Ro J Pediatr. 2023;72(2)
DOI: 10.37897/RJP.2023.2.1

Comprehensive review of EEG data classification techniques for ADHD detection using machine learning and deep learning

Nitin Ahire[1,2], Raval N. Awale[3], Abhay Wagh[4]

[1]Doctoral School, Veermata Jijabai Technological Institute, Mumbai, India
[2]Xavier Institute of Engineering, Mumbai, India
[3]Veermata Jijabai Technological Institute, Mumbai, India
[4]Directorate of Technical Education, Mumbai, India

ABSTRACT

Children who have Attention-Deficit/Hyperactivity Disorder (ADHD) have a chronic behavioral disease. Children with ADHD have a hard time focusing and controlling their actions. One of the most difficult problems in controlling and treating this condition is early detection. There is yet to be discovered a reliable professional procedure for early detection of this condition. The electroencephalogram (EEG) is a useful neuroimaging technique for researching ADHD; one of the key goals is to define the EEG of ADHD youngsters. Numerous methods based on EEG signals have been put out in the literature to address this issue since they are an effective neuroimaging approach for studying ADHD. The best recording formats and channels for diagnosing ADHD, however, have not been the subject of many research. Machine learning (ML) and Artificial Intelligence (AI) strategies for identifying ADHD using EEG-based tools are discussed in this paper. Although, in the case of ADHD, the utilization of ML and AI approaches is restricted. However, the data clearly imply that combining EEG technologies with ML/AI may be utilized to detect ADHD. For categorizing adult ADHD subtypes based on EEG power spectra, ML algorithms that incorporate several classifiers are presented. A widely used deep learning (DL) method is the convolutional neural network (CNN). The use of DL approaches in ADHD research, on the other hand, is currently restricted. EEG has been used in studies to look for ADHD neurological connections. Recent advances in deep learning algorithms, particularly CNN, are anticipated to overcome the issue.

Keywords: EEG, ADHD, CNN, Machine learning, Artificial Intelligence, Deep learning

INTRODUCTION

Although it has recently been shown to persist into adulthood, attention deficit hyperactivity disorder (ADHD) is a complex neurobehavioral condition that is often diagnosed in children and adolescents. Inattention, hyperactivity, and impulsivity are symptoms of the disease, according to the DSM-IV (American Psychiatric Association, 1994). 2.5% of adults and 8.4% of children are thought to have ADHD. Despite adopting a different nomenclature, hyperkinetic disorder (HD), the ICD-10 [1] specifies identical criteria for the illness. Childhood prevalence is predicted to be between 5 and 9 percent [2]. More than half of children with ADHD still have clinically significant symptoms as adults, even though symptoms may become better with maturity. This indicates that roughly 5% of adults throughout the globe are impacted. Young children often exhibit many of the traits associated with ADHD, including a high degree of activity, difficulty being still for long periods of time, and short attention spans. Children with ADHD are distinguished by their excessive hyperactivity and inattention relative to their age, which may be distressing and/or make it difficult for them to function at home, school, or with friends.

Corresponding author:
Nitin Ahire
E-mail: nitin.a@xavier.ac.in

Article History:
Received: 19 June 2023
Accepted: 30 June 2023

Scientists have not yet pinpointed the precise causes of ADHD. There is evidence that heredity may play a part in ADHD. Every third kid with ADHD has a family member who also has the condition. ADHD may also be brought on by premature delivery, brain damage, the mother's use of alcohol, tobacco, or high levels of stress while pregnant. It can be difficult to diagnose, particularly in youngsters. No test will detect it. Doctors diagnose ADHD in children and adolescents after in-depth discussions about symptoms with the child, parents, and teachers and after closely observing the child's behaviour. The American Psychiatric Association's guidelines, which are based on the quantity and persistence of a patient's symptoms, are used by doctors. They will also rule out other potential causes of the symptoms, such as health disorders or daily difficulties. To confirm a diagnosis of ADHD or learning problems, a battery of neurological and psychological tests may be administered to a kid. The tests should be administered by a doctor or mental health specialist who has expertise identifying and treating ADHD. You could be referred to a psychiatrist, psychologist, or psychotherapist by your primary care physician.

Researchers have been looking for links between ADHD and electroencephalography (EEG), the oldest tool for systematically evaluating human brain cortex activity, since the early 1990s. An increase in the strength of slow waves, a decrease in the power of fast waves, or both, are the most often reported EEG abnormalities in children with ADHD, which are frequently assessed using the theta-beta ratio (TBR). Given the complexity of the multivariate EEG profile and the diversity of the ADHD population, we suggest that EEG abnormalities in children with ADHD should be customized at high frequency resolution. In other words, rather of looking for abnormalities within certain frequency ranges, EEG anomalies should be detected separately at each frequency point. When applied to ADHD children, data-driven classifiers may extract individual-specific information in addition to identification judgements, therefore deep learning approaches may play a crucial role in attaining this goal. If it can detect specific variances in each person, this kind of diagnostic instrument may be simpler to use.

Most EEG studies on ADHD children focus on indicators like lower alpha and beta bands, as well as higher theta and delta bands, to distinguish ADHD children from healthy control groups [3-10]. The findings of the few EEG investigations including ADHD adults [11-15] are extremely different. This may be because both the developmental aspect of the disease and the quantitative EEG characteristics were examined. Because of this, diagnosing an adult's condition still requires the skills and expertise of the doctor.

The bulk of methods for identifying pertinent discriminators between control and ADHD groups use data acquired from EEG readings using simple statistical techniques like the ANOVA test. In studies of ADHD children and adolescents, these methodologies have produced consistent findings among researchers, but this is not the case in studies of ADHD adults. Deep learning approaches have made enormous strides in computer vision and natural language processing over the past few decades, thanks to faster processors and bigger annotated data sets. For diverse types of data, multiple deep learning architectures have been designed. The most popular deep learning architecture is a multi-layer perceptron (MLP), which can produce results for a variety of inputs. Other common deep learning architectures include convolutional neural networks (CNNs), which are good for image- like data, and recurrent neural networks (RNNs), which are good for sequential data (such as text). The community's interest in applying intelligent algorithms to the challenge of diagnosing ADHD in children has been piqued by the algorithms' capacity for learning.

Convolutional neural networks (CNNs), in particular, function as black boxes in deep learning models, and this makes it difficult to understand the results, which is the main drawback of using these methods in clinical practise. Diverse visualisation approaches, including as saliency maps, activation maps, noise iteration, and picture occlusion, have been presented as a means of illuminating the mystery of CNN models. However, not all ways are effective owing to the inherent discrepancies between pictures and EEG data, and depending on how the EEG is represented, the findings of the visualisation may have a variety of clinical interpretations even if an appropriate practical interpretation is not available. As a result, it's critical to find the best approach to present the data as well as decipher what it signifies dependent on the EEG data type.

BACKGROUND

In this context, machine learning (ML) refers to the use of computer algorithms that can pick up certain tasks from example data without explicit written instructions, i.e., the diagnosis of neurological disorders. To produce the most precise predictions for new data, this area of artificial intelligence uses complex statistical techniques to find predictive or discriminating patterns in training data. Several data mining algorithms have been created as a result of data mining research. These algorithms may be directly applied to a dataset for the purpose of constructing models or deriving crucial conclusions and inferences from that dataset. Among other prominent data mining methods, there are decision trees, Nave Bayes, k-means, and artificial neural networks.

Artificial neural networks, Nave Bayes, k-means, and decision trees are a few popular techniques for data mining. Machine learning algorithms or methods are categorised into five subfields using learning approaches: unsupervised learning, semi-supervised learning, supervised learning, deep learning, and reinforcement learning. Approaches to supervised learning use labelled input data with predetermined outcomes. Methods in this category create a connection between the input and output properties of a labelled dataset. Labelled data facilitates the development of more dependable and accurate models, although the process is computationally intensive. Although the output characteristics are unknown before to the analysis, unsupervised learning techniques try to analyse the data structure in an unlabeled input dataset and provide a mapping between the input and output attributes. Semi-supervised Learning techniques utilised both labelled and unlabelled datasets to create models for intelligence inference. The objective of these strategies is to maximise the rewards from the outcome. This approach of reinforcement learning generates a series of decisions that maximise rewards. Deep Learning focuses on the unification of artificial intelligence and machine learning. It utilises common data to deliver valuable insights. It solves input datasets that contain fewer tagged data.

Machine learning algorithms or techniques are also classified using learning problems as: Classification, Optimization, Clustering, and Regression. Classification is a grouping strategy dependent on the goal value and dataset. It quantifies and classifies the dataset based on the supplied goal value. Clustering is a method that finds intelligence-generating patterns from datasets by identifying intriguing patterns. In regression method, knowledge or data is extracted from prior learning experiences. An equation is constructed that corresponds to the majority of the data points, and data points that do not fit the curve are discarded. Optimization is a technique for enhancing the system's performance in terms of several attributes. In clustering, unlike classification, the goal value is either not supplied as an input or is an unknown parameter.

K Nearest Neighbour

A fundamental supervised learning technique is K-Nearest Neighbor. The most comparable category is chosen based on the assumption that the new instance or data is equivalent to previous examples. The method stores all accessible data and compares it to previously stored data. Using the KNN method, fresh data may be quickly sorted into a suitable category. Problems with classification and regression may both be resolved utilising the KNN approach. It is sometimes referred to as a "lazy learner" algorithm since data from the training set is kept instead of being immediately learnt and then utilised to categorise. The KNN algorithm simply stores the dataset and categorises it into a group that is similar to the incoming data during the training phase. A new data point, x1, and two categories, A and B, are assumed. To solve this issue, a KNN approach is required. We can easily identify a dataset's category or class using KNN.

KNN is a basic algorithm that saves all examples and categorises new cases based on similarity. i) Case- based reasoning (KNN) ii) inductive reasoning iii) case study learning. Since 1970, KNN algorithms have been employed in statistical estimation and pattern identification. It has two types. i) KNN without structure ii) structure-based KNN. Structure-less KNN divides data into training and test samples. The shortest distance between two points is termed the nearest neighbour. These include the orthogonal structure tree (OST), ball tree, k-d tree, suggested a KNN classifier for plant leaf disease detection and classification [16]. The categorization uses textural characteristics taken from leaf disease photos. Alternaria, anthracnose, bacterial blight, leaf spot and canker are among the illnesses that KNN classifier will classify. The suggested method can accurately detect and identify illnesses with 96.76 percent accuracy.

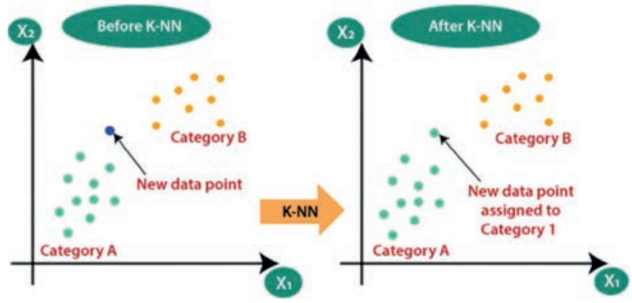

FIGURE 1. K Nearest Neighbor (KNN) algorithm for machine learning

Extreme Gradient Boosting (XGB)

In regression and classification problems, for example, gradient boosting is applied. It generates a group of flimsy prediction models, often decision trees. When a decision tree is the weak learner, gradient-boosted trees perform better than random forests in general. Unlike previous boosting approaches, gradient-boosted trees allow optimization of an arbitrary differentiable loss function. This algorithm creates consecutive decision trees. Weights are vital in XGB. Weights are assigned to each independent variable before being included in the decision tree that forecasts outcomes. The second deci-

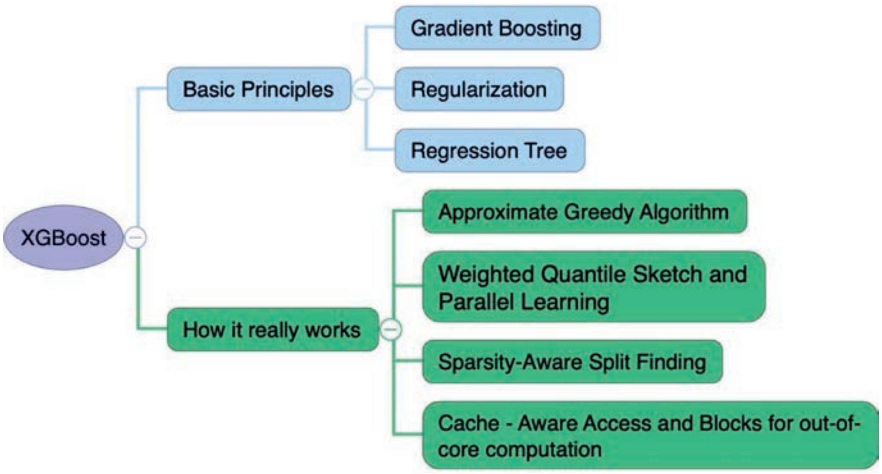

FIGURE 2. Extreme gradient boosting (XGBoost) algorithm for machine learning

sion tree receives higher weight for variables that the first one predicted erroneously. Then, several classifiers and predictors are integrated to create a more accurate model. A user-defined prediction issue can be solved using this tool.

Based on weighted XGB, developed a hierarchical categorization approach. A vast number of characteristics from six groups are recovered from the pre-processed heartbeats [17]. Then, to choose features, recursive feature elimination is carried out. The classification process is then followed by the development of a hierarchical classifier. Threshold and boost classifiers make up the hierarchical classifier. And weights are added to the XGBoost classifiers to improve them. The results demonstrated that the XGB has the highest precision (> 0.94) and greatest prediction accuracy (94 percent). These findings indicate that the XGB is a strong candidate for patient classification. These results suggest a potential new treatment option for individuals with acute bronchiolitis: XGB systems trained on clinical data.

Decision Tree

A decision tree is a decision support tool that uses a tree-like graph or model to represent actions and their possible effects, including utility and the results of chance occurrences. It's one way of showing an algorithm. Operations research regularly uses decision trees, especially in decision analysis, to help determine a course of action that is most likely to succeed. The usage of it in machine learning is also widespread. By mapping from the root node to each child node separately, a decision tree may easily be transformed into a set of rules. Finally, by adhering to these criteria, one can draw suitable judgments.

Random Forest [4] is a popular approach for supervised learning. Used in Machine Learning Classification and Regression. It makes use of ensemble learning, a technique that combines a number of classifiers to tackle complicated issues and enhance model performance. In order to increase the precision of a dataset's predictions, a Random Forest makes use of a number of decision trees on different subsets of the provided dataset. The random forest

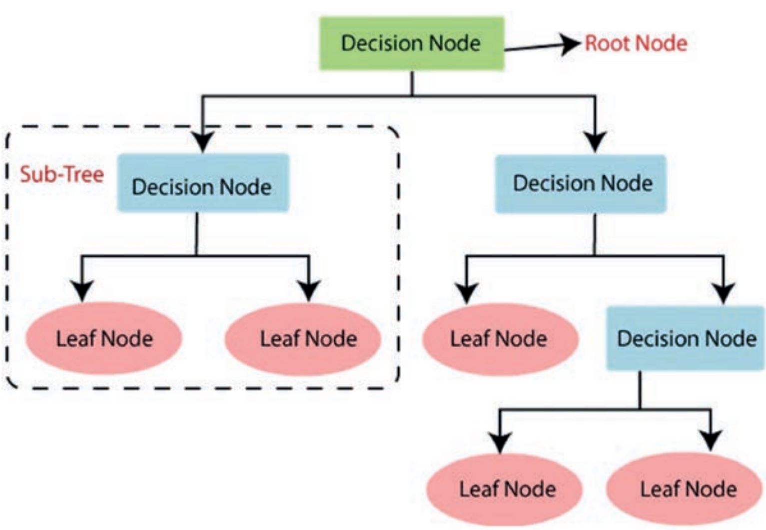

FIGURE 3. Decision tree algorithm for machine learning Random Forest classifier

considers the forecasts from each tree and predicts the final result depending on the majority vote rather than relying on a single decision tree. More trees in the forest enhance precision and reduce overfitting. There are several scientific applications of random forest [10,18,19,20,21].

To predict the class of the dataset, the random forest makes use of many trees. While some decision trees may be accurate, others might not. But as a group, the trees make wise predictions. So, for a better random forest classifier, here are two underlying presumptions. For the classifier to accurately predict results, the feature variable in the dataset has to have some genuine values. The predictions from each tree must be substantially connected. RF is a grouping or ensemble of Classification and Regression Trees (CART) that have been trained using datasets called bootstraps that are the same size as the training set. A collection of bootstraps is used as a test set after a tree has been constructed. The classification error rate across all test sets serves as the OOB estimate of generalization error. It was demonstrated that the OOB error for bagged classifiers is the same as using a test set of the same size as the training set. Using the OOB estimate eliminates the requirement for a test set [4].

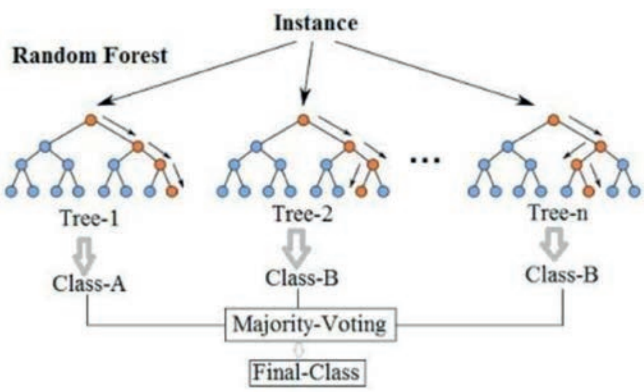

FIGURE 4. Random forest classifier

Logistic regression

One of the most often used ML methods, supervised learning includes logistic regression. Using a given collection of independent factors, it is employed to predict the categorical dependent variable. A dependent variable with a categorical output is anticipated. Thus, the result must be a discrete or categorical value. It may be Yes or No, 0 or 1, true or False, etc., although probabilistic values between 0 and 1 are presented. Logistic Regression and Linear Regression are quite similar, with the exception of their distinct uses. Problems involving regression are solved using linear regression, whereas those requiring classification are solved using logistic regression.

In order to estimate the prediction error more precisely [22] present a parametric bootstrap model based on significant research in differentially expressed genes, particularly the local false discovery rate. The suggested technique guides model selection on two crucial issues: the number of genes to include in the model and the appropriate penalised logistic regression shrinkage. They show that picking more than 20 genes reduces prediction error just little. With Golub's leukaemia and our cervical cancer data, we get extremely accurate predictions. As assessed the performance of classification algorithms to predict coronary artery disease (CAD) [23]. They contrasted self-organizing feature maps, LR, CART, MLP, and SOFM. The predictor variables were age, sex, family history of CAD, smoking status, diabetes, systemic hypertension, hypercholesterolemia, and BMI. The ROC curve, HCA, and Multidimensional Scaling were used to compare classification performance (MDS). There are 0.783 AUROC curves for MLP and 0.753 AUROC curves for CART, RBF and SOFM. In this data set, MLP performed the best in terms of classifying CAD. So FM did not perform well in predicting CAD in HCA and MDS.

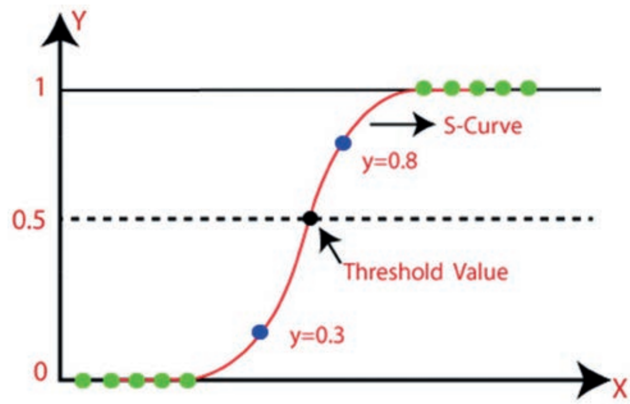

FIGURE 5. Logistic regression in machine learning

Support Vector Machine

It is a supervised learning method that uses a hyperplane to split a dataset into two groups. SVM algorithm serves a similar function to C4.5, excluding the need of Decision trees. In order to reduce the likelihood of misclassification, the SVM algorithm tries to raise the margin (the distance between the hyper plane and the two closest data points from each corresponding class). Popular svm - based implementations include scikit-learn, MATLAB, and LIBSVM.

Literature Survey

Machine learning methods have been used in a few studies to differentiate ADHD patients from control groups. There have been studies that have

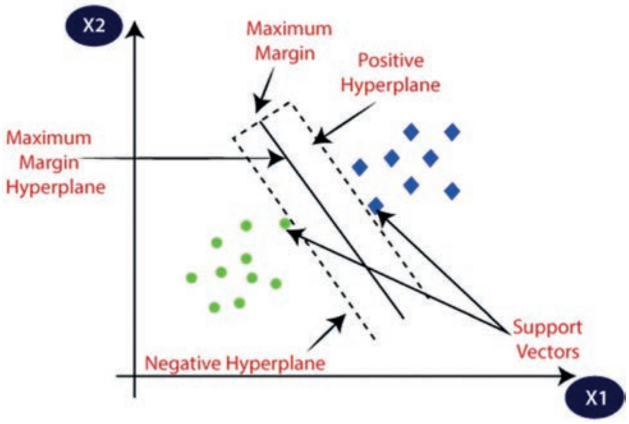

FIGURE 6. Support Vector Machine (SVM) algorithm

used linear classifiers with reasonable accuracy to distinguish ADHD from control groups [3,7,24-27]. Data non-linear correlations may be found via non-linear classifiers like support vector machines and artificial neural networks. According to [28] a support vector machine classifier might be used to distinguish between ADHD patients and control groups using EEG scans of event- related potentials.

As recorded EEG data from 31 children with ADHD and 30 healthy youngsters. In this study, a convolutional neural network was used to create a deep learning model with substantial performance in image processing applications [29]. Next, pre-processed EEG data to reduce noise and artefacts for this purpose. Then, subdivide the pre-processed samples into additional samples. By removing the theta, alpha, beta, and gamma frequency bands from each sample segment, they created a three-channel RGB colour picture. In order to extract features and identify the photos, a 13-layer convolutional neural network was fed the data. Using 5-fold cross validation on train, evaluation, and test data, the proposed model was examined. For segmented samples, the average accuracy was 99.06 percent, 97.81 percent, and 97.47 percent, respectively. Subject-based test samples had an average accuracy of 98.48 percent.

Using EEG-based brain networks developed a generic DL framework for diagnosing children with ADHD and illustrated the framework's applicability and outlined some essential considerations for its implementation [8]. The suggested framework demonstrated impressive performance with validation data accuracy of 98.17 percent and test data accuracy of 94.67 percent. In general, CNN techniques to diagnosing ADHD seem promising, and the current framework may be changed to support additional medical applications and/or brain signal recordings.

The effects of photic stimuli at various frequencies and channels on the diagnosis of ADHD were examined [30]. With a primary objective to determine the best effective channel and recording status for ADHD diagnosis. Using power spectrum densities and spectral entropy values, the dataset utilised in these studies was compiled. These numbers were acquired from both ADHD and non-ADHD patients. The LSTM classifier had the greatest accuracy when these data were put into support vector machine (SVM), long short-term memory (LSTM), and artificial neural network classifiers. The computed accuracy of LSTM on the "Fp1,F7" channel was 88.88 percent, and in the eyes-closed resting state, it was 92.15 percent. It was discovered that spectral entropy contributed favourably to the accuracy. Consequently, the potential difference among "Fp1, F7" sensors in the eyes-closed, resting condition proved useful for detecting ADHD.

The work was the first effort to categorise people with ADHD using a support vector machine, indicating that classification using non-linear approaches is viable for clinical populations. A visual GO/NOGO test with two stimuli was performed by 74 ADHD sufferers and 74 controls, two sets of age- and gender-matched participants. ICA was used to separate ERP responses into distinct components [31]. A collection of independent component features was defined by a feature selection algorithm and put into a SVM model. Five latency measurements in certain time frames, taken from four different, independent components, made up the feature set. There were also two executive function-related components, a sensory-related component, and an independent novelty component. The accuracy of categorisation using a 10-fold cross-validation approach was 92 percent. In addition, it has been demonstrated that separate ERP components give characteristics that may be utilised to characterise clinical groups.

The powerful algorithms of machine learning to deduct the ADHD in adults. The sample studied contains 117 adults (67 ADHD, 50 controls). Two resting conditions (with eyes open and closed) and two neuropsychological tasks total four measurements (visual continuous performance test and emotional continuous performance test). From the sample, they produce four data sets, one for each condition. Each data set is used to train four different SVM classifiers, and the results are combined using a logical expression built from a Karnaugh map. The findings demonstrate that this method enhances differentiation across ADHD and control conditions, including between ADHD subgroups. Furthermore, on the basis of EEG power spectra acquired under different measurement settings, they present a model for classifying people with ADHD and control groups. The EEG power spectrum is derived from scalp electrode-recorded EEG data and displays the distribution of the signal's squared amplitude over all of its frequency ranges. Neuronal activity in the brain results in the production of EEG signals [32].

According to [11] ADHD is difficult to diagnose since many of its symptoms are like those of other diseases. It is connected with gambling problem and obesity, with around 20% of overlap between each diagnosis. Using proven diagnostic equipment, it is essential for clinical practise to differentiate between illnesses presenting similar symptoms. Using all 26 items of the Conner's Adult ADHD Rating Scales, the authors were able to differentiate between subjects with ADHD, obesity, and the control group with an overall accuracy of 0.80; precision (positive predictive value) ranged from 0.78 (gambling) to 0.92 (obesity); and recall (sensitivity) ranged from 0.58 (obesity) to 0.87 (control group) (ADHD). The models with the top 5 and best 10 components had fewer excellent fits. The CAARSS appears to be a viable tool for use in clinical practise for multi-classifying illnesses with ADHD-like symptoms. The study's flaws were also pointed up by the author. To begin with, they were unable to recruit patient groups other than those with obesity and compulsive gambling to participate in this research. Second, only the CAARS short self-rating version was employed. In order to study the diagnostic features of various populations, future research should strive to include people with a diversity of disorders and diagnostic methods.

For the automated identification of ADHD, [33] offer a robust machine learning approach that uses pupil-size dynamics as an objective biomarker. In our method, pupil metric feature engineering and visualisation were combined with cutting-edge binary classification algorithms and univariate feature selection. Ten-fold nested cross-validation (CV) yielded 85.6 percent AUROC, 77.3 percent sensitivity, and 75.3% specificity for the support vector machine classifier on declassified datasets of 50 patients. Pupil-size dilation velocity was discovered to be one of 218 statistically significant differentiators (p 0.05) among 783 engineered features. These other novel behavioural insights into associations between pupil-size dynamics and the presence of ADHD included Fourier transform metrics, absolute energy, consecutive quantile changes, approximate entropy, aggregated linear trends, and other. The strong AUROC values demonstrate the binary classifiers' robustness in identifying ADHD despite the limited sample size; hence, the sensitivity and specificity metrics may be significantly enhanced with additional data. For the first time, this work uses machine learning approaches to diagnose Through the use of oculometric paradigms and machine learning, Pupilometrics emphasises its potential as a discriminative biomarker and paves the way for novel diagnostic applications to assist in the diagnosis of ADHD.

This machine learning-based methodology offers a reliable, time-efficient way to diagnose ADHD using an objective biomarker rather than subjective clinical judgments. Qualitative observations are used in examinations lasting many hours. Our findings help hasten clinical diagnosis and provide machine learning researchers a fresh perspective on pupillometry and ADHD. They can help doctors correctly diagnose ADHD. We demonstrate that pupillometrics may be utilised to identify between ADHD positive and negative groups using machine learning. Future research to improve model performance and robustness may include the integration of pupillometrics and eye gaze directions to provide a more precise ADHD risk score. When additional data is made accessible, deep learning-based algorithms may allow for a more scalable and potent study of oculometric data for ADHD identification. A multi-classification approach that includes On-ADHD, Off-ADHD, and Ctrl individuals might provide a more comprehensive framework.

The used CNN and Gradient-weighted Class Activation Mapping (Grad-CAM) to detect a customised spatial-frequency imbalance in the EEGs of children with ADHD. A total of 57 age- and handedness-matched control children and 50 ADHD-afflicted children (9 females, mean age: 10.44 0.75 years) were enrolled [9]. As input, the power spectrum density of EEGs was utilised. They proposed an understandable representation of multichannel EEG data that can be trained using CNN algorithms. With their anomalies removed, the relative power distributions in various frequency bands were compared to in children with ADHD. The study proved that it is possible to identify ADHD using CNN approaches with an accuracy of 90.290.58%. There were significant disparities in the individualised spatial-frequency abnormalities of ADHD patients. The aberrations were consistent with both group- and individual-level power distributions. A genetic and positron emission tomography (PET) imaging classification model for ADHD and healthy controls is proposed (HC) by [34]. Serotonin transporter (SERT') binding potential was measured with [11C] DASB in 16 ADHD patients and 22 healthy controls using PET. Thirty SNPs in the HTR1A, HTR1B, HTR2A, and TPH2 genes were genotyped in all subjects. Cortical and subcortical regions of interest (ROI) were found using a ten-fold cross-validation model, and feature selection and classification were carried out using random forest (RF) machine learning. The ROIs of the posterior cingulate gyrus, cuneus, precuneus, pre-, para-, and postcentral gyri, as well as the SNPs HTR2A rs1328684 and rs6311, and HTR1B rs130058, were shown to be the most effective in differentiating between the presence of ADHD and the presence of HC. The validation sets' average accuracy across all iterations was 0.82 (0.09), while their respective sensitivity and specificity were

0.75 and 0.86. The data supporting the suggested model demonstrates the significance of the SERT gene, as well as the HTR1B and HTR2A genes, in ADHD and suggests disease-specific consequences. A trustworthy computer- assisted diagnostic tool for disorders that originate in the serotonergic system would help clinicians make better judgments due to the large number of co-occurring ailments and the difficulty of discriminating between them, particularly in ADHD.

The study of [35] involved 83 adolescents with ADHD. Parents completed the ADHD Rating Scale- IV and the Disruptive Behaviour Disorder rating scales at baseline, and participants underwent the continuous performance test, the Stroop colour word test, and resting functional MRI scans. Additionally, the amounts of cotinine and lead in the urine and blood were assessed. The subjects took part in a methylphenidate research with an open label that lasted for eight weeks. For data analysis, four distinct machine learning methods were employed. The classification accuracy of support vector machines was 84.6 percent (area under the receiver operating characteristic curve: 0.84). Age, weight, ADRA2A MspI and DraI polymorphisms, lead level, Stroop colour word test performance, and oppositional symptoms were shown to be the most distinctive sets of characteristics on the DBD grading scale. Their results support the translational development of SVM as an informative tool for predicting treatment response in ADHD, however additional work is needed to improve classification performance.

In a prior work of [36] discovered that four machine-learning algorithms could accurately (area under the curve (AUC) >0.96) differentiate ASD from ADHD using only a limited fraction of Social Responsiveness Scale items (SRS). They use a freshly collected crowdsourced data set that includes answers to their top 15 SRS-derived questions from parents of children with ASD (n = 248) or ADHD (n = 174) in order to improve their model's ability to generalise to new, "real-world" data. By combining this unique survey data with their original archive sample (n=3417) and doing repeated cross-validation with subsampling, they were able to construct a classification system with an AUC of 0.890.01 that uses just 15 questions.

As trained (a) a cross-sectional random forest (RF) model using data accessible at age 17 to predict SUD diagnosis between years 18 and 19; and (b) a longitudinal recurrent neural network (RNN) model with the Long Short-Term Memory (LSTM) architecture to predict new diagnoses at each age. The area under the receiver operating characteristic curve (AUC) for the random forest model (RF) was 0.73 (95 percent confidence interval [CI]: 0.70-0.76). The RF model nevertheless yielded significant AUCs in the 18-19 age range whether predicting all SUD diagnoses (0.69, 95 percent CI 0.66-0.72) or new diagnoses after excluding past diagnoses from the covariates (0.67, 95 percent CI 0.64, 0.71). The model that predicted new diagnoses had excellent model calibration, with a low Brier score of 0.086. 10 years before the first diagnosis, the Longitudinal LSTM model was able to forecast future SUD risks as early as 2 years of age. For longitudinal models that forecasted new diagnoses one, two, five, and ten years in the future, AUC was 0.63 [37].

The study of [38] used machine learning to analyse parent/teacher evaluations, behavioural and neurological measures of executive function (EF) in predicting ADHD in 162 young children (ages 4–7, mean age 5.55, 82.6% Hispanic/Latino). Evaluations of EF teachers were the best indicators of ADHD. The current study discovered that cortical anatomy metrics from research studies and cognitive measures of EF that are frequently found in repeated evaluations don't add much to the ability to tell chil-

TABLE 1. Comparison of the model accuracy with some state-of-the-art studies in this field

Study	Year	Dataset	Methods & Feature Extraction	Classifier	Accuracy
Allahverdy et al. [2]	2016	Same as this study	Exponent of Lyapunov Fractal dimension of Katz Dimension of the Higuchi fractal Fractal dimension of Sevcik	MLP NN	96.7%
Mohammadi et al. [2]	2016	Same as this study	Estimated entropy The Most Notable Petrosian Fractal Dimension of Lyapunov Dimension of the Higuchi Fractal	MLP NN	93.65%
Chen et al. [8]	2019	51 healthy youngsters and 50 children with ADHD	Mutual understanding a matrix of connections	Deep CNN	94.67%
Chen et al. [9]	2019	50 kids with ADHD 57 wholesome kids	Gradient-weighted Class Activation Mapping: Common Spatial Patterns	Deep CNN	90.29%
Dubreuil-Vall et al. [39]	2019	20 adult ADHD sufferers; 20 healthy controls	Wavelet-based ERP spectrograms of event-related potentials (ERP) during the Flanker Task	Deep CNN	88%
Moghaddari et al. [29]	2020	30 healthy youngsters and 31 children with ADHD	separation of the frequency bands creates RGB images	Deep CNN	98.48%

dren with ADHD from those who are developing normally. This is despite the possibility that a more thorough examination of neural metrics, such as diffusion-weighted imaging, may shed more light on the underlying cognitive deficits linked to ADHD. This impact may be amplified by several EF questions in the BRIEF, as well as ADHD symptoms. Future study analysing the usefulness of such indicators in predicting academic and social impairment might provide light on their involvement in ADHD.

CONCLUSION AND FUTURE DIRECTIONS

This review examined a variety of ADHD diagnostic methods, including trials in which ML and DL AI techniques were employed to make the diagnosis. The bulk of research focused on hospital-based modalities such as MRI and EEG, whereas relatively few studies reported on the remaining modalities. Using data mining techniques, the healthcare business may successfully "mine" important information from the vast volumes of data it generates. Rather of using a single mining strategy to a data set, these papers demonstrate that a variety of mining techniques yields considerably superior results. The rapid and efficient creation of a system for the treatment of ADHD disease is the result of the careful selection of a combination of data mining techniques and their precise application to the data set. The needed dataset is separated into two halves, one of which is utilised for mining and the other for confirming. Frequently, the 10-fold cross validation procedure is utilised. Some of the studies use a dataset to examine various categorization algorithms in order to identify whether a patient has a likelihood of developing ADHD. Others have investigated the aetiology of ADHD diseases by "mining" a specific dataset. Fuzzy logic, association rule mining, Nave Bayes, artificial neural networks, and decision trees are some classification techniques. In addition to analysing these regularly employed strategies, several recent publications have examined "hybrid models." The purpose of a hybrid model is to improve outcomes by combining various well-known classification and selection strategies into a single model. It has been noted that hybrid models provide extremely high precision if the appropriate combinations of multiple algorithms are used.

The use of AI as a therapeutic decision support tool for ADHD might be further researched via a variety of additional research avenues. Last but not least, we wish to support the creation of a cloud-based platform that can save data from all ADHD diagnostic instruments in a single location. As a consequence, medical professionals would have immediate access to all the information required to confirm a diagnosis. For instance, parents, teachers, or ADHD patients may complete surveys independently, with the data being analysed by AI models in a cloud-based system for psychiatrists or doctors. So, to aid in the design and monitoring of medicines, researchers may be able to apply AI as a component of a diagnostic procedure as well as a precision medicine clinical decision support system.

New efforts should ensure that ML & DL algorithms are explicable. AI approaches such as ML and DL might be difficult to comprehend. Due to the complex process utilised to produce the result, DL models are referred to as "black boxes," and their outputs are difficult for doctors to comprehend. AI algorithms have been hampered in their application as clinical decision support aids in healthcare due to their limited interpretability. As a result, future research on DL models should focus on the model's explainability. Several methods, such as LIME, SHAP, and integrated gradients, may help ML or DL models be more interpretable. We intend to aid in the creation of an accurate AI model for identifying and tracking ADHD. This model will be used in a cloud-based system. In healthcare systems, wearable sensors are becoming increasingly prevalent. Collecting and analysing a stream of data from wearable computing devices in real time in order to make chronic illness forecasts.

Conflict of interest: none declared
Financial support: none declared

REFERENCES

1. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. 1993;2. Available: https://apps.who.int/iris/handle/10665/37108
2. Polanczyk G, De Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am J Psychiatry.* 2007;164(6):942-8. doi: 10.1176/ajp.2007.164.6.942
3. Barry RJ, Clarke AR, Johnstone SJ. A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography. *Clin Neurophysiol.* 2003 Feb;114(2):171-83. doi: 10.1016/s1388-2457(02)00362-0
4. Clarke AR, Barry RJ, McCarthy R, Selikowitz M. EEG analysis in Attention-Deficit/Hyperactivity Disorder: a comparative study of two subtypes. *Psychiatry Res.* 1998 Oct 19;81(1):19-29. doi: 10.1016/s0165-1781(98)00072-9
5. Clarke AR, Barry RJ, McCarthy R, Selikowitz M. Electroencephalogram differences in two subtypes of attention-deficit/hyperactivity disorder. *Psychophysiology.* 2001 Mar;38(2):212-21. PMID: 11347867.
6. Clarke AR, Barry RJ, McCarthy R, Selikowitz M. Age and sex effects in the EEG: development of the normal child. *Clin Neurophysiol.* 2001 May;112(5):806-14. doi: 10.1016/s1388-2457(01)00488-6
7. Clarke AR, Barry RJ, McCarthy R, Selikowitz M. EEG analysis of children with attention-deficit/hyperactivity disorder and comorbid reading disabilities. *J Learn Disabil.* 2002 May-Jun;35(3):276-85. doi: 10.1177/002221940203500309

8. Lubar JF. Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders. *Biofeedback Self Regul.* 1991 Sep;16(3):201-25. doi: 10.1007/BF01000016
9. Monastra VJ, Lubar JF, Linden M, VanDeusen P, Green G, Wing W, Phillips A, Fenger TN. Assessing attention deficit hyperactivity disorder via quantitative electroencephalography: an initial validation study. *Neuropsychology.* 1999 Jul;13(3):424-433. doi: 10.1037/0894-4105.13.3.424
10. Pop-Jordanova N, Markovska-Simoska S, Zorcec T. Neurofeedback treatment of children with attention deficit hyperactivity disorder. *Prilozi.* 2005 Aug;26(1):71-80. PMID: 16118616.
11. Bresnahan SM, & Barry RJ. Specificity of quantitative EEG analysis in adults with attention deficit hyperactivity disorder. *Psychiatry research.* 2002;112(2):133-44. doi: 10.1016/s0165-1781(02)00190-7
12. Reichenberg A, Gross R, Weiser M, Bresnahan M, Silverman J, Harlap S, Susser E. Advancing paternal age and autism. *Arch Gen Psychiatry.* 2006;63(9):1026-32. doi: 10.1001/archpsyc.63.9.1026
13. Clarke B. Normal bone anatomy and physiology. *Clin J Am Soc Nephrol.* 2008;3(Suppl 3):S131-9. doi: 10.2215/CJN.04151206
14. Koehler M, Punya M. What is technological pedagogical content knowledge (TPACK)? Contemporary issues in technology and teacher education 9, no. 1 (2009): 60-70. doi: 101177/002205741319300303
15. Markovska-Simoska S, Pop-Jordanova N. Face and emotion recognition by ADHD and normal adults. *Acta Neuropsychologica.* 2010;8(2):99-122.
16. Hossain E, Hossain MF, Rahaman MA. A color and texture-based approach for the detection and classification of plant leaf disease using KNN classifier. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) 2019; pp. 1-6. *IEEE.* doi: 10.1109/ECACE.2019.8679247
17. Shi H, Wang H, Huang Y, Zhao L, Qin C, Liu C. A hierarchical method based on weighted extreme gradient boosting in ECG heartbeat classification. *Comput Methods Programs Biomed.* 2019 Apr;171:1-10. doi: 10.1016/j.cmpb.2019.02.005
18. Guo L, Chehata N, Mallet C, Boukir S. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing.* 2011;66(1);56-66. doi: 10.1016/j.isprsjprs.2010.08.007
19. Gislason PO, Benediktsson JA, Sveinsson JR. Random forests for land cover classification. *Pattern Recognition Letters.* 2006;27(4):294-300. doi: 10.1016/j.patrec.2005.08.011
20. Ghimire B, Rogan J, Miller J. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters.* 2010;1(1):45-54. doi: 10.1080/01431160903252327
21. Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. Random forests for classification in ecology. *Ecology.* 2007;88(11):2783-92. doi: 10.1890/07-0539.1
22. Liao JG, Chin KV. Logistic regression for disease classification using microarray data: model selection in a large p and small n case. *Bioinformatics.* 2007;23(15):1945-51. doi: 10.1093/bioinformatics/btm287
23. Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Systems with Applications.* 2008;34(1):366-74. doi: 10.1016/j.eswa.2006.09.004
24. Wender PH. Attention-Deficit Hyperactivity Disorder in Adults. *Psychiatr Clin North Am.* 1998;21(4):761-74. doi: 10.1016/s0193-953x(05)70039-3
25. Robaey P, Breton F, Dugas M, Renault B. An event-related potential study of controlled and automatic processes in 6-8 year-old boys with attention deficit hyperactivity disorder. *Electroencephalogr Clin Neurophysiol*. 1992;82(5):30-340. doi: 10.1016/0013-4694(92)90003-z
26. Satterfield JH, Braley BW. Evoked potentials and brain maturation in hyperactive and normal children. *Electroencephalogr Clin Neurophysiol*. 1977;43(1):43-51. PMID: 68871. doi: 10.1016/0013-4694(77)90193-6
27. Smith JL, Johnstone SJ, Barry RJ. Aiding diagnosis of attention-deficit/hyperactivity disorder and its subtypes: discriminant function analysis of event-related potential data. *J Child Psychol Psychiatry.* 2003;44(7):1067-75. PMID: 14531589. doi: 10.1111/1469-7610.00191
28. Mueller A, Candrian G, Grane VA, Kropotov JD, Ponomarev VA, Baschera GM. Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: a validation study. *Nonlinear Biomed Phys.* 2011;5(1):1-18. doi: 10.1186/1753-4631-5-5
29. Moghaddari M, Lighvan MZ, Danishvar S. Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG. *Comput Methods Programs Biomed.* 2020;197: 105738. PMID: 32927404. doi: 10.1016/j.cmpb.2020.105738
30. Tosun M. Effects of spectral features of EEG signals recorded with different channels and recording statuses on ADHD classification with deep learning. *Phys Eng Sci Med.* 2021;44(3):693-702. PMID: 34043150. doi: 10.1007/s13246-021-01018-x
31. Mueller A, Candrian G, Grane VA, Kropotov JD, Ponomarev VA, Baschera GM. Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: a validation study. *Nonlinear Biomed Phys.* 2011;5(1):1-18. doi: 10.1186/1753-4631-5-5
32. Tenev A, Markovska-Simoska S, Kocarev L, Pop-Jordanov J, Müller A, Candrian G. Machine learning approach for classification of ADHD adults. Int J Psychophysiol. 2014;93(1):162-6. PMID: 23361114. doi: 10.1016/j.ijpsycho.2013.01.008
33. Das W, Khanna S. A Robust Machine Learning Based Framework for the Automated Detection of ADHD Using Pupillometric Biomarkers and Time Series Analysis. *Sci Rep.* 2021;11(1):1-12. doi: 10.1038/s41598-021-95673-5
34. Kautzky A, Vanicek T, Philippe C, Kranz GS, Wadsak W, Mitterhauser M, Lanzenberger R. Machine learning classification of ADHD and HC by multimodal serotonergic data. *Transl Psychiatry.* 2020;10(1):1-9. doi: 10.1038/s41398-020-0781-2
35. Kim JW, Sharma V, Ryan ND. Predicting methylphenidate response in ADHD using machine learning approaches. *Int J Neuro Psychopharmacol.* 2015;18(11). PMID: 25964505. PMCID: PMC4756719. doi: 10.1093/ijnp/pyv052
36. Duda M, Ma R, Haber N, Wall DP. Use of machine learning for behavioral distinction of autism and ADHD. *Transl Psychiatry.* 2016;6(2):e732. PMID: 26859815. PMCID: PMC4872425. doi: 10.1038/tp.2015.221
37. Zhang-James Y, Chen Q, Kuja-Halkola R, Lichtenstein P, Larsson H, Faraone SV. Machine-Learning prediction of comorbid substance use disorders in ADHD youth using Swedish registry data. *J Child Psychol Psychiatry.* 2020;61(12):1370-9. doi: 0.1111/jcpp.13226
38. Öztekin I, Finlayson MA, Graziano PA, Dick AS. Is there any incremental benefit to conducting neuroimaging and neurocognitive assessments in the diagnosis of ADHD in young children? A machine learning investigation. *Dev Cogn Neurosci.* 2021;49:100966. doi: 10.1016/j.dcn.2021.100966
39. Dubreuil-Vall L, Ruffini G, Camprodon JA. A deep learning approach with event-related spectral EEG data in attentional deficit hyperactivity disorder. *MedRxiv.* 2019;19005611. doi: 10.1101/19005611