.75

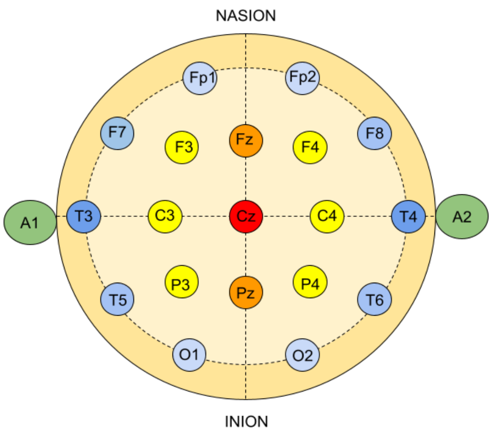

Figure 1.The placement of 10-20 system-based electrodes[12].

.Dataset 2: This dataset is publicly available via OSF [13], was collected from a total of 144 children; comprising 44 healthy controls and 100 diagnosed with neurological disorders. EEG recordings were obtained using 60 Ag/AgCl electrodes and a Brain-Amp amplifier (Brain Products Inc., Munich, Germany). Due to their tendency to produce unreliable readings from high electrode impedances, electrodes P9, P10, P11, and P12 were excluded from the analysis, resulting in data from 56 electrodes. Following segmentation, the dataset was categorized as follows: 10,129 trials for the control group, 13,031 trials for the ADD group, and 10,742 trials for the ADHD group. Table 3 shows information about the participants in datasets2.

Table 3. Dataset2 participants' details

CONDITION/GENDER	FEMALE	MALE	AGE	IQ
ADD	10	42	10.9 ± 2.4	100 ± 12
AD(H)D	12	36	10.6 ± 1.9	103 ± 13
HEALTH CONTROL	15	29	11.3 ± 2d.2	103 ± 12

https://doi.org/10.71058/jodac.v10i01004

2.2 Data Pre-processing

EEG signals from dataset 1 exhibit various artefacts and noise that should be eliminated prior to analysis. For the pre-processing phase, a band pass filter was employed with cut-off frequencies set at 0.5 Hz and 63 Hz. Additionally, a 50 Hz notch filter was utilized to eliminate power line interference. The data was subsequently segmented into 2-second intervals for each channel, with a 1-second overlap. This resulted in a data structure of (16749, 19, 256), where the dimensions represent the number of samples, channels, and timestamps, respectively. This pre-processing procedure is essential for ensuring that the analysis produces reliable and accurate outcomes, as it reduces the influence of unwanted signals on the EEG data.

2.3 Feature Extraction

Feature extraction plays an important role in the classification process, as it involves transforming and reducing raw data into a meaningful and informative set of features. Selecting the right features significantly enhance the performance of the applied model. In this study, 24 were extracted features from EEG signals, categorized into time-domain, frequency-domain, morphological, and non-linear features. For each window, these 24 features were extracted for every channel, except for Power Spectral Density (PSD), which was also computed for each frequency band. This was particularly relevant since ADHD patients exhibit variations in the power of certain frequency bands.

2.4 Feature Selection

Feature selection is the process of reducing the number of input variables by choosing the most significant features and removing highly correlated ones. We employed LASSO with Logistic Regression (LASSO-LR) [14] model with a 5-fold CV. LASSO-LR offers several advantages over other feature selection methods, particularly when dealing with high-dimensional data such as EEG signals. One of its main strengths is its ability to perform automatic feature selection by applying L1 regularization, which shrinks some feature coefficients to exactly zero, effectively removing irrelevant or less important features. This not only reduces dimensionality but also helps in preventing overfitting, resulting in a model that generalizes better to unseen data. LASSO also leads to simpler, more interpretable models; especially valuable in fields like healthcare or neuroscience where understanding the impact of features is essential. As an embedded method, LASSO integrates feature selection directly into the model training process, making it more efficient than separate filter or wrapper methods. Moreover, it handles multicollinearity well by selecting only one feature among highly correlated ones, helping to eliminate redundancy. With the ability to tune its regularization strength via cross-validation, LASSO is both flexible and scalable.

We applied Grid- Search to select the optimal alpha value corresponding to minimum mean squared error. Alpha (α) is the penalization factor that represents the amount of shrinkage that will be applied in the equation [15]. The cost function of the LASSO model is shown in the following equation:

$$\frac{1}{2N}\sum_{N}^{i=1}(y_{real}^{(i)} - y_{pred}^{(i)}) + \propto \sum_{j=1}^{n}|a_j| \qquad (1)$$

https://doi.org/10.71058/jodac.v10i01004

While features with an importance value higher than 0.02 were selected. Table 4 represents more details about the selection process.

Table 4. Information about the feature selection procedure.

Data	Number of features		Alpha
	Before Selection	After Selection	
Dataset 1	532	96	0.001
Dataset 2	1568	82	0.001

2.5 Classification Models

Various machine learning algorithms were used for classification tasks. The algorithms utilized included Logistic Regression [16], Random Forest [17], K-Nearest Neighbours (KNeighbors) [18], Gradient Boosting [19], Decision Tree [20], and MLP classifier (Multi-Layer Perceptron) [21]. Additionally, several deep learning models were incorporated, such as Long Short-Term Memory (LSTM) [22], Bidirectional LSTM (BI-LSTM) [23], Gated Recurrent Unit (GRU) [24], and One-Dimensional Convolutional Neural Network (1D-CNN) [25].

2.6 Evaluation Metrics

Five performance metrics were utilized to evaluate the models: accuracy (ACC), precision (Prec), recall (Rec), F1 score, and the area under the ROC (Receiver Operating Characteristic) curve (AUC). These metrics are derived from the values of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), false positive rate (FPR), and true positive rate (TPR), and are calculated using the following equations:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Prec = \frac{TP}{TP+FP} \quad (3)$$

$$Rec = \frac{TP}{TP+FN} \quad (4)$$

$$F1_{score} = 2 * \frac{Prec * Rec}{Prec + Rec} \quad (5)$$

$$AUC = \sum_{i=1}^{n-1} \frac{(FPR_{i+1} - FPR_i)((TPR_{i+1} - TPR_i)}{2} \quad (6)$$

Where n is the number of thresholds, and $FPR_i$and TPR_i are the FPR and TPR values for the i threshold.

3 - Results

In this section, the results and discussion will be presented. First, the input data and features are normalized using the give formula.

$$z = \frac{X - \mu}{\sigma} \quad (7)$$

https://doi.org/10.71058/jodac.v10i01004

Where X is the original feature vectors or the input data. μ and σ are the mean and standard deviation of the respective vectors. z is the standardized value, and its values lies between 0 to 1. Then, a 5-fold cross-validated grid-search was applied to fine-tune the hyper-parameters of the utilized ML algorithms. This method helps improve the performance and reliability of the system by selecting the best parameter combinations through repeated evaluation.

Dataset1:The results for dataset 1 are detailed in Table 5, which illustrates the effectiveness of classical machine learning classifiers on dataset 1 after partitioning the data in an 80:20 ratio(for training and testing) and employing 5-fold cross-validation. It was observed that the KNN classifier outperformed the other classifiers in terms of performance metrics. Specifically, it achieved a test accuracy of 97.88%, a precision of 97.77%, a recall of 98.45%, an F1 score of 98.11%, and an AUC of 0.98. Furthermore, the MLP classifier also demonstrated strong performance, along with the SVM utilizing the RBF kernel, with accuracies of 96.93% and 96.09%, respectively.

Table 5. Classical ML classifiers' results using dataset1.

Classifier	Accuracy (%)	Precision(%)	Recall (%)	F1 score (%)	AUC	Confusion matrix			
						TP	TN	FP	FN
LR	80.21	80.65	84.84	82.69	0.80	1584	1103	380	283
RF	93.52	91.46	97.48	94.37	0.93	1820	1313	170	47
KNN	97.88	97.77	98.45	98.11	0.98	1838	1441	42	29
GB	5.46	94.59	97.43	95.99	0.95	1819	1379	104	48
DT	83.37	85.18	84.95	85.06	0.83	1586	1207	276	281
SVM	96.09	96.31	96.68	96.50	0.96	1805	1414	69	62
MLP	96.93	97.77	98.45	97.24	0.97	1815	1432	51	52

On the other hand, the results of applying deep learning models are presented in Table 6 and Table 7. For each model, a batch size of 32 was employed and 100 epochs were conducted for training. As shown in Table 6, using the raw EEG signals led to poor performance, with high false positive rates. The best accuracy achieved in this case was around 55.88% using the BI-LSTM model. To improve performance, we retrained all the models using features extracted from the signals. This led to a significant improvement, as seen in Table 7. Among the tested models, LSTM and CNN performed the best, with very similar results. LSTM achieved an accuracy of 81.25%, while CNN followed closely at 80.66%. In terms of precision, recall, F1 score, and AUC, both models showed strong and consistent performance. A broader comparison of previous studies that used the same dataset (Dataset 1) for ADHD classification is provided in Table 8 of the appendix.

Table 6. Results of DL models on raw signals of dataset1.

Classifier	Accuracy (%)	Precision(%)	Recall (%)	F1 score (%)	AUC	Confusion matrix			
						TP	TN	FP	FN
BI-LSTM	55.88	58.22	73.81	65.09	0.54	1378	494	989	498
LSTM	55.73	58.02	74.40	65.20	0.53	1389	478	1005	478
GRU	51.76	55.42	68.72	61.36	0.50	1283	451	1032	584
CNN	53.07	55.25	83.18	66.40	0.49	1553	225	1258	314

https://doi.org/10.71058/jodac.v10i01004

Table 7. Results of DL models on extracted features of dataset1.

Classifier	Accuracy (%)	Precision(%)	Recall (%)	F1 score (%)	AUC	Confusion matrix			
						TP	TN	FP	FN
BI-LSTM	75.52	77.19	79.59	78.38	0.75	1627	1075	408	240
LSTM	81.25	80.06	88.38	84.01	0.80	1486	1044	439	381
GRU	73.28	74.85	78.41	76.59	0.73	1650	1072	411	217
CNN	80.66	79.95	87.15	83.39	0.80	1464	991	492	403

Dataset 2: The results shown in Table 8, highlight how classical machine learning models performed on dataset 2. KNN and SVM exhibited the highest accuracy, achieving approximately 98.30 % and 98.20% respectively. Both models also presented strong results across other key metrics like precision, recall, F1-score, and AUC-ROC, reflecting their overall robustness. While the other classifiers give less accuracy than KNN and SVM, they still showed strengths in other areas such as precision, recall, and F1-score, making them useful in different contexts.

Table 8. Classical ML classifiers' results using dataset 2.

Classifier	Accuracy (%)	Precision(%)	Recall (%)	F1 score (%)	AUC(%)	Confusion matrix			
						TP	TN	FP	FN
LR	82.29	84.50	91.64	87.92	75.87	1208	4372	802	399
RF	92.80	90.94	99.71	95.12	88.06	1536	4757	474	14
KNN	98.30	98.78	98.81	98.79	97.96	1952	4714	58	57
GB	86.93	85.32	98.34	91.37	79.10	1203	4692	807	79
DT	83.22	88.03	88.14	88.08	79.83	1438	4205	572	566
SVM	98.20	98.28	99.18	98.73	97.53	1927	4732	83	39
MLP	97.91	98.19	98.85	98.52	97.26	1923	4716	87	55

When it comes to deep learning models, the results show some interesting trends. As detailed in Table 11, CNN performed best on raw EEG signals with an accuracy of 72.04%, closely followed by BI-LSTM (71.26%), LSTM (70.70%), and GRU (69.92%). However, when we shifted to using extracted features (Table 10), GRU achieved the highest accuracy of 76.43%, while BI-LSTM, LSTM, and CNN followed with 73.99%, 71.89%, and 70.36% respectively. Notably, CNN achieved a perfect recall of 100%, highlighting its strength in identifying all true positive cases. These findings provide valuable insight into the strengths of various deep learning models on this dataset.

Table 9. Results of DL models on raw signals of dataset 2

Classifier	Accuracy (%)	Precision(%)	Recall (%)	F1_score (%)	AUC (%)	Confusion matrix			
						TP	TN	FP	FN
BI-LSTM	71.26	73.13	93.50	82.07	55.98	371	4461	1639	310
LSTM	70.70	72.66	93.57	81.80	54.99	330	4464	1680	307
GRU	69.92	73.92	88.45	80.53	57.19	521	4220	1489	551
CNN	72.04	74.42	91.83	82.21	58.45	504	4381	1506	390

 https://doi.org/10.71058/jodac.v10i01004

Table 10. Results of DL models on extracted features of dataset 2

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)	AUC (%)	Confusion matrix			
						TP	TN	FP	FN
BI-LSTM	73.99	75.41	93.52	83.50	60.57	555	4462	1455	309
LSTM	70.36	70.36	100.00	82.60	50.00	0	4771	2010	0
GRU	76.43	77.75	93.17	84.76	64.94	738	4445	1271	326
CNN	71.89	72.95	95.43	82.69	55.73	322	4553	1688	218

4. Discussion

In the current study, the aim is to build a robust and comprehensive approach to ADHD classification using EEG data. To do this, a diverse combination of features was extracted from the signals, including time domain, frequency domain, morphological, and non-linear features. This extensive feature set provided a richer and more complete representation of the brain activity patterns than previous studies, which often focused on a narrower feature subset. These features captured much relevant information as possible from the EEG signals to improve model performance. However, working with a huge number of features can introduce noise and lead to overfitting. To address this, the LASSO method was applied for feature selection. LASSO helps by automatically removing less informative features, keeping only those that contribute significantly to the classification task. This step was crucial in reducing dimensionality and improving model generalization. Notably, this method was also used by Maniruzzaman et al. [7], demonstrating its effectiveness in similar contexts. With the most relevant features selected, several machine learning models were trained. The KNN classifier achieved the highest accuracy of 97.88%on dataset1 and 99.30 on dataset2. KNN's performance can be attributed to its strength in handling well-separated data clusters and its ability to make decisions based on the proximity of data points, making it a good fit for this kind of structured EEG data.

While many previous studies in this field have predominantly focused on traditional machine learning approaches, in this work, we broadened the investigation to include deep learning techniques. We trained four deep learning models using the same set of carefully selected features obtained after applying LASSO for feature reduction. Among these models, the best achieved an accuracy of 81.25%, which, although lower than the top-performing machine learning model (KNN), was an anticipated outcome. Deep learning models typically excel when trained on very large datasets, as they rely on extensive data to learn complex patterns and avoid overfitting. In our case, while the dataset was rich in relevant information, its size may not have been sufficient to fully leverage the depth and capacity of the neural architectures employed.

Nevertheless, incorporating deep learning into the study was a valuable step. It allowed to better understand how these models perform relative to traditional machine learning models when applied to the same problem and feature space. Moreover, it provided important insights into the limitations and challenges of using deep learning in smaller biomedical datasets, such as EEG-based ADHD classification. Our findings suggest that while machine learning models like KNN

https://doi.org/10.71058/jodac.v10i01004

remain highly effective under current conditions, deep learning models hold great promise if larger and more diverse datasets become available.

In future work, the exploration of end-to-end deep learning pipelines; where models learn directly from raw EEG signals without extensive feature engineering could potentially lead to even greater performance. Additionally, strategies such as data augmentation, transfer learning, and the use of pre-trained models could help mitigate the limitations imposed by smaller datasets. Overall, this work demonstrates the importance of selecting the appropriate modelling strategy based on data characteristics and highlights exciting future opportunities for advancing ADHD classification research through deep learning.

In this study, we compared our findings on Dataset 2 with those reported in prior research that utilized the same dataset [26]. The earlier work adopted a ternary classification framework, categorizing subjects into three distinct groups: ADHD, ADD, and healthy controls. However, their deep learning models encountered challenges in accurately distinguishing between ADD and ADHD patients. In contrast, our approach involved combining the ADD and ADHD categories into a single group. This binary classification strategy achieved a significant improvement in classification accuracy and predictive performance. These results suggest that fusion ADD and ADHD cases may offer a more robust and effective framework for analysing Dataset 2, particularly when employing deep learning methodologies.

5 - Conclusion

The advancement of machine learning (ML) and deep learning models contributes significantly to the diagnosis of ADHD and the subsequent alleviation of its severe symptoms. The EEG data was investigated in distinguishing between individuals diagnosed with ADHD and those who are healthy control participants. Several ML methods were tested to improve the performance of ADHD diagnosis, by applying distinct brain patterns that were collected through EEG recordings. This paper explores the accuracy of these models in detecting ADHD based on various biological factors. The current research indicates that KNN classification outperforms all other algorithms, achieving a classification accuracy of 98.3%. The KNN method demonstrates superior performance compared to alternative algorithms in predicting brain ADHD, as evidenced by the findings of the study utilizing 5-fold cross-validation. In the future, a broader dataset will be employed alongside improvements to Deep Learning framework models, to assist the public in assessing ADHD in paediatric patients.

References

[1] "Center For Inclusive Child Care," [Online]. Available: https://www.inclusivechildcare.org/resource-library/website/national-resource-center-adhd.

[2] "The National Institute of Mental Health (NIMH)," [Online]. Available: https://www.nimh.nih.gov/health/topics/attention-deficit-hyperactivity-disorder-adhd.

[3] P. S. a. C. B. Herbert H. Jasper, "Electroencephalographic analyses of behavior problem children.," *American Journal of Psychiatry,* vol. 95, no. 3, pp. 641-658, 1938.

https://doi.org/10.71058/jodac.v10i01004

[4] J. F. Lubar, " Discourse on the Development of EEG Diagnostics and Biofeedback for Attention-Deficit/Hyperactivity Disorders. Biofeedback and Self-Regulation," *Biofeedback and Self-Regulation,* vol. 16, pp. 201-225, 1991.

[5] M. H. M. T. a. V. A. Farnaz Ghassemi, "Using non-linear features of EEG for ADHD/normal participants' classification," vol. 32, pp. 148-152, 2012.

[6] M. H. I. Anika Alim, "Automatic Identification of Children with ADHD from EEG Brain Waves," *Signals,* vol. 40, pp. 193-205, 2023.

[7] M. A. M. H. N. A. a. J. S. Md Maniruzzaman, "Optimal Channels and Features Selection Based ADHD Detection From EEG Signal Using Statistical and Machine Learning Techniques," *IEEE,* vol. 11, pp. 33570 - 33583, 2023.

[8] A. K. A. M. N. S. R. M. B. a. H. Z. Mohammad Reza Mohammadi, "EEG classification of ADHD and normal children using non-linear features and neural network," *Biomedical Engnineering Letters,* vol. 6, pp. 66-73, 2016.

[9] A. B. R. ,. S. a. C. B. Amirali Vahid, "Deep Learning Based on Event-Related EEG Differentiates Children with ADHD from Healthy Controls," *Journal of clinical medicine,* vol. 8, no. 7, 2019.

[10] K. S. K. Amandeep Kaur, "Accurate Identification of ADHD among Adults Using Real-Time Activity Data," *Brain Sciences,* vol. 12, no. 7, 2022.

[11] M. Y. E. a. F. Latifoğlu, "Detection of ADHD from EEG signals using new hybrid decomposition and deep learning techniques," *Journal of Neural Engineering,* vol. 20, no. 3, 2023.

[12] A. A. M. S. M. R. M. Ali Motie Nasrabadi, "EEG data for ADHD / Control children," *IEEE Dataport,* 2020.

[13] C. Beste, "Deep learning EEG," 2018.

[14] R. R. R Muthukrishnan, in *IEEE International Conference on Advances in Computer Applications*, Coimbatore, India, 2016.

[15] R. T. M. W. Trevor Hastie, Statistical Learning with Sparsity, The Lasso and Generalizations, New York: Chapman and Hall/CRC, 2015.

[16] J. Berkson, "Application of the Logistic Function to Bio-Assay," *Journal of the American Statistical Association,* vol. 39, no. 227, pp. 357-365, 1944.

[17] L. BREIMAN, "Random Forests," *Machine Learning,* vol. 45, 2001.

[18] P. H. T. Cover, *IEEE Transactions on Information Theory ,* vol. 13, no. 1, pp. 21-27.

[19] J. H. .Friedman, "Greedy function approximation: A gradient boosting machine.," *Annals of Statistics,* vol. 29, no. 5, p. 1189–1232, 2001.

[20] J. QUINLAN, "Induction of decision trees," *Machine learning,* vol. 1, pp. 81-106, 1986.

[21] G. H. a. R. J. D.E. Rumelhart, "Learning representations by back-propagating errors.," *Nature,* vol. 323, no. 6088, pp. 533-536, 1986.

[22] J. S. Sepp Hochreiter, "Long Short-Term Memory," *Neural Computation,* vol. 9, no. 8, pp. 1735 - 1780.

[23] J. S. Alex Graves, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks,* vol. 18, pp. 602-610, 2005.

https://doi.org/10.71058/jodac.v10i01004

[24] B. v. M. C. G. D. B. F. B. H. S. Y. B. Kyunghyun Cho, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” *arXiv preprint arXiv:1406.1078,* vol. 1, 2014.

[25] Y. B. ,. G. H. Yann LeCun, “Deep learning,” *Nature,* vol. 521, p. pages436–444, 2015.

[26] A. B. V. R. S. S. C. B. Amirali Vahid, “Deep Learning Based on Event-Related EEG Differentiates Children with ADHD from Healthy Controls,” *Journal of clinical medicine,* vol. 8, no. 7, 2019.

[27] N. K. J. S. R. K. G. Anshu Parashar, “Machine Learning Based Framework for Classification of Children with ADHD and Healthy Controls,” *Intelligent Automation & Soft Computing ,* vol. 28, no. 3, pp. 669-682, 2021.

[28] A. M. N. M. M. Ali Ekhlasi, “Classification of the Children with ADHD and Healthy Children Based on the Directed Phase Transfer Entropy of EEG Signals,” *Frontiers in Biomedical Technologies ,* vol. 8, no. 2, pp. 115-122 , 2021.

[29] “AI-techniques-for-ADHD-detection,” [Online]. Available: https://github.com/bangkook/AI-techniques-for-ADHD-detection/blob/39749bc0138eb94a610a07e745c5245e06f9b6ef/Final_Data_1.ipynb.

[30] “ADHD-detection-from-EEG-signals-using-ML-and-DL-models,” [Online]. Available: https://github.com/sara-gaballa/ADHD-detection-from-EEG-signals-using-ML-and-DL-models.

Appendix

Table 11 Comparison with our work on dataset 1 against existing published similar work on the same dataset.

Authors	Extracted features	Feature selection	Classifiers	ACC(%)
Mohammadi et al. [8]	NL	mRMR, DISR	MLP	93.70
Parashar et al. [27]	---	---	AB, RF, SVM	84.00
Ekhlasi et al. [28]	----	GA	ANN	89.70
Maniruzzaman et al. [7]	MI, TD	t-test, LASSO	SVM,KNN, MLP, LR	94.20
Maniruzzaman et al. [7]	TD, MI,	LASSO	GPC, RF, KNN, MLP, DT, LR.	97.50
Alim & Imtiaz [6]	TD, FD	ANOVA	SVM	94.20
Our work	TD, MI, NL, FD	LASSO	ML: LR, RF, KNN, FD GB, DT, SVM, MLP.	97.88
			DL: BI-LSTM, LSTM, GRU, CNN.	81.25

NL: Non-linear; TD: Time domain; MI: Morphological; FD: Frequency domain; mRMR: Minimum redundancy and maximum relevance; DISR: Double input symmetrical relevance; GA: Genetic algorithm.

https://doi.org/10.71058/jodac.v10i01004

A Data availability statement

The datasets and notebooks are freely available on GitHub : Dataset 1 [12]-

Notebook [29], Dataset 2 [13] – Notebooks [30].

B Conflict-of-interest

The authors have no relevant financial or non-financial interests to disclose.

C An overview of the classification models employed in our study.

- Logistic Regression: Logistic Regression is a linear classification algorithm that models the relationship between the input features and the binary outcome using the logistic function.
- Random Forest: Random Forest is an ensemble learning technique that combines multiple decision trees to improve predictive accuracy. Aggregating the outcomes of individual trees helps mitigate overfitting and enhance model robustness.
- K-Nearest Neighbours (Neighbours): K-Nearest Neighbours it is instance-based learning model classifies a data point based on the class of its k-nearest neighbours in the feature space. It relies on the similarity of instances to determine the class of a new observation.
- Gradient Boosting: Gradient Boosting classifier is an ensemble technique that constructs a strong predictive model by iteratively adding weak learners, such as decision trees. Each new learner focuses on the errors made by the previous ones, leading to improved accuracy.
- Decision Tree: The Decision Tree classifier creates a hierarchical structure of nodes, each representing a feature test, to make classification decisions. It divides the feature space into distinct regions and assigns class labels, accordingly, offering interpretability and adaptability.
- Support Vector machine (SVM): SVM is an instance-based learning model that can be used for classification problems. It works by finding the hyperplane that best separates the two classes of data points. The hyperplane is the line or curve that has the maximum margin between the two classes. The SVC algorithm finds the hyperplane that minimizes the misclassification error, while also maximizing the margin between the two classes.
- MLP (Multi-Layer Perceptron): The MLP is an artificial neural network comprising interconnected layers of nodes. Its ability to capture intricate relationships suits data with non-linear decision boundaries.
- Long Short-Term Memory (LSTM): LSTM is a type of recurrent neural network (RNN) that can process sequential data by maintaining a hidden state that can store long-term dependencies. The LSTM has a cell structure that consists of three gates: an input gate, a forget gate, and an output gate. These gates regulate the flow of information into and out of the cell, allowing the LSTM to learn what to remember and what to forget.
- Bidirectional LSTM (BI-LSTM): BI-LSTM is a variant of the LSTM that consists of two LSTMs: one that processes the input sequence in the forward direction, and another that processes it in the backward direction. BI-LSTM can capture both past and future contexts from the input sequence, which can improve the performance.

https://doi.org/10.71058/jodac.v10i01004

- Gated Recurrent Unit (GRU): GRU is a type of RNN that can process sequential data with long-term dependencies. GRU is similar to LSTM but has a simpler structure that consists of two gates: a reset gate and an update gate. The update gate controls how much of the previous state (reset gate) to carry forward.
- One-Dimensional Convolutional Neural Network (1D-CNN): 1D-CNN is a type of convolutional neural network (CNN) that can process one-dimensional data, such as text, audio or signals. 1D-CNN applies convolutional filters along the temporal dimension of the input sequence, extracting local features at different levels of abstraction. 1D-CNN can reduce the need for manual feature engineering and handle variable-length inputs.