An artificial intelligence techniques for classifying attention deficit hyperactivity disorder (ADHD) based on electrophysiology

Lamiaa A. Amar[1*], Mennatallah Magdy[2], Sara Gaballa[3], Mariam Hossam[3], Ahmed.M.Otifi [4]

[1*]Department of Networks and Distributed Systems, Informatics Research Institute, City of Scientific Research and Technological Applications, SRTA-CITY, Alexandria , 21934, Egypt.
[2]Department of Healthcare Analytics , Faculty of Computers and Data Science, Alexandria University , Alexandria University, Alexandria, 21554, Egypt.
[3]Department of Computer and Systems Engineering, Faculty of Engineering , Alexandria University, Alexandria, 21544, Egypt.
[4]Department of data science , Faculty of Computers and Data Science, Alexandria University , Alexandria University, Alexandria, 21554, Egypt.

*Corresponding author(s). E-mail(s): lamy2003@hotmail.com; Contributing authors: Mennamego555@gmail.com ; saragaballa2002@gmail.com; m.hossam2551@gmail.com; ahmedotifi11112000@gmail.com;

Abstract

ADHD as a prevalent psychiatric disorder in childhood. The prevalence of this disease in the community has risen steadily from the past to the present. While psychiatric tests are used to make the ADHD diagnosis, there's lack of active clinically used objective diagnostic tool for ADHD. Previous studies suggesting the potential of EEG signals for ADHD diagnosis.The goal of this study was to create an objective screening tool for ADHD utilising electroencephalography (EEG) signals. We used two benchmark dataset with large number of features, extracted 24 features from the EEG signals belonging to different categories: time domain, frequency domain, morphological and non-linear. We employed the LASSO-LR model to select the best features and used as input data for machine learning the deep learning algorithms developed in the study. by comparing between the models the KNN algorithm achived the best accuracy in the two data in distinguishing between ADHD patients and healthy people. The current study adds to the literature through originality, and machine learning methods could be used as a clinical instrument in the near future, as they are more suitable for small eeg data.

Keywords: ADHD, classification, machine learning, deep learning

1 Introduction

An electroencephalogram (EEG) records brain activity using sensors attached to the scalp that record electrical signals produced by the brain [1]. It has been a useful tool for the diagnosis of brain disorders such as Alzeheimer's disease, epilepsy, and sleep disorders by detecting abnormalities in the brain waves for example Attention-deficit/hyperactivity disorder (ADHD).

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders of childhood which affects 11 percent of U.S. school-age children [2]. It is marked by an ongoing pattern of inattention, impulsivity and/or hyperactivity that interferes with functioning or development [3]. It is generally first diagnosed in children starting from the age of 3 and often lasts into adulthood. Males are more prone to have ADHD than females but with different behavioral patterns. Without identification and proper treatment, ADHD may have serious consequences, including academic failure, family stress and disruption, depression, problems with relationships, and job failure. Early identification and treatment are of great value [2].

Jasper et al were the first to use EEG in ADHD over 75 years ago [4] . Then, J. Lubar in 1973 conducted the first abnormalities study by EEG signals in ADHD [5]. He reported increased activity in the theta band (4–8 Hz) with dramatic reduction in the power of beta (13–30 Hz) band. A common pattern in the brain waves of ADHD patients is abundance of slow (delta or theta) brain waves and a shortage of fast (beta) brain waves. Therefore, they have a high theta-to-beta ratio which could be used to automatically detect ADHD from the brain waves. [6]

Recently, fMRI (functional magnetic resonance imaging) and MRI have been used by researchers to diagnose ADHD. EEG, on the other hand, is more widely used, less expensive, non-invasive, portable, and provides understandable insights into how the brain functions.

The EEG waves include a big and complex quantity of data. Because of this, manually detecting abnormalities is difficult for a human. This is the situation when machine learning is helpful [6]. In general, machine learning is a developing technology that allows computers to automatically learn from previous data and can be applied to the present task.

In this study, we investigate how different machine learning and deep learning techniques would perform on the classification of ADHD children and healthy controls. We apply an EEG ML/DL pipeline summarized as follows:

- We extracted different types of time domain, frequency domain, morphological and non-linear features from EEG signals.
- We employed LASSO-LR method for feature selection and reduction.
- We selected a set of seven traditional ML classifiers: LR, RF, KNN, GB, DT, SVM, and MLP, and a set of four DL models: LSTM, BI-LSTM, GRU and CNN. The models were trained on each one of the two datasets to identify ADHD patients from healthy controls.
- We performed 5-fold cross validation to avoid bias and overfitting. Finally, five performance metrics were used to evaluate our approach.

The rest of the paper is organized as follows. First, Literature Review is presented in Section 2. Materials and methods are explained in Section 3 including data description, data preprocessing, feature extraction, feature selection, classification models and evaluation metrics.The experimental results and discussion are presented in Section 4. Finally, conclusion is presented in Section 5.

2 Literature Review

Several studies have been conducted to suggest an automated system for the early identification of patients with ADHD. They are summarized in Table 1.

Table 1 Summary of recent and related works.

Reference	Data Size	Feature Extraction	Feature Selection	Model	ACC
[7]	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%
[5]	120	Linear	ANOVA	ML(SVM)	94%
[11]	121	Time Domain Morphological Non-linear	LASSO	ML(GPC, RF, KNN, MLP, DT, LR)	97.53%
[12]	-	-	-	DL	87%

Gahssemi et al. [7] worked on the classification of adult normal and ADHD participants who went through a continuous performance test (CPT). They extracted three non-linear features: Wavelet entropy, Correlation dimension, and Lyapunov exponent to train the KNN classifier. They achieved an accuracy of 96% on test data using only two features. Mohammadi et al [8] utilized the EEG signals collected from 30 children with ADHD and 30 healthy children for classification. They extracted non-linear features which are fractal dimension, approximate entropy, and Lyapunov exponent from the EEG signals. They achieved the accuracy of 92.28% and 93.65% using minimum Redundancy Maximum Relevance (mRMR) and double input symmetrical relevance (DISR) using multilayer perceptron (MLP), respectively. A different work done by Vahied et al. [9] was to classify ADHD subtypes - ADD and ADHD - as well. They used the data collected from 144 participants – 44 healthy children, 52 with ADD, and 48 with ADHD – to perform a three-class classification. Using the deep learning model EEGNET, they achieved an accuracy up to 83%. Kaur and Kahlon [10] worked on the identification of ADHD among adults using a dataset collected from 51 patients with ADHD and 52 clinical controls. They extracted a total of 788 statistical features then applied PCA to select the significant features. These features are then fed into six ML models which are C4.5 Decision Tree, Random Forest, LogitBoost, SVM and NAive Bayes. The evaluation showed a higher accuracy of 98.43% using SVM.

For recent studies, Alim and Imtiaz [5] extracted linear features from EEG data to train a Gaussian SVM-based model. They reduced the computational load by eliminating signals of more than 30 Hz and used only the first four subbands of EEG achieving a mean accuracy of 94%. Maniruzzaman et al [11] focused on the optimal selection of channels and features. They combined two separate methods (support vector machine and t-test) to select optimal channels. After that, LASSO logistic regression-based model was used to select the important features. Finally, six ML-based classifiers were applied for the detection of ADHD achieving an accuracy rate of 97.53%. In the work of Esas and Latifoğlu [12], EEG signals were decomposed into subbands by robust local mode decomposition and variational mode decomposition techniques. They designed a deep learning algorithm that resulted in classification accuracy over 87.

We worked on two datasets with a different number of channels (19, 56). By training different ML and DL models on each data set, we could evaluate our approach to each one. Then, we make a comparison between the obtained results from each data set. We focus on how the identification of ADHD children could be affected by the size of the data set, validity of the data, and number of channels. Finally, we hold a comparison between our work on each data set and previous similar works on the same data set if found.

3 Materials and Methods

3.1 Data Description

This paper utilized the input EEG signals from two datasets described as follows: **Dataset 1:** The dataset is available online in the IEEE data port [13]. The dataset was collected from 121 children – 60 healthy children and 61 children with ADHD. Their ages lied in the range from 7 to 12 years. The EEG signals were collected from 19 channels: (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2) with A1 and A2 electrodes as references on earlobes.

The children were shown some images and asked to count the number of age-suitable characters. The duration of the task differed for every subject. The data was collected from two different sessions; thus, it is divided into two folders for each type. Each folder contains files in mat format showing the EEG recording of the corresponding subject.

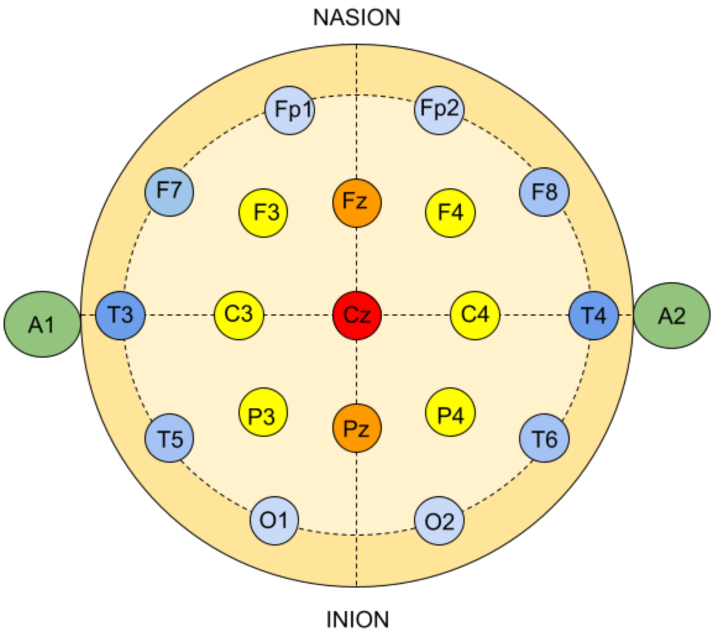

Fig. 1 The placement of 10-20 system-based electrodes.

Table 2 Information about the participants of dataset 1.

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

Dataset 2: The dataset is available online at OSF [14] .Data was collected from 144 children - 44 healthy children and 100 with disorder.

EEG was recorded from 60 Ag/AgCl electrodes using a BrainAmp amplifier (Brain Products Inc, Munich, Germany). For further analysis, electrodes P9, P10, P11, and P12 were removed from the data set due to their high susceptibility to high electrode impedances and thus unreliable EEG recordings. Thus, 56 electrodes were used.
After data segmentation, it is classified as follows: N = 10,129 trials were included for the control group, N = 13,031 trials were included for the ADD group, and N = 10,742 trials were included for the ADHD group.

Table 3 Information about the participants of dataset 2.

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	$11.3 \pm$ 2d.2	103 ± 12

3.2 Data Preprocessing

EEG signals of dataset 1 contain different artifacts and noises that should be removed before analysis. The sampling frequency of the EEG signals is 128 Hz. For the preprocessing method, we used a bandpass filter with cutoff frequencies at 0.5 Hz and 63 Hz. A 50 Hz notch filter was applied to remove power line noise. Finally, the data was divided into 2s segments for every channel with 1s overlap.

The data was converted to the following shape (16749, 19, 256), where the dimensions correspond to the number of samples, number of channels, and number of timestamps respectively.

3.3 Feature Extraction

Feature extraction was proven to be a very important step in classification. It is the process of converting and reducing raw data to a set of useful and meaningful information. The proper selection of features enhances the performance of the applied model.

In this work, we extracted 24 features from the EEG signals belonging to different categories: time domain, frequency domain, morphological and non-linear. Table 4 gives a brief description of these features.

For each window, we extracted the 24 features for every channel except the PSD (Power Spectral Density) which was extracted for every frequency band as well, since ADHD patients show a variant in the power of some bands.

3.4 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 the LASSO-LR [15] model with a 5-fold CV. We applied GridSearchCV 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 [16]. The cost function of the LASSO model is shown in the following equation:

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

Then, Features with an importance value higher than 0.02 were selected. See Table 5 for more details.

Table 4 Description of EEG features extracted in this study.

Feature	Description	Formula
Mean	Mean	$\frac{1}{n}\sum_{t=1}^{n} z_t$
Std	Standard deviation	$\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(z_t-\mu)^2}$
Ptp	Peak to peak	$max\lvert z_t\rvert - min\lvert z_t\rvert$
Var	Variance	$\frac{1}{n-1}\sum_{t=1}^{n}(z_t-\mu)^2$
Min	Minimum value	$min\lvert z_t\rvert$
Max	Maximum value	$max\lvert z_t\rvert$
RMS	Root mean square	$\sqrt{\frac{\sum_{t=1}^{n}(z_t)^2}{n}}$
AD	Absolute difference	$\sum_{t=1}^{n-1}\lvert z_{t+1}-z_t\rvert$
Median	Median	$Median(z_t)$
Q1	1st quartile	$Q1(z_t)$
Q3	3rd quartile	$Q3(z_t)$
Skewness	Skewness	$\frac{\sum_{t=1}^{n}(z_t-\mu)^3/n}{\sigma^3}$
Kurtosis	Kurtosis	$\frac{\sum_{t=1}^{n}(z_t-\mu)^4/n}{\sigma^4} - 3$
Mobility	Hjorth Parameter Mobility	$\sqrt{\frac{var(z`(t))}{var(z(t))}}$
Complexity	Hjorth Parameter Complexity	$\sqrt{\frac{Mob(z`(t))}{Mob(z(t))}}$
ENG	Energy	$\sum_{t=1}^{n} z_t^2$
POW	Power	$\frac{1}{n}\sum_{t=1}^{n} z_t^2$
PFD	Pertrosian Fractal Dimension	$\frac{\log_{10}(N)}{log_{10}(N)+log_{10}(\frac{N}{N+0.4N_\delta})}$
KFD	Katz Fractal Dimension	$\frac{\log_e(N-1)}{log_e(N-1)+log_e(\frac{d}{L})}$
HFD	Higuchi Fractal Dimension	$\frac{1}{K}\sum_{M=1}^{k} L_m(k)$
DFA	Detrend Fluctuation Analysis	$L^H * \sigma(z_t)$
SEN	Spectral Entropy	$\frac{-\sum_{n=0}^{N-1} P_n \log_2(P_n)}{log_{10}(N)}$
Hurst	Hurst exponent[]	$\log_n(E[\frac{R(n)}{S(n)}]/C)$
PSD	Power Spectral Density	

N: Length of the time series; N_δ: No. of sign changes in the signal derivatives; L: Sum of the spacing between sequential points; d: Distance between the first and farthest point of EEG signal; H: Hurst parameter; E: Expected value; $R(n)$: The range of the first n cumulative deviations from the mean; $S(n)$: Sum of the first n standard deviations; n: Number of data points, and C is constant.

Table 5 Details about the feature selection process.

Data	Number of features		Alpha
	Before Selection	After Selection	
Dataset 1	532	96	0.001
Dataset 2	1568	85	0.001

The selected features were then used in ML classification models to differentiate between healthy and ADHD children.

3.5 Classification Models

An overview of the classification models employed in our study. Each model offers a unique approach to tackling the classification task, capturing various aspects of the data.

- **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 Neighbors (KNeighbors):** K-Nearest Neighbors it is instance-based learning model classifies a data point based on the class of its k-nearest neighbors 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 Classifier (SVC):** The Support Vector Classifier is a 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.

- **MLPClassifier (Multi-Layer Perceptron):** The MLPClassifier 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.

- **Gated Recurrent Unit (GRU):** GRU is type of RNN that can process sequential data with long-term dependencies. GRU is similar to the LSTM, but has a simpler structure that consists of two gates: a reset gate and an update gate. The reset gate controls how much of the previous state to forget, and the update gate controls how much of the previous state 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.

- **Multi-headed models:** Multi-headed models are deep learning models that have multiple output heads, each output head is responsible for predicting a different output, such as a different class label or a different feature. The main advantage of multi-headed models is that they can learn more complex relationships between the input and output data. This is because each output head can focus on learning a different aspect of the relationship and it can built on top of a recurrent neural network (RNN) like [BI-LSTM, LSTM, GRU] or a convolutional neural network (CNN) like 1D-CNN.

3.6 Evaluation Metrics

Five performance metrics were used to evaluate the models which are accuracy (ACC), Precision (Prec), Recall (Rec), F1_score and area under the ROC (Receiver Operating

Characteristic) curve (AUC). Using the values of true positive (TP), true negative (TN), false positive (FP), false negative (FN), false positive rate (FPR), and true positive rate (TPR), the metrics are calculated according to the following equations:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Prec = \frac{TP}{TP + FP} \tag{3}$$

$$Rec = \frac{TP}{TP + FN} \tag{4}$$

$$F1\_score = 2 * \frac{Prec * Rec}{Prec + Rec} \tag{5}$$

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

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

4 Experimental Results and Discussion

In this work, first, we normalized the input data and features using the formula given at equ.7.

$$z = \frac{X - \mu}{\sigma} \tag{7}$$

- X is the original feature vectors or input data.
- μ and σ are the mean and standard deviation of the respective vectors.
- z is the standardized value and its values lie between 0 to 1.
  Then, we applied a 5-fold cross-validated grid-search in the feature selection and classification processes. It is used to fine-tune the hyper-parameters of the utilized ML algorithms by selecting the combination of parameters giving the best score from a set of different hyper-parameters in each model.

Dataset 1: The results are shown in Table 6 based on applying the ML classifiers on **dataset 1** after splitting the data with the ratio 80:20 and using a 5-fold CV. We noticed that KNN gave relatively better performance scores than other classifiers. The KNN classifier provided 97.88% test accuracy, 97.77% precision, 98.45% recall, 98.11% F1_score, and 0.98 AUC. Similarly, MLP provided high scores along with the RBF kernel-based SVM and performed better than GB, RF, DT, and LR. The accuracies of MLP and SVM were 96.93% and 96.09% respectively.

Whereas the results of applying the DL models are shown in Tables 7 and 8. We employed a batch size of 32 and 100 epochs while training each model.

Table 7 shows the results of applying the models to the raw signals. It resulted in degraded performance with large values of false positives. The maximum accuracy reached was roughly 55.88% with BI-LSTM.

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	95.46	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

Table 6 Results of ML classifiers on dataset 1

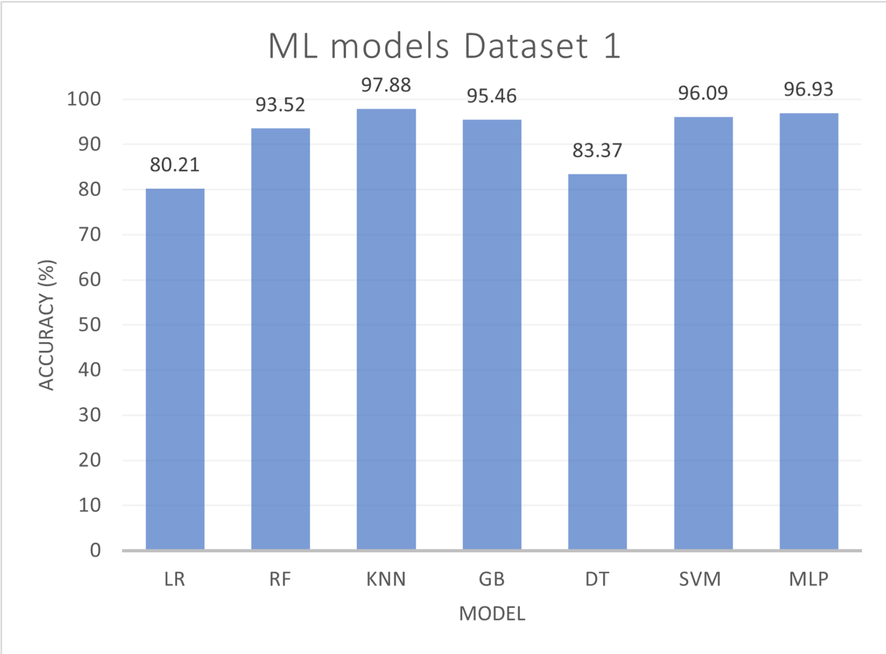

Fig. 2 Accuracy of dataset 1 ML models

We trained the models again on the extracted features of the data which improved the performance considerably as shown in Table 8.

After feature extraction, LSTM and CNN outperformed BI-LSTM and GRU models with nearly equal performance scores. LSTM and CNN provided 81.25%, 80.66% accuracy, 80.06%, 79.96% precision, 88.38%, 87.15% recall, 84.01%, 83.39% f1_score and 0.80, 0.80 AUC, respectively.

The comparison of different studies on the classification of ADHD using the data provided by dataset 1 is briefly illustrated in Table 10.

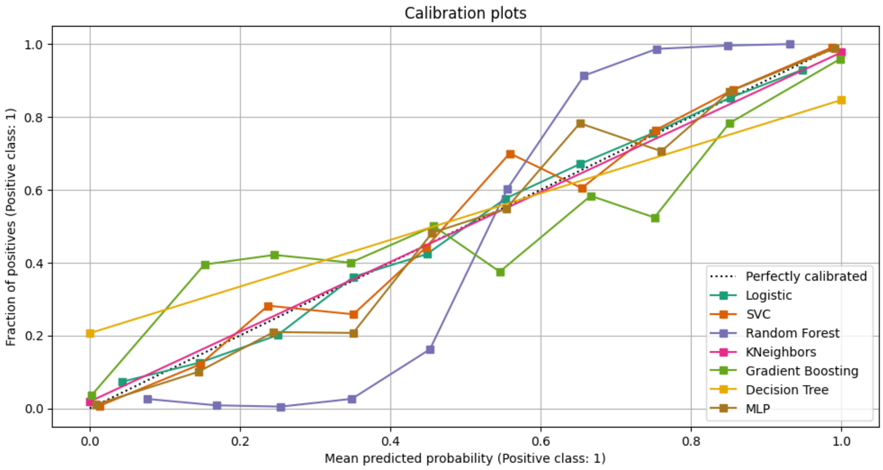

Fig. 3 Calibration curve for ML classifiers on data set 1.

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	489
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

Table 7 Results of DL models on raw signals of dataset 1.

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

Table 8 Results of DL models on extracted features of dataset 1.

In our work, we extracted a different combination of features including time domain, frequency domain, morphological, and non-linear, which includes more features than any other paper in the literature. We applied LASSO for feature selection which was also used by Maniruzzaman et al. [11]. Then, we fed the selected features into different machine-learning models. We achieved an accuracy up to 97.88% with KNN classifier, higher than the previous works in the literature. Most papers used ML models for classification. Therefore, we investigated some deep learning architectures by training four DL models on the extracted features and achieved an accuracy of 81.25%.

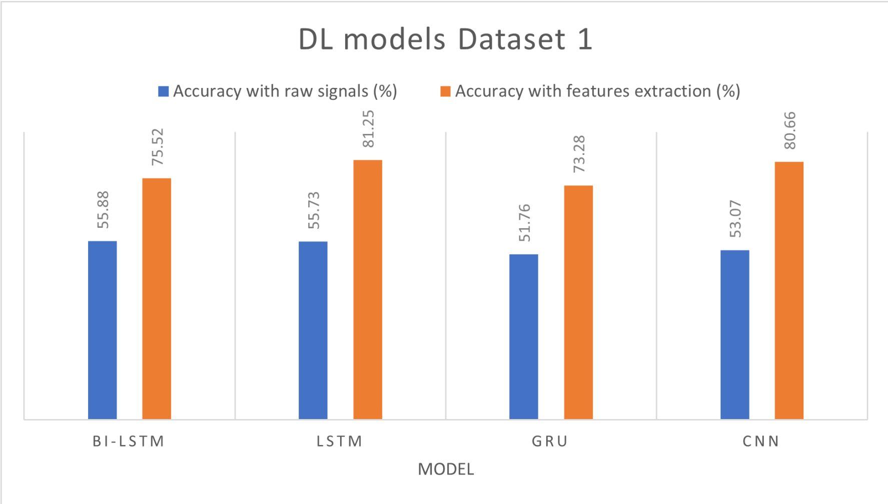

Fig. 4 Comparison of DL model performance on dataset 1 with raw signals against extracted features.

Table 9 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	93.7
Parashar et al. [17]	-	-	AB, RF, SVM	84.0
Ekhlasi et al. [18]	-	GA	ANN	89.7
Maniruzzaman et al. [11]	MI, TD	t-test, LASSO	SVM, KNN, MLP, LR	94.2
Maniruzzaman et al. [11]	TD, MI, NL	LASSO	GPC, RF, KNN, MLP, DT, LR	97.5
Alim & Imtiaz [5]	TD, FD	ANOVA	SVM	94.2
Our proposed study	TD, MI, NL, FD	LASSO	ML: LR, RF, KNN, 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.

Dataset 2: The results in table 10 demonstrate varying levels of performance across the different machine learning models on dataset 2. K-Nearest Neighbors (KNeighbors) and Support Vector Classifier (SVC) exhibited the highest accuracy, achieving approximately 98.30 % and 98.20% accuracy, respectively. These models also displayed impressive precision, recall, F1-score, and AURoc values. the MLP and RF classifiers demonstrated strong performance, with accuracies of 97.91% and

92.80%, respectively. Despite exhibiting lower accuracy, the GB, DT, and LR classifiers showcased diverse strengths in precision, recall, and F1 scores.

In addition to the previously discussed DL models, Table 11 showcases their performance on raw signals. CNN performed the best with an accuracy of 72.04%, followed closely by BI-LSTM (71.26%), LSTM (70.70%), and GRU (69.92%).

Furthermore, Table 12 analyzes DL models on extracted features. GRU displayed the highest accuracy (76.43%), followed by BI-LSTM (73.99%), LSTM (71.89%), and CNN (70.36%). Notably, CNN achieved a perfect recall of 100.0%.

In conclusion, these tables provide a comprehensive overview of the performance of both machine learning and deep learning models on different aspects of dataset 2, allowing for a detailed comparison of their capabilities.

						Confusion matrix			
Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)	AUC (%)	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
SVC	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

Table 10 Results of ML classifiers on dataset 2

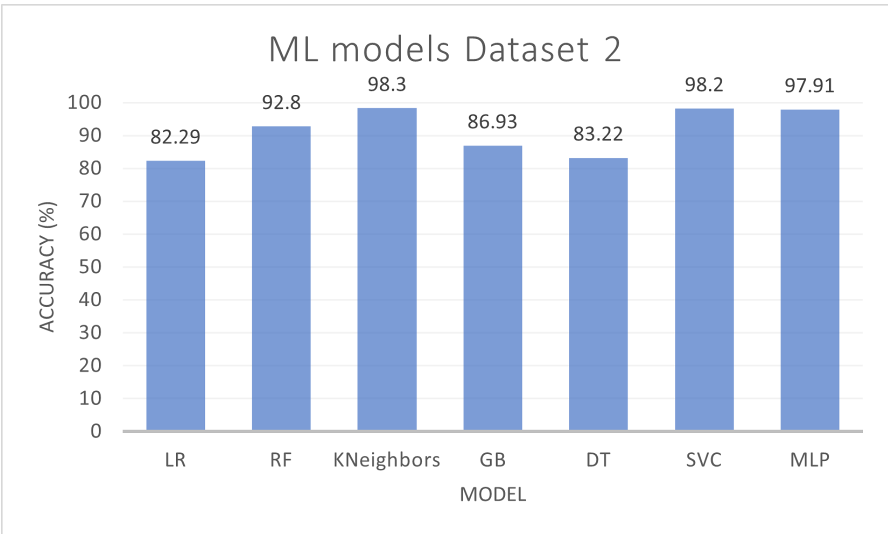

Fig. 5 Accuracy of dataset 2 ML models

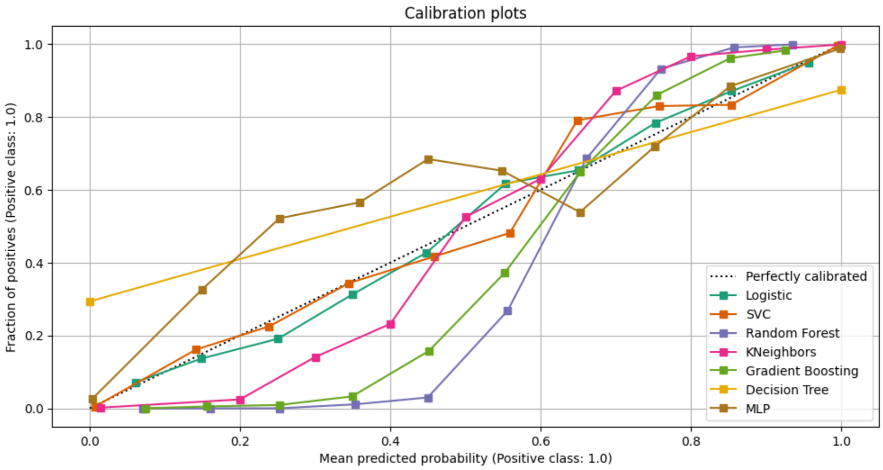

Fig. 6 Calibration curve for ML classifiers on dataset 2.

						Confusion matrix			
Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)	AUC (%)	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

Table 11 Results of DL models on raw signals of dataset 2

						Confusion matrix			
Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)	AUC (%)	TP	TN	FP	FN
BI-LSTM	73.99	75.41	93.52	83.50	60.57	555	4462	1455	309
CNN	70.36	70.36	100.0	82.60	50.0	0	4771	2010	0
GRU	76.43	77.75	93.17	84.76	64.94	738	4445	1272	326
LSTM	71.89	72.95	95.43	82.69	55.73	322	4553	1688	218

Table 12 Results of DL models on extracted features of dataset 2

Finally, we discuss the findings of our study in relation to Dataset 2 and compare them to the results presented in the prior research that utilized the same dataset [19]. The previous study employed a ternary value classification scheme, categorizing subjects into three groups: ADHD, ADD, and healthy control. In their analysis, the deep learning approach encountered challenges in distinguishing between patients with ADD and those with ADHD.

In contrast, in our study, we explored an alternative approach by combining the ADD and ADHD groups. Our results indicated a noteworthy improvement in classification accuracy and predictive performance when compared to the ternary classification

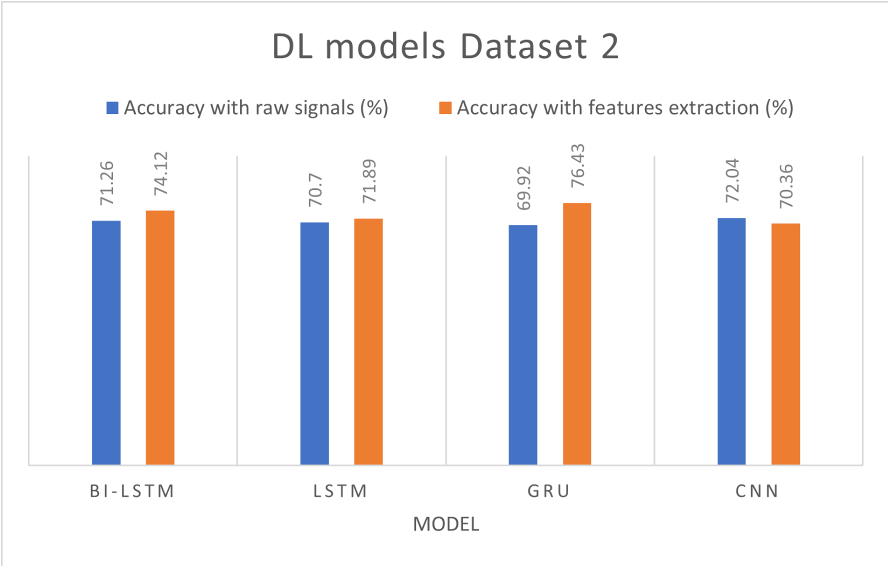

Fig. 7 Comparison of DL models Performance on dataset 2 with raw signals against extracted features.

scheme. This suggests that the fusion of ADD and ADHD data might offer a more robust and effective approach for the analysis of Dataset 2.

5 Conclusion

We investigate the use of EEG data in the classification of ADHD subjects and healthy control participants. We used machine learning (ML) methods to improve the performance of ADHD diagnosis by applying distinct brain patterns that were collected through EEG recordings. Developing the use of a machine learning (ML) and Deep Learning model could play a role in the diagnosis of ADHD and the resulting amelioration of its severe symptoms. The present study explores the accuracy with which different (ML) and (DL) algorithms detect ADHD based on various biological factors. KNN classification trumps all of the other evaluated algorithms with a classification accuracy of 98.3%. The KNN technique outperforms other algorithms when anticipating brain ADHD using 5 fold cross-validation, as demonstrated by the study. The study's future goals could include using a larger dataset or applying the same model to numerous other datasets, as well as improving Deep Learning framework models. The artificial intelligence architecture may assist the general public in assessing ADHD in a child patient.

=====================================

6 Data availability statement

The datasets(1,2) and notebooks are freely available on GitHubs : Dataset 1 [13] - **Notebook** , Dataset 2 [14] - **Notebooks** accessed 22 July 2023.

References

[1] Electroencephalogram (EEG). https://mayfieldclinic.com/pe-eeg.htm Accessed 2023-07-27

[2] http://www.help4adhd.org/NRC.aspx Accessed 2023-07-27

[3] Attention-Deficit/Hyperactivity Disorder. https://www.nimh.nih.gov/health/topics/attention-deficit-hyperactivity-disorder-adhd Accessed 2023-07-27

[4] Lenartowicz A, L.S.: Use of eeg to diagnose adhd. Current psychiatry reports **16**(11), 498 (2014) https://doi.org/10.1007/s11920-014-0498-0

[5] Alim, A., Imtiaz, M.H.: Automatic identification of children with adhd from eeg brain waves. Signals **4**(1), 193–205 (2023) https://doi.org/10.3390/signals4010010

[6] Lubar, J.F.: Discourse on the development of eeg diagnostics and biofeedback for attention-deficit/hyperactivity disorders. Biofeedback Self-Regul **16**, 201–225 (1991)

[7] Ghassemi, F., Hassan-Moradi, M., Tehrani-Doost, M., Abootalebi, V.: Using non-linear features of eeg for adhd/normal participants' classification. Procedia - Social and Behavioral Sciences **32**, 148–152 (2012) https://doi.org/10.1016/j.sbspro.2012.01.024

[8] Mohammadi, K.A.N.A.M.e.a. M.R.: 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

[9] Vahid A, R.V.S.S.B.C. Bluschke A: Deep learning based on event-related eeg differentiates children with adhd from healthy controls. Journal of Clinical Medicine **8**(7), 1055 (2019) https://doi.org/10.3390/jcm8071055

[10] Kaur A, K.K.: Accurate identification of adhd among adults using real-time activity data. Brain Sci. **12**(7), 831 (2022) https://doi.org/10.3390/brainsci12070831

[11] M. Maniruzzaman, N.A. M. A. M. Hasan, Shin, J.: Optimal channels and features selection based adhd detection from eeg signal using statistical and machine learning techniques. IEEE Access **11**, 33570–33583 (2023) https://doi.org/10.1109/ACCESS.2023.3264266

[12] Esas MY, L.F.: Detection of adhd from eeg signals using new hybrid decomposition and deep learning techniques. J Neural Eng. **20**(3) (2023) https://doi.org/10.1088/1741-2552/acc902

[13] Motie Nasrabadi, A., Allahverdy, A., Samavati, M., Mohammadi, M.R.: EEG Data for ADHD / Control Children. https://doi.org/10.21227/rzfh-zn36 . https://dx.doi.org/10.21227/rzfh-zn36 Accessed 2023-07-13

[14] Beste, C.: Deep learning EEG. OSF (2020). osf.io/6594x

[15] R, M., Rohini, R.: Lasso: A feature selection technique in predictive modeling for machine learning, pp. 18–20 (2016). https://doi.org/10.1109/ICACA.2016.7887916

[16] Statistical Learning with Sparsity: The Lasso and Generalizations (1st ed.). Chapman and Hall/CRC (2015). https://doi.org/10.1201/b18401

[17] Parashar, A., Kalra, N., Singh, J., Goyal, R.: Machine learning based framework for classification of children with adhd and healthy controls. Intelligent Automation Soft Computing **28**, 669–682 (2021) https://doi.org/10.32604/iasc.2021.017478

[18] Ekhlasi, A., Motie Nasrabadi, A., Mohammadi, M.: Classification of the children with adhd and healthy children based on the directed phase transfer entropy of eeg signals. Frontiers in Biomedical Technologies **8**(2), 115–122 (2021) https://doi.org/10.18502/fbt.v8i2.6515

[19] Vahid, A., Bluschke, A., Roessner, V., Stober, S., Beste, C.: Deep learning based on event-related eeg differentiates children with adhd from healthy controls. Journal of clinical medicine **8**(7), 1055 (2019)