Systematic Review of The Effectiveness of Machine Learning in Discriminating ADHD From ASD

Yuansi Wu *

Psychology and Brain Science, University of California, Santa Barbara, CA, United States, 93117

* Corresponding Author Email: wonsara0109@gmail.com

Abstract. Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that significantly impacts psychological, physical, and social functioning, especially in children. Traditional diagnostic methods—such as interviews, questionnaires, and Continuous Performance Tests (CPTs)—struggle to accurately distinguish ADHD from similar conditions, affecting treatment outcomes and quality of life. Recently, advancements in artificial intelligence (AI), particularly machine learning (ML), have been applied to improve ADHD classification. This review assesses the design and effectiveness of ML models in six studies for classifying ADHD and Autism Spectrum Disorder (ASD), highlighting promising approaches to enhance reliability and accuracy in these tasks.

Keywords: ADHD; ASD; machine learning; diagnosis; pediatric.

1. Introduction

Attention-Deficit/Hyperactivity Disorder (ADHD) is a pervasive neurodevelopmental disorder that typically emerges in childhood and can persist into adulthood. Recent research has identified ADHD as an impairment in the brain's executive functions, shifting the focus from behavioral symptoms to underlying cognitive dysfunctions [1-2]. Studies indicate a steady increase in ADHD diagnoses since 1997 [3-4]. An analysis of the National Health Interview Survey reported that the prevalence of ADHD among children rose from 6.1% in 1997–1998 to 10.2% in 2015–2016 [3]. A 2024 study of U.S. data estimated that approximately 11.4% of children have received an ADHD diagnosis, with around 6.5 million children living with the condition [5]. Although specific 2024 data is still emerging, it is likely that this upward trend will continue.

The traditional diagnosis approach involves a comprehensive analysis of interview, questionnaires, behavioral checklists, and neuroimaging data by psychiatric, neurologists [6], and family practitioners [7]. In recent year, more researchers start to incorporate machine learning into the medical diagnosis process to unveil relationships between behavioral symptoms and changes in brain activities and improve cost and time traditional method takes to make a diagnosis [8, 6]. The present study aims to use systematic review to explore how machine learning uses complex neuron networking to enhance the discrimination of ADHD from ASD based on physiological and behavioral datasets.

2. Literature Review

2.1. Autism Spectrum Disorder (ASD)

ASD is a neurodevelopmental disability characterized by impairments in social interaction, stereotyped behaviors, and communication [9]. The symptoms of ASD manifest in both verbal and nonverbal forms. For instance, deficits in social communication and interaction may be evident in an inability to engage in essential nonverbal behaviors, such as maintaining eye contact. Additionally, restricted and repetitive behavioral patterns often present as repetitive language use and highly focused interests, as outlined in the DSM-V [10].

Individuals with ASD often experience impairments in executive functioning, leading to challenges in attention regulation. Specifically, preschool-aged children with ASD may show disengagement from certain activities while demonstrating intense focus on others, often accompanied by either hypersensitivity or hyposensitivity to sensory stimuli. Current diagnostic methods for ASD involve a variety of assessments, including clinical interviews, rating scales, family-centered care, and brain imaging [11]. Though this is a scientific-based process and have been repeatedly experimented by researchers, it is still difficult to accurately discriminate ASD patients from neurotypical groups and disorders with similar comorbidities as reports from psychiatric are often too subjective and simple, which might cause many ASD -risk groups go under diagnosed, leading to more severe impairment in workplace and reduce life quality as adults.

2.2. Attention-Deficit/Hyperactivity Disorder (ADHD)

ADHD, characterized by hyperactivity, inattention, and impulsivity [9], is a neurodevelopmental disorder impacting a child's focus, ability to filter distractions, and behavior regulation in social settings. The DSM-5 expands the ADHD diagnostic criteria to include adults, specifies symptom severity, and provides a more comprehensive framework compared to DSM-IV [12], which must persist for at least six months and cause social, academic, or occupational impairment, include inattention (e.g., difficulty focusing, forgetfulness) and hyperactivity-impulsivity (e.g., fidgeting, excessive talking). Present before age 12, symptoms are categorized into Combined, Predominantly Inattentive, and Predominantly Hyperactive-Impulsive presentations, with severity levels of mild, moderate, or severe. While ADHD and ASD assessments share similar methods, ADHD evaluations focus on how symptoms impact daily life. If ADHD persists into adulthood, individuals may face challenges in employment and education, affecting family life and increasing poverty risk [13].

2.3. ADHD and ASD Comorbidity

ASD is one of the most common comorbid conditions in children diagnosed with ADHD. Approximately 20-50% of children with ADHD meet the diagnostic criteria for ASD, and both disorders frequently present in males [14]. Both ADHD and ASD are neurodevelopmental disorders that impair executive function and social interaction. Studies have shown individuals with both ADHD and ASD experience greater social and cognitive problems and more adverse effect on daily life [15-16].

While there are currently no direct assessment tools to definitively differentiate ADHD from ASD, researchers have found key behavioral, neurological, and observational differences between the two disorders. These distinctions provide valuable information for clinical pre-screening and classification efforts.

2.4. Differentiating Features

2.4.1 Behavioral differences

Emotional recognition experiments are widely used to assess cognitive abilities in children with neurodevelopmental disorders, often employing eye-tracking devices to gather cognitive data. Eye-tracking fixation data provides insights into the nervous system and cognitive behaviors in individuals with cognitive impairments. For example, studies have found associations between pupil size, cognitive ability, attention, and stress levels [17]. Pupil diameter, in particular, has been identified as a key factor distinguishing ASD from ADHD, as it reflects differences in attention levels and cognitive load during tasks [18]. ASD participants generally exhibit larger pupil diameters, suggesting sustained attention and greater cognitive effort, while ADHD participants show more fluctuation in pupil diameter with task difficulty, indicating variable cognitive effort and attention levels. The consistent pupil variation in ASD may relate to their narrowly focused attention or less reactive autonomic nervous system functioning [18].

2.4.2 Self-report Questionnaire

Machine learning (ML) can greatly reduce the cost and time involved in clinical diagnoses, with researchers leveraging online reports combined with ML to classify ADHD and ASD. Duda et al. were the first to use the Social Responsiveness Scale (SRS) as an ML input for distinguishing ADHD from ASD. The SRS, a standardized questionnaire of 65 questions completed by caregivers, teachers, or parents, enabled researchers to select five items with the highest discriminative power, achieving over 96% accuracy using Enet and LDA models [19]. In 2023, a European study inspired by Duda's work compiled online reports from parents and teachers using CPRS-R, CBCL, SRS, and CTRS-R to evaluate ADHD and ASD in children and adolescents, examining their consistency with on-site clinical diagnoses. This study found that caregiver reports on ADHD symptoms provided the highest discriminative power in clinical diagnostic processes [20].

2.4.3 Physiological differences

Reactivity in individual's physical body call deeper insight into the difference between ADHD and ASD [21]. Electroretinogram (ERG) is widely used in physical measure by producing ERG waveform, the cumulative activities generated from retinal cells, enhancing understanding of disorders from neurochemical perspective. A spectral analysis of ERG applied time-domain features and VFCDM statistical feature analysis to classify ADHD and ASD. Using the GradBoost algorithm, the study showed a high accuracy of 0.87 and found a higher performance using 446.Right Eye configuration [21]. Yet, there are discrepancies in ERG application. For instance, ERG paired up with random forest model to classify ADHD and ASD after model optimization using repeated cross validation and synthetic minority over-sampling technique only yielded accuracy of 0.69 [21]. This differences might be contributed to data imbalance and demographic variability.

2.5. Machine Learning

Machine learning (ML) is a powerful pre-screening tool widely used in healthcare and research to support diagnostics for conditions like Alzheimer's, schizophrenia, Parkinson's disease, dementia, ADHD, and other psychological disorders [22]. This review centers on supervised ML, which aims to predict outcomes based on input data and identify relationships between features and outcomes in new datasets [23].

Model validity in ML is often assessed using sensitivity and specificity metrics, which together measure validity and generalizability. Sensitivity measures the model's ability to detect true positives, while specificity reflects its accuracy in identifying true negatives [24]. The Receiver Operating Characteristic (ROC) curve visualizes model validity by plotting the true positive rate against the false positive rate. The Area Under the Curve (AUC) of the ROC summarizes the model's discriminative power, with values ranging from 0.5 (random performance) to 1.0 (perfect discrimination) [25].

Cross-validation is essential in machine learning for preventing overfitting and improving model robustness. It involves training and testing a model on various data subsets to provide a reliable performance estimate. Leave-One-Out Cross-Validation (LOOCV) maximizes data use by testing on each data point individually but is computationally intensive and prone to high variance, especially with small datasets. In contrast, 10-fold Cross-Validation, which splits the data into ten parts, balances bias and variance more effectively. It is less computationally demanding and enhances generalizability by averaging results across folds, though it still requires significant resources with large datasets.

2.6. Research Questions

While ADHD and ASD share overlapping characteristics, several methods can support medical practitioners in accurately differentiating between the two. Machine learning models have emerged as promising decision-support tools in the diagnostic process for ADHD and ASD highlighted that ML applications deepen researchers'understanding of the intersections between neuroscience and clinical psychology by analyzing complex patterns of neurobiological activity relevant to these disorders [26]. Previous researchers have used self-report data, behavioral observations, and neuroactivity data to train and test machine learning models for classification, yielding promising results.

Despite the increased accessibility of ML in clinical trials, research on these applications remains highly variable, with differences in sample sizes, training methodologies, and model selection, yielding a diagnostic AUC range of 60% to 90% [22]. Furthermore, limited research focuses specifically on ML classification of ADHD and ASD due to the inherent challenge of differentiating these disorders given their symptom overlap.

Consequently, there is a critical need to evaluate the existing performance of various ML models by comparing metrics such as accuracy, specificity, sensitivity, AUC, and training methodologies. A systematic comparative analysis would facilitate a more structured understanding and improved interpretability of ML models within clinical contexts. This systematic review aims to assess the effectiveness of ML models in distinguishing pediatric ADHD from ASD by examining data complexity, validation methods, and performance metrics. The review seeks to address the following questions regarding the application of ML in this domain:

1. How do researchers enhance the reliability and accuracy of ML models for ADHD and ASD classification?
2. What constitutes the most effective ML model design for accurately classifying ADHD and ASD?

3. Methodology

The aim of the systematic review was to evaluate the effectiveness of machine learning in classifying ADHD and ADHD in children. The review followed the PRISMA 2020 guidelines and PRISMA 2020 flow chart for the searching strategy [27].

3.1. Eligibility Criteria

The eligibility criteria for this systematic review included studies focused on ADHD diagnosed according to DSM-IV or DSM-5 criteria in children aged 2–18 years, published between 2015 and 2024. European studies were eligible if diagnoses were conducted under ICD-10 standards. Studies had to involve representative populations, be empirical, provide full-text access, and focus on machine learning techniques for ADHD and ASD classification, published in SSCI or SCI-E journals with reported validity and reliability data. Exclusions applied to studies involving small or referral populations, children with severe mental retardation (IQ < 55), studies published before 2015, reviews, articles irrelevant to classification, and those lacking full-text access, SSCI/SCI-E indexing, or validity and reliability metrics (Table 1).

However, due to the limited number of articles relevant to machine learning classification of ADHD and ASD, studies reporting at least one confusion metric were allowed to ensure the review’s robustness. For studies presenting only the AUC measure, this metric illustrates the relationship between specificity and sensitivity graphically. Studies that only reported accuracy were compared based on accuracy in relation to different feature selection methods. Furthermore, if a study did not employ cross-validation (CV) in model optimization, its exclusion provided an opportunity to analyze the impact of CV on machine learning performance and implications for future improvements.

Table 1. Inclusion and Exclusion Criteria

Inclusion	**Exclusion**
Studies diagnosing ADHD based on DSM-IV and/or DSM-V and/or ICD-10	ADHD and ASD with symptoms based on other diagnosing criteria
Participants aged 2-18	Adults aged above 18
Studies published between 2015-2024	Published before 2015
Empirical studies	Reviews
Relevant to review questions	Irrelevant to machine learning classification of ADHD and ASD
Machine learning for diagnosing or classifying ADHD and ASD	Machine learning for studying or treating ADHD and ASD
Studies published in journals indexed in SSCI or SCI-E	Studies not included in SSCI or SCI-E journals
Access to full articles	No access to full articles

3.2. Search Strategy

The primary database used for this systematic review was Web of Science. Additionally, relevant journals were sourced from IEEE, Springer Nature, Scientific Reports, Translational Psychiatry, and Applied Computational Intelligence and Soft Computing. The search was restricted to publications from 2015 to 2024, using keywords such as “ADHD,” “ASD,” “machine learning,” “classification,” and “children” to filter for relevant studies. Filters were applied on Web of Science to exclude articles published before 2016, non-English publications, and review articles, ensuring a focused and relevant selection of studies (Figure 1).

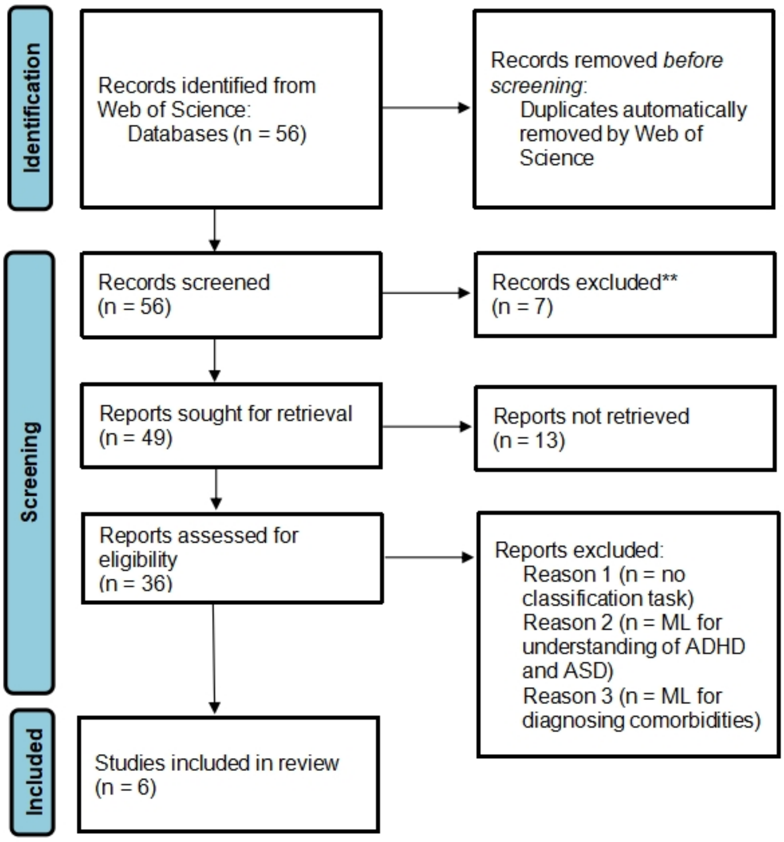

Figure 1. Prisma flow diagram

3.3. Data Collection

The data collection process was conducted using six selected articles that met the defined inclusion and exclusion criteria. Data were organized in a spreadsheet in Excel, capturing details such as publication title, publication date, journal name, measurements of validity and reliability, participant information, and types of machine learning models employed. To ensure objectivity and minimize bias, two independent researchers carried out the data collection. Any discrepancies encountered were resolved through discussions between the researchers (Table 2).

Table 2. Coding framework

Code	Subcode	Crowdsourced validation of a machine-learning classification system for autism and ADHD	A data driven machine learning approach to differentiate between autism spectrum disorder and attention-deficit/hyperactivity disorder based on the best-practice diagnostic instruments for autism	ADHD and ASD Classification Based on Emotional Recognition Data	Statistical Analysis and Multimodal Classification on Noisy Eye Tracker and Application Log Data of Children with Autism and ADHD	Spectral analysis of Electroretinography to differentiate autism spectrum disorder and attention deficit hyperactivity disorder.	Detecting Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder Using Multimodal Time-Frequency Analysis with Machine Learning Using the Electroretinogram from Two Flash Strengths
Background	Publication data	2017	2022	2016	2018	2023	2024
	Journal name	TP	SR	IEEE	IACS	IEEE	SN
	Country of primary affiliation	United State	Germany	Turkey	Turkey	UK, Australia, United States	United States
Participants	Age range	2-17	5-18	9–10	9-10	3-13	11-14
	Health condition	Archival: ASD=2775, ADHD=150, Survey:	ASD=574, ADHD=164, ASD/ADHD=133, Control=344	ASD=18, ADHD=30 Control=13	ADHD=12, ASD=12, Control=10	ADHD=46, ASD=94	ASD=94, ADHD=46, Control=146

		ASD=248, ADHD=174					
Methodolgy	Setting	C + W	C	C	C	C	C
	Data collection	Clinical diagnosis; Survey	ADOS, ADI-R	ER	SMI + TE	ERG	ERG
	Input modality	Archival + crowdsourced survey data	Diagnostic sores + behavioral features	APL	APL + Fixation	ERF waveform	ERG waveform
	Data type	B	B	B	B	P	P
Model optimization	Validation	5-fold CV	10-fold CV	Nan	Leave-one-out CV	10-fold CV	10-fold CV
	Feature selection	Enet, LAD	SVM	ReliefF	RF	VFCDM	Time, time-frequency analysis
	Hyperparameter tuning	3-fold CV, Grid Search	Nan	Nan	Nan	GS + 3-fold CV	GS + 3-fold CV
	Data sampling	Subsampling	Nan	Random sampling	Tomek links		Synthetic minority over-sampling technique
Machine learning	Best ML model	Enet, LDA	SVM	Adaboost	RF	GradBoost	RF
	Features	5 SRS questions	5 features from both ADOS and ADI-R	RC + RT	RC + RT + Raw fixation, ET_log data	VFCDM	Time Domain indices DWT VFCDM
Results	Specificity	Nan	0.82	Nan	0.8333	0.87	Nan
	Sensitivity	Nan	0.73	Nan	0.4167	0.79	Nan
	AUC	0.89 +- 0.01	≥ 0.91	Nan	Nan	0.87	0.69
	Accuracy			0.9	0.7083	0.84	0.8

Note: TP = Translational Psychiatry; SR = Springer Nature; IEEE = Institute of Electrical and Electronics Engineers; IACS = Applied Computational Intelligence and Soft Computing; C = Clinical;

W = Real-world; ADOS = Autism Diagnosis Observation Schedule; ADI-R = Autism Diagnostic Interview, Revised; SMI = SMI eye tracking glasses; TE = TrackEmo software program; ER = emotional recognition experiment; STOTE = synthetic minority over-sampling technique; SRS = social responsiveness scale; ERG = electroretinogram; B = behavioral; P = physiological; CV = cross validation; APL = application log data; VFCDM = Mean, max, variance, kurtosis, skewness, entropy, inter-quartile range; Time Domain indices = Ta, Va , Tb, Vb; DWT = DWT coefficients DWT coefficients in a20, a40, b20, b40, OP80, OP160 descriptors; RF = random forest; GradBoost = gradient boosting; Adaboost = adaptive boosting; SVM = support vecter machine; ENet = elastic net; LDA = linear discriminant analysis; RC = respond correctness; RT = response time; GS = GridSearch; Nan = information not found;

4. Results

The emergence of various machine learning techniques has yielded different results in the classification tasks for ADHD and ASD. Variations in the application of machine learning have complicated the selection of the most promising training approach, necessitating a structured analysis of these studies' use of machine learning, confusion metrics, and model optimization techniques. This section highlights the main findings from the six selected studies and examines the methodologies used to improve the interpretability and generalizability of machine learning performance in diagnosing ADHD and ASD.

The six journals reviewed were mainly associated with institutions in the United States, Germany, the United Kingdom, Australia, and Turkey. They were published in journals with impact factors ranging from 2.4 to 25, indicating a reliable level of quality. The studies included participants aged 2–18 years, categorized as healthy, ADHD, ASD, or ADHD with ASD, according to diagnostic criteria from DSM-IV, DSM-V, ADOS, ADI-R, or ICD-10. Although multiple machine learning algorithms were applied in these studies, this review focuses on the models that achieved the highest performance scores.

Four studies conducted experiments in controlled clinical settings with supervision, while two studies combined clinical diagnoses with real-world reports. Behavioral data included online parental self-reports (SRS) and assessments through ADOS and ADI-R, which are standardized tools for evaluating social interaction, emotional regulation, and daily functioning in children with these conditions. Additionally, one study from Turkey used APL data paired with the Adaboost algorithm to compare Adaboost's performance in classifying ADHD and ASD, with and without the use of ReliefF for feature selection.

In a study utilizing TrackEmo software and SMI eye-tracking glasses, researchers collected APL (activity pattern log) data and eye fixation data to classify ADHD and ASD using a random forest model [28]. Physiological data such as this provides insights into neuroactivity differences between ADHD and ASD. Two other studies used ERG waveforms to detect retinal activity variations in response to light, employing 10-fold cross-validation and VFCDM to train and select statistical features from ERG data. Neuroimaging data was excluded due to the limited number of relevant studies and the lengthy process required for neuroimaging in classifying ADHD and ASD.

Physiological and behavioral data are effective for large-scale data collection, as they can be gathered remotely or in controlled settings, which can enhance data quality. Therefore, this review primarily focuses on these two data types. This section details the procedures used in each selected study, presenting key findings such as validity measures, model optimization techniques, and participant demographics. These insights will form the basis for comparing and analyzing the performance of various machine learning models in distinguishing between ADHD and ASD.

A novel study conducted in Turkey aimed to classify individuals with ASD and ADHD through an emotion recognition task [29]. The study included 18 participants with ASD, 30 with ADHD, and 13 controls, with average ages of 10.5, 9.46, and 9.22 years, respectively. Participants viewed facial images expressing various emotions and identified the depicted emotion. Researchers recorded response accuracy (RC) and response time (RT) for each response. To improve classification accuracy, they used group redefinition to create binary classification scenarios that highlighted distinctive traits of each group. The study employed ReliefF, a feature selection algorithm, to rank features based on relevance, specifically identifying the most informative images by comparing within-class and between-class proximities. Results showed that the AdaBoost algorithm achieved the highest accuracy, reaching up to 90% using only RT without feature selection. Combining RC and RT with ReliefF achieved 80% accuracy. Overall, the study underscored the effectiveness of feature selection and group redefinition in improving classification performance, with different combinations of RC, RT, or both enhancing results.

Ozturk et al. (2018) explored a multimodal approach by combining eye-tracking data with application log data to classify ADHD and ASD, reaching a peak accuracy of 70.83% with a random forest model. This emotion recognition experiment involved 12 ASD participants, 12 ADHD participants, and 10 neurotypical controls (ages 9-10, all with IQs above 70). Participants viewed images of faces and identified emotions while wearing eye-tracking devices to capture fixation data. Data recorded included RT and RC as application log data, combined with fixation data and ET_log data. Feature selection was applied to improve predictive quality, and Tomek links were used for undersampling to address data imbalance. To avoid overfitting, the study used leave-one-out cross-validation (LOOCV). Three models—Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR)—were trained, with RF showing the highest validity, achieving 41.67% sensitivity and 83.33% specificity in ADHD versus ASD classification. Notably, pupil diameter emerged as the most important feature, surpassing RT and RC in predictive value. While Tomek links slightly improved performance for ASD versus control classification, they had minimal effect on RF's specificity in distinguishing ADHD from ASD. Despite RF's accuracy of 70.83%, the model's discriminatory power was limited, possibly due to overlapping symptoms in ADHD and ASD, which complicates precise classification. This study highlights the promise and challenges of using multimodal data to distinguish ADHD and ASD, particularly noting pupil diameter's value in classification, while acknowledging that symptom overlap poses significant challenges to achieving high accuracy.

Duda et al. (2017) conducted a crowdsourced study using both archival data from clinically diagnosed individuals with ADHD or ASD and parent-reported survey data gathered through a crowdsourcing approach. They used these datasets to train and test five machine learning models, assessing their classification accuracy and generalizability. Each model was trained using 5-fold cross-validation and subsampling, with a 3-fold cross-validation employed to optimize parameter selection. The study tested three conditions: training on archival data and testing on survey data, training on survey data and testing on archival data, and training and testing on a combined dataset. The two top-performing models were ENet, a logistic regression model with lasso and ridge regularization, and Linear Discriminant Analysis (LDA). ENet included an in-built feature selection process, which identified five key SRS survey questions, and, along with LDA, consistently outperformed simpler models, achieving an AUC of 0.89. Notably, the models achieved the highest AUC (over 0.9) when trained solely on the archival dataset.

In 2022, Wolff et al. enhanced the previous research by incorporating ADOS and ADI-R assessments to distinguish ASD from healthy controls, ADHD, and comorbid ASD/ADHD. This German study used a larger, diverse sample of 574 ASD, 164 ADHD, 133 ASD/ADHD, and 344 control participants. The dataset was split into 10 folds, with feature selection conducted on the 9-fold training sets using Support Vector Machine (SVM) models [30]. The top behavioral features centered on social communication and restrictive/repetitive behaviors. SVM achieved an AUC of above 0.91, with 82% specificity and 73% sensitivity. The study found overlapping features in social and repetitive behavior domains across ASD and ADHD classifications, such as ADOS item A8 (conversation) and eye contact, indicating risks of misdiagnosis.

In contrast to these behavior-focused studies, Grazioli et al. (2024) and Manjur et al. (2023) explored electroretinography (ERG) data for classifying ADHD and ASD. Grazioli et al. used spectral analysis within a Gradient Boosting framework, testing 142 participants (46 ADHD, 94 ASD) with time-domain ERG parameters and frequency-band decomposition using Variable Frequency Complex Demodulation (VFCDM). Random Forest (RF) was employed for feature selection, while various machine learning models were tested, with Gradient Boosting and XGBoost performing the best. The GradBoost model achieved 87% sensitivity, 79% specificity, and an AUC of 0.87.

Expanding on ERG-based classification, Manjur et al. (2024) involved a larger cohort of 286 participants (146 controls, 94 ASD, 46 ADHD). Using both time-domain and time-frequency analyses (Discrete Wavelet Transform and VFCDM), the study employed a rigorous 10-fold cross-validation to avoid overfitting. Among models, RF achieved the highest accuracy (70%) with a flash strength of 446 Td.s, reaching an AUC of 0.69 in distinguishing ADHD from ASD, emphasizing ERG's potential as a biomarker for these neurodevelopmental disorders.

5. Discussion and Conclusion

Multidisciplinary research on classifying ADHD and ASD is crucial for improving the clinical diagnosis process, offering a more accessible, cost-effective, and objective alternative to traditional methods. Conventional diagnosis often depends on clinical interviews, checklists, and questionnaires, which can be subjective and influenced by clinicians' perspectives. Machine learning (ML) provides a promising approach, offering greater reliability and precision in diagnosing ADHD and ASD as models are continuously optimized for clinical use. Discrepancies across studies are addressed in the following section.

Performance differences in ML models for ADHD and ASD classification generally stem from three main factors. The first factor, data complexity, includes dataset diversity, participant count, and participant conditions. The second factor, cross-validation, helps prevent overfitting and improve generalizability. Techniques like GridSearch with 3-fold cross-validation are used for hyperparameter tuning, optimizing settings such as learning rates, regularization, and kernel types. Lastly, feature selection refines models by identifying the most predictive features. While other optimization methods like SMOTE, Tomek links, and subsampling are effective, this review focuses on the main patterns observed, with additional methods discussed as supplementary considerations for future studies.

Using diverse participants is vital for generalizability, allowing models to learn from varied features. Studies including both ADHD and ASD participants have shown higher performance than those focusing on only one condition. For example, Wolff et al.'s study achieved 82% specificity and 73% sensitivity, with an AUC over 0.91, likely due to the inclusion of both ADHD and ASD participants. This performance exceeded that of Duda et al.'s research, underscoring the benefits of diverse participant conditions. Furthermore, Wolff et al. and Duda et al. had more participants than studies focused on emotion recognition, eye-tracking, or ERG, which indicates that increasing sample sizes could enhance model robustness and accuracy. However, the datasets in Wolff et al. and Duda et al. were obtained from clinical databases or online surveys, which can yield larger, quicker datasets but may introduce biases [19, 30]. This trade-off between dataset size and data quality suggests a need for careful filtering of large datasets to ensure high-quality samples.

Validation methods in machine learning vary by dataset size and study objectives, with 10-fold cross-validation commonly used for its ability to enhance generalizability and provide stable performance estimates, particularly in studies with sample sizes over 100. For hyperparameter tuning, some studies have applied GridSearch with 3-fold cross-validation, as seen in two ERG studies. For example, Manjur et al. (2023) used this method to optimize statistical features from ERG waveforms, allowing GradBoost to achieve an accuracy of 84%, surpassing RF (80%). GradBoost's sequential learning structure effectively captures complex data patterns, making it well-suited for detecting nuanced ERG signal variations, whereas RF's independent tree structure offers robustness but lacks precision with subtle interactions. These differences between GradBoost and RF emphasize the importance of model architecture and data structure on performance. While Duda et al. also used

GridSearch with 3-fold cross-validation, their higher AUC score underscores the role of dataset diversity and model selection as key performance factors.

Feature selection is integral to improving model performance, with many machine learning models having built-in mechanisms for identifying predictive features. For instance, RF evaluates each feature's effectiveness in decision trees, and Tomek links, an undersampling technique, can remove ambiguous data, enhancing RF's ranking accuracy. In the eye-tracking study, RF identified pupil diameter, pupil size, and fixation positions as top predictive features, although classification accuracy remained lower than for ASD/control and ADHD/control classifications. SVM also incorporates feature selection, with Wolff et al. reporting high AUC values with balanced specificity and sensitivity, whether or not feature selection was applied. This could be due to the high discriminative power of clinical assessments like ADOS and ADI-R. Ultimately, while feature selection boosts accuracy, other optimization tools, such as Tomek links and high-quality datasets, can further support this process.

This systematic review has several limitations that should be acknowledged. First, the studies included participants diagnosed according to DSM-5, DSM-IV, or ICD-10 standards, introducing variability in the definitions and diagnoses of these conditions, which may affect the generalizability of the findings. Second, inconsistencies in reporting performance metrics across studies make it challenging to directly compare the effectiveness of different machine learning models. Lastly, this review included only six studies, which may limit the comprehensiveness of the analysis. Future reviews should aim to include a broader range of studies with standardized diagnostic criteria and comprehensive performance metrics to strengthen the findings.

The current review highlights the importance of machine learning in ADHD and ASD classification, with Wolff et al.'s study exemplifying best practices. Using high-quality clinical data from structured assessments, comprehensive feature selection, and cross-validation, Wolff's model achieved high accuracy, AUC, sensitivity, and specificity. Its inclusion of a diverse participant pool, covering both ADHD and ASD, enhances the model's ability to capture subtle distinctions. Future research could expand on this by incorporating complex data from controlled settings and mechanisms like Tomek links to improve interpretability. Additionally, exploring ADHD and ASD as potential confounding variables could address challenges in distinguishing these conditions.

References

[1] Brown, T. E. (2008). ADD/ADHD and Impaired Executive Function in Clinical Practice. Current Psychiatry Reports, 10(5), 407–411. https://doi.org/10.1007/s11920-008-0065-7

[2] Lange, K. W., Reichl, S., Lange, K. M., Tucha, L., & Tucha, O. (2010). The history of attention deficit hyperactivity disorder. ADHD Attention Deficit and Hyperactivity Disorders, 2(4), 241–255. https://doi.org/10.1007/s12402-010-0045-8

[3] Li, Y., Yan, X., Li, Q., Li, Q., Xu, G., Lu, J., & Yang, W. (2023). Prevalence and Trends in Diagnosed ADHD Among US Children and Adolescents, 2017-2022. JAMA Network Open, 6(10), e2336872. https://doi.org/10.1001/jamanetworkopen.2023.36872

[4] Abdelnour, E., Jansen, M. O., & Gold, J. A. (2022). ADHD Diagnostic Trends: Increased Recognition or Overdiagnosis? Missouri Medicine, 119(5), 467–473. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9616454/

[5] Danielson, M. L., Claussen, A. H., Bitsko, R. H., Katz, S. M., Newsome, K., Blumberg, S. J., Kogan, M. D., & Ghandour, R. (2024). ADHD Prevalence Among U.S. Children and Adolescents in 2022: Diagnosis, Severity, Co-Occurring Disorders, and Treatment. Journal of Clinical Child & Adolescent Psychology. https://www.tandfonline.com/doi/abs/10.1080/15374416.2024.2335625

[6] Oh, S., Joung, Y.-S., Chung, T.-M., Lee, J., Seok, B. J., Kim, N., & Son, H. M. (2024). Diagnosis of ADHD using virtual reality and artificial intelligence: an exploratory study of clinical applications. Frontiers in Psychiatry, 15, 1383547. https://doi.org/10.3389/fpsyt.2024.1383547

[7] Gualtieri, C. T., & Johnson, L. G. (2005). ADHD: Is Objective Diagnosis Possible? Psychiatry (Edgmont), 2(11), 44–53. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2993524/

[8] Mohd Javaid, Abid Haleem, Ravi Pratap Singh, & Rajiv Suman. (2022). Significance of machine learning in healthcare: Features, pillars and applications. https://www.researchgate.net/publication/361105025_Significance_of_machine_learning_in_healthcare_Features_pillars_and_applications

[9] American Psychiatric Association (Ed.). (1998). Diagnostic and statistical manual of mental disorders: DSM-IV; includes ICD-9-CM codes effective 1. Oct. 96 (4. ed., 7. print).

[10] Lordan, R., Storni, C., & De Benedictis, C. A. (2021). Autism Spectrum Disorders: Diagnosis and Treatment (A. M. Grabrucker, Ed.). Exon Publications. http://www.ncbi.nlm.nih.gov/books/NBK573609/

[11] Yu, Y., Ozonoff, S., & Miller, M. (2023). Assessment of Autism Spectrum Disorder. Assessment, 31(1), 24. https://doi.org/10.1177/10731911231173089

[12] Substance Abuse and Mental Health Services Administration. (2016). DSM-5 Changes: Implications for Child Serious Emotional Disturbance. Substance Abuse and Mental Health Services Administration (US). http://www.ncbi.nlm.nih.gov/books/NBK519708/

[13] Miller, L. L., Gustafsson, H. C., Tipsord, J., Song, M., Nousen, E., Dieckmann, N., & Nigg, J. T. (2018). Is the association of ADHD with socio-economic disadvantage explained by child comorbid externalizing problems or parent ADHD? Journal of Abnormal Child Psychology, 46(5), 951. https://doi.org/10.1007/s10802-017-0356-8

[14] Scandurra, V., Gialloreti, L. E., Barbanera, F., Scordo, M. R., Pierini, A., & Canitano, R. (2019). Neurodevelopmental Disorders and Adaptive Functions: A Study of Children With Autism Spectrum Disorders (ASD) and/or Attention Deficit and Hyperactivity Disorder (ADHD). Frontiers in Psychiatry, 10, 673. https://doi.org/10.3389/fpsyt.2019.00673

[15] Rusting, R. (2018). Decoding the overlap between autism and ADHD. Spectrum. https://doi.org/10.53053/KCZY8213

[16] Leitner, Y. (2014). The Co-Occurrence of Autism and Attention Deficit Hyperactivity Disorder in Children – What Do We Know? Frontiers in Human Neuroscience, 8, 268. https://doi.org/10.3389/fnhum.2014.00268

[17] Baltaci, S., & Gokcay, D. (2016). Stress Detection in Human–Computer Interaction: Fusion of Pupil Dilation and Facial Temperature Features. International Journal of Human–Computer Interaction, 32(12), 956–966. https://doi.org/10.1080/10447318.2016.1220069

[18] Poyraz Findik, O. T., Arman, A., Bulut, G., Findik, O., Yilmaz, S., Aslan Genç, H., Yazgan, Y., Teker, U., & Cataltepe, Z. (2018). Statistical Analysis and Multimodal Classification on Noisy Eye Tracker and Application Log Data of Children with Autism and ADHD. Intelligent Automation and Soft Computing, 1. https://doi.org/10.31209/2018.100000058

[19] Duda, M., Haber, N., Daniels, J., & Wall, D. P. (2017). Crowdsourced validation of a machine-learning classification system for autism and ADHD. Translational Psychiatry, 7(5), e1133. https://doi.org/10.1038/tp.2017.86

[20] Grazioli, S., Crippa, A., Rosi, E., Candelieri, A., Ceccarelli, S. B., Mauri, M., Manzoni, M., Mauri, V., Trabattoni, S., Molteni, M., Colombo, P., & Nobile, M. (2024). Exploring telediagnostic procedures in child neuropsychiatry: addressing ADHD diagnosis and autism symptoms through supervised machine learning. European Child & Adolescent Psychiatry, 33(1), 139–149. https://doi.org/10.1007/s00787-023-02145-4

[21] Manjur, S. M., Hossain, M.-B., Constable, P. A., Thompson, D. A., Marmolejo-Ramos, F., Lee, I. O., & Posada Quintero, H. F. (2023). Spectral analysis of Electroretinography to differentiate autism spectrum disorder and attention deficit hyperactivity disorder. 1–4. https://doi.org/10.1109/BHI58575.2023.10313406

[22] Quaak, M., van de Mortel, L., Thomas, R. M., & van Wingen, G. (2021). Deep learning applications for the classification of psychiatric disorders using neuroimaging data: Systematic review and meta-analysis. NeuroImage: Clinical, 30, 102584. https://doi.org/10.1016/j.nicl.2021.102584

[23] Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, 2(3), 160. https://doi.org/10.1007/s42979-021-00592-x

[24] Florkowski, C. M. (2008). Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. The Clinical Biochemist Reviews, 29(Suppl 1), S83. https://pmc.ncbi.nlm.nih.gov/articles/PMC2556590/

[25] Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. Turkish Journal of Emergency Medicine, 23(4), 195. https://doi.org/10.4103/tjem.tjem_182_23

[26] Eslami, T., Almuqhim, F., Raiker, J. S., & Saeed, F. (2021). Machine Learning Methods for Diagnosing Autism Spectrum Disorder and Attention- Deficit/Hyperactivity Disorder Using Functional and Structural MRI: A Survey. Frontiers in Neuroinformatics, 14, 575999. https://doi.org/10.3389/fninf.2020.575999

[27] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

[28] Ozturk, M., Arman, A., Bulut, G., Findik, O., Yilmaz, S., Genc, H., Yazgan, M., Teker, U., & Cataltepe, Z. (2018). Statistical Analysis and Multimodal Classification on Noisy Eye Tracker and Application Log Data of Children with Autism and ADHD. Intelligent Automation and Soft Computing, 1--1. https://doi.org/10.31209/2018.100000058

[29] Ozturk, M. U., Arman, A. R., Yilmaz, S. S., Findik, O. P., & Cataltepe, Z. (2016). ADHD and ASD Classification Based on Emotion Recognition Data. 15th IEEE International Conference on Machine Learning and Applications. http://www.xueshufan.com/publication/2583477452

[30] Wolff, N., Kohls, G., Mack, J. T., Vahid, A., Elster, E. M., Stroth, S., Poustka, L., Kuepper, C., Roepke, S., Kamp-Becker, I., & Roessner, V. (2022). A data driven machine learning approach to differentiate between autism spectrum disorder and attention-deficit/hyperactivity disorder based on the best-practice diagnostic instruments for autism. Scientific Reports, 12(1), 18744. https://doi.org/10.1038/s41598-022-21719-x