VAWKUM Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2025, ISSN(e): 2308-8168, ISSN(p): 2411-6335
Volume 13, Number 1, January-June 2025 pp:178-189

Advancements in EEG-Based Machine Learning Techniques for Early Autism Spectrum Disorder Diagnosis: A Review

Saad Akbar [1*], **Shahid Munir Shah** [1], **Rubaishe Tahir** [1], **Muhammad Mujtaba Khan** [1], **Nuha Aamir** [1]

[1] Faculty of Engineering Sciences and Technology, Department of Computing, Hamdard University, Karachi, Pakistan

Keywords: *Autism Spectrum Disorder (ASD), Electroencephalography (EEG), Machine Learning (ML), Deep Learning (DL), Early ASD Diagnosis.*

Journal Info:
Submitted:
March 10, 2025
Accepted:
April 27, 2025
Published:
May 09, 2025

Abstract

Autism Spectrum Disorder (ASD) is a neuro developmental condition characterized by impairments in social interaction, communication, and repetitive behaviors. Electroencephalography (EEG) has gained prominence as a reliable tool for ASD diagnosis, capturing critical behavioral patterns. Researchers have applied various Machine Learning (ML) techniques to enhance ASD detection, achieving notable accuracy. Studies using feature selection with ML classifiers have reported up to 100% accuracy in children as young as 6–12 months. Other approaches integrating EEG with behavioral features such as eye gaze, facial gestures, and body movements have attained classification accuracies as high as 87.5%. Additionally, resting-state EEG studies have explored microstate differences between ASD and neurotypical individuals. Several ML models, including Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and ensemble methods like Random Forest and Naïve Bayes, have demonstrated classification precision between 90% and 99%. However, challenges such as data heterogeneity and limited sample sizes hinder clinical implementation. This review highlights the most notable EEG-based ML studies for ASD diagnosis and emphasizes the need for further research to refine these techniques for broader clinical adoption.

***Correspondence author email address:** akbersaad@yahoo.com
DOI: 10.21015/vtcs.v13i1.2095

1 Introduction

ASD is one of the neurological disorders that operate on ways in which an individual processes information and acts and often obvious itself in early childhood [1]. ASD is a group of neurodevelopment disorders that affect communication, social interaction and causes repetitive behaviors [2]. It is important, specifically in the early years, since the health related approaches when applied early enough comes with a very good result with regards to development. Disability identification in early childhood previously focused on behavioral observations, which is unfair and can cause a child to be diagnosed later. Several approaches to quantify and visualize the brain networks have been made which includes the resting state and task-based EEG. Functional integration analysis involves the use of resting state EEG so as to identify the brain activity in a situation in which the patient is not carrying out an activity. EEG signals are electrical voltage extracted on the electrodes by brain electromagnetic signal (BEMS) [3]. These are electrical voltage created on the electrodes by the BEMS [2]. EEG records brain electrical activity and presents tangible opportunity to investigate neural patterns of individuals with ASD. When used in conjunction with machine learning algorithms, the arrays of data that can be derived from an EEG can be screened for those that may suggest the presence of ASD. Together with the improved diagnostic precision and earlier detection methods, this is quite appealing. This perspective implies the idea that basic neural abnormalities which make up ASD are only temporary and are subsequently difficult to diagnose after the developmental phase [4].

The identification of ASD at a very young age usually initiate early functioning that has effects that are socially positive. But traditional treatment approaches cannot easily provide such accurate outcomes in diagnosing the condition at an early stage. An evaluation done by pediatrics with the help of a documented analysis mentioned that an autistic child at the age of around 24 months are highly incapable to construct two functional words which are not imitate or reiterate [2]. Thus, increasing attention is paid to the use of more objective neurophysiological parameters including EEG to search for biomarkers of ASD. Table 1. below shows the EEG Brain Signal of Normal Human.

Table 1. EEG Brain Signal of Normal Human Being [2]

Level	Range of Frequency	Estimated EEG Classification
1.1	(0.5–4 Hz)	Delta
1.2	(4–8 Hz)	Theta
1.3	(8–13 Hz)	Alpha
1.4	(13–30 Hz)	Beta
1.5	(30–100 Hz)	High Gamma

Currently, WHO estimates that 1 child in every 160 has an ASD, the rate of diagnosed autism rises by 40% increasing from one in 100 to about one in 70 in Australia. With current data, the CDC pointed out that one child in about 54 children is diagnosed with an ASD in the United States in 2020. As detail in the World Health Organization, one child in every 160 suffers from ASD all over the world [5], Autism pre valance rate escalate approximate 40% from one in 100 to an around one in 70 in Australia . The core for Disease Control and Prevention (CDC) declared that one child in about 54 children is found with an ASD in the U.S. in 2020.

As suspected, DL is a subset of ML, according to the researchers, DL is characterized by the use of many hidden neurons and layers, often more than two, as an architectural advantage in addition to a new training approach when data replenishment is available, DL begins to build up and complete the areas where explanation is impossible for a human-driven system. Still it may be machine learning or deep learning both contain supervised learning and unsupervised learning. Supervised learning is constructing a predictive method out of the input and output data provided. While unsupervised learning is a method of grouping and interpreting data based only on input data [2]. This research survey focuses on the application of EEG data alongside with machine learning for enhancing the diagnosis of ASD. From the studies reviewed in this research, there is a correlation of the usage of EEG in capturing neural activity, and the involvement of models such as CNN and SVM where CNN on the spectrogram of EEG performs a classification with an accuracy reaching 99.15%. Entropy based algorithms, wavelet transforms and time-frequency analysis are used to extract features to distinguish ASD patterns [6].

According to Mehmet Baygin et al. lightweight feature evocation, they propose a novel deep hybrid approach on extracting deep features, and achieve an accuracy of 96.44% with the SVM classifier [7] [8]. ML based classification success lies in the extraction of meaningful features from the EEG signals and until now, different researchers have tried this approach for ASD classification. In a later study [9], they used STFT and statistical analysis with KNN, with a data set of 28 subjects (17 ASD, 11 control group) they reported accuracy of 96.4%. Bosl et al[10], described the diagnostic framework to employ EEG data for the biomarker in children at higher risk for ASD. The extracted features they used MMSE and KNN, naïve bayes and received classification accuracy higher than 90% on 79 infants (46 HRA and 33 controls) between the age of 6 to 24 months [11]. The conclusion of these many works is that a child with ASD stays typical in the neural activity during the frequency band simulation. Therefore, there is high hope seek more findings to identify the exact area of the brain where the alteration commences, according to the EEG signal. This is further supported by [2], where it will be further explained that if the rate at which the brain responds to visual and/or audio input is determined that it will help in classifying the autism and diagnose the disorder earlier.

2 Related Studies

The present review is devoted to the consideration of the application of EEG data in ASD's early identification by critically assessing prior research and epidemiological data. Data collected via EEG has be-come pertinent in ASD diagnosis because it involves real time analysis of activity in brain that standard behavioral assessments cannot provide. Quantitative assessment of functional development of the brain is challenging primarily because of the structural and organizational complexity of the cortex as the center of the nervous system and secondly due to the lack of adequate noninvasive approaches to monitor and quantify the function in babies. New nonlinear approaches to analyze the brain electrical activity recorded with scalp electrodes might help detect the differences in infant brain connectivity. For instance, the entropy of EEG electrodes with an electrode distance of greater than 3 cm was coarser in children with autism compared to a group of normal developing children [12], which is consistent with the weak FC theory of autistic brains [10], Researchers collected EEG data from both ASD and control participants with varying age, gender, and ASD symptom severity in order to increase the model's robustness and accuracy across different subpopulations. Population of epidemiological publications that joined the criteria was recognized via organized re-view technique and inputs from prior first-stage of systematic reviews of epidemiological surveys were incorporated to advance prior disruptions of ken. Complete or final diagnostic results and the minimum of two EEG recordings were obtained for 188 children and used in this study. In this research, all visits were considered as singular interactions In other words all the visits were assumed to be separate and did not build on one another. For instance, all observations of the EEG made at 12-month visits are used to predict outcomes regardless of the measures recorded at other ages in the same child. While a growth trajectory analysis was outside the aims of the presented research, one classification test was completed by joining the measurements from 6 months and 9 months into one set of features for the subjects who completed 6 and 9 month visits [4]. In three age groups (4 years old), the prevalence estimates in studies with children older than 4 years old were significantly lower than estimates in younger children (odds ratio: 0.32; 95 percent CI: 0.16, 0.68). Nevertheless, this result was based on three very large studies only. However, if three studies with the largest sample size were excluded this association was not evident (P= 0.33). There was no statistically meaningful correlation between the prevalence estimate and group sample size: 5000 donors, 5000 to 7500, and greaterthan 7500. No other covariate influences the prevalence estimates provided for analysis in this study [13]. These reviews provided a strong foundation for understanding ASD prevalence and variation across demographics.

To precede analysis, pre-processing techniques are used when data have been gathered on patients and healthy individuals. The initial steps of filtering carry out the international standard norms and algorithms for EEG elimination of noise and artifacts. The use of various higher order methods are employed for extracting signal features such as wavelet transformations, entropy and spectrogram images of time-frequency domain. Data were collected from 79 different infants: 46 children who were at high risk for ASD, as per the outcome measure of having an older sibling with now confirmed ASD, and 33 typically developing children with no family history of neuro developmental disorders. Many children were tested more than once and ranged in age from 6 to 24 months and the testing sessions comprised infants[10]. Due to rarity of these methods, it extracts important EEG features that have the potential to diagnose ASD, further improving the model's discriminative powers between ASD specific neural connectivity patterns. Basically, the most popular algorithms used to classify the EEG data to diagnose ASD are SVM and CNN. They also pointed out that CNNs have been popular due to their ability to automatically extract features and recognize patterns in a system, which is a plus for learning in ASD research.

Recent years have also witnessed a surge of activities related to the multi modal data fusion in the medical domain and this technique is not only used for the diagnosis of ASD, but is also used for disease diagnosis such as Parkinson, Alzheimer, and Depression. In the context of ASD, EEG data is expanded by other data points that are eye-tracking data, body movement and metrics, and behavior in order develop a complex diagnostic model. This fusion is helpful in improving diagnostic accuracy and complementing EEG findings by including neural and behavioral signals associated with ASD; capturing features which are hard to observe from EEG signals only In recent years, multi modal fusion has garnered much interest especially in SEO application and extends to the diagnosis of ASD [14], [15] as well as other diseases, such as Parkinson [16] , Alzheimer [17] and Depression [18] multi modal data fusion is considered, integrating other sources like eye-tracking or behavioral metrics to enhance diagnostic accuracy. Evaluation tools include accuracy, precision, recall and F1-score; cross checks by validation and test data confirms the models ability to generalize. Table 2. below shows the summarize findings of the studies.

Table 2. Key Studies and Findings

Ref.	Datasets	Data Source	Preprocessing	Methods	Main Findings	Applications
.[1]	EEG data from children with ASD and control group	Public dataset	Noise removal, signal normalization	ML techniques, including SVM	ML effectively classified ASD children with high accuracy.	Early detection of ASD, clinical diagnosis support.
[2]	EEG data from children with ASD	Publicly available dataset	Signal filtering, feature extraction	DL (CNN, RNN)	High classification accuracy for ASD.	Autism Detection, clinical tools.
[3]	Resting-state EEG data from young adults with ASD	Research Database	Artifact removal, signal segmenting	Microstate Analysis	Distinct microstate patterns observed in ASD participants.	Diagnostic biomarkers, ASD characterization in adults.
[4]	EEG data from infants at risk for ASD	Infancy database	Data normalization, feature extraction	Data-driven analytics, ML	Early detection of ASD using EEG.	Early identification and intervention in ASD.
[5]	Time-frequency spectrogram images of EEG	Public EEG dataset	Spectrogram transformation	CNN with time-frequency images	Successful detection of ASD with high accuracy.	Early ASD diagnosis, clinical tools.
[6]	Neuroimaging datasets from psychiatric disorders	Public neuroimaging datasets	Data standardization, neuroimaging preprocessing	Deep learning (CNNs, RNNs)	Classified psychiatric disorders including ASD.	Psychiatric disorder classification, ASD detection.
[19]	EEG data from infants with ASD	Infant database	Nonlinear signal processing, ERP analysis	Nonlinear EEG measures, prediction models	Language-related EEG predicted early ASD diagnosis.	Early ASD prediction, language-related disorder detection.
[7]	EEG and behavioral data	Combined EEG and behavioral dataset	Signal preprocessing, artifact removal	Multimodal approach	Improved ASD detection accuracy.	Multimodal ASD diagnosis, enhanced accuracy.
[8]	EEG data from children with ASD	Public EEG Dataset	Preprocessing with feature extraction	Hybrid deep learning model (CNN + RNN)	Strong ASD detection performance.	Real-time ASD detection, mobile health.
[11]	Spectrogram images from EEG ASD data	Public EEG dataset	Spectrogram transformation, feature extraction	CNN on spectrogram images	High accuracy ASD detection from spectrograms.	Early ASD detection, clinical tools.

The review also establishes a clear inclusion and exclusion criterion for the studies considered. The included studies had to meet the following conditions:

a) The sample involved participants with clinically diagnosed ASD, including conditions such as infantile autism, ASD, PDD-NOS, or Asperger's Syndrome (AS), with or without intellectual disabilities.

b) The analysis involved movement features, for example, eye- tracking or body movements.

In the studies, data sources such as qualitative behavioral data, scores for traditional assessment tools, parental reports, medical/genetic data, and vocal patterns were not considered. Excluded were papers that examined only the effects of rehabilitation, addressed one or several behaviors, or used ML models considering biomarkers obtained while performing tasks that presuppose prior skills. Table 3 and Table 4 shows the summary of the discussion in this section.

Therefore, to improve the reliability of the developed models, the following performance measures are applied: accuracy, precision, recall, and F1 score. These scenarios are reiterated during cross-validation and testing on unseen data sets to reinforce the business case for generalization. The last objective is therefore to replicate these findings in real-life clinical settings to evaluate the potential of this methodology in the diagnosis of early signs of ASD.

Table 3. Comparison of EEG-Based Techniques and Machine Learning Models for ASD Detection

Ref	Preprocessing	Techniques	Datasets	Sampling Value	Precision	Attributes
.[19]	Yes	Support Vector Machine and Correlation	EEG dataset for ages 6 and 12-month	(40 Non ASD) and (14 ASD)	100%	Person correlation in an SVM classifier
[1]	Yes	RF, SVM, KNN, Naïve Bayes for hybrid fusion	Fixation on eyes and facial expression	80 (40 Non ASD and 40 ASD)	87.50% (hybrid), 83.75% (EEG + RF)	Combining behavioral and physiological characteristics
[3]	Yes	Mann Whitney Comparative	EEG microstate analysis tests	Not specified (ASD and controls)	Unspecified	Maps examined for GEV and TP
[2]	Yes	Convolution Layers (CNN)	EEG datasets (Normal, Autism)	20 datasets	80%	CNN with 6 layers
[7]	Yes	MMSDAE, SDAE models	EEG and ET data	Not provided	95.56%	4-layer SDAE models
[11]	Yes	SVM	Spectrogram dataset	Not specified	95.25%	Linear SVM with cross-validation
[20]	Yes	DWT + Shannon entropy, ANN	KSA dataset	Not specified	99.7%	10-fold cross-validation
[5]	Yes	SVM	EEG spectrogram	Not specified	95.25%	SVM with 10-fold validation, TCENTRIST features
[4]	Yes	SVM	EEG signals	100 samples	95%	ASD vs LRC-infant EEG signal classification

Table 4. Summary of Methods and Techniques Used for EEG-Based ASD Detection with Results

Ref.	Method Used in Papers	Technique Name	Results
[19]	Feature selection combined with SVM classifier on EEG data	Prediction of nonlinear measures of language related EEG	Achieved 100% accuracy for infants at 6 and 12 months
[1]	EPOC neuroheadset, FFT for EEG features, CNN for facial expression, K-means for eye fixation analysis	Application of ML to detect Children's Autism	87.50% (Hybrid Fusion), 83.75% (EEG with RF)
[3]	Two-step spatial clustering using Mann-Whitney-Wilcoxon tests	EEG Microstates Analysis in Young Adults With Resting-State	Not specified
[2]	Confusion Matrix for classification	ASD classification on EEG signal using deep learning algorithm	Maximum 80% accuracy
[7]	Multimodal stacked denoising autoencoder (MMSDAE) combining EEG and eye tracking data	A Multimodal Approach for Identifying Autism Spectrum Disorders in Children	Achieved 95.56% accuracy, 92.5% sensitivity
[11]	ML (SVM, RF, KNN) and DL (CNN) on EEG spectrogram images	A spectrogram image-based intelligent technique for automatic detection of ASD from EEG	SVM achieved 95.25% accuracy
[20]	Discrete Wavelet Transform with statistical features and entropy for classification using ANN	EEG-Based Computer Aided Diagnosis of ASD Using Wavelet, Entropy, and ANN	Best accuracy of 99.8% using overlapping segments
[5]	Time-frequency spectrogram generation, TCENTRIST and PCA feature extraction, SVM classification	Diagnosis of ASD using a time–frequency spectrogram image-based approach	Best accuracy: 90.95%
[4]	Nonlinear analysis of EEG signals as early predictors of ASD	EEG Analytics for Early Detection of ASD	Accuracy achieved through LRC: 95.12%

3 Gap Analysis

3.1 Combined Analysis of EEG and Behavioral Data

Many works have investigated how EEG data can be taken together with behavioral data in [1]], [7], [20]. However, there is generally a lack of approach for merging several modalities [1], which used a weight naive Bayes approach to fusion while [7] used a stacked de-noising auto encoders (SDAE) for feature extraction. But such methods do not give consistent enhancements in accuracy of classification, and there is no unique way to fuse the classifiers.

Subsequent studies should aim at establishing more stable, single systems that combine EEG with other techniques of eye-tracking, facial conducted and any other sign of behavior. One should also pay attention to the time and context to support these modalities, which include the diagnostic performance. Developing these approaches more could enhance stronger outcome across those section with variably distinct population and settings, thus a better diagnostic for ASD.

3.2 Small Sample Sizes

Evaluations of the diagnostic tools and methods in ASD research also show that many of the current research projects were conducted on small samples of subjects [19], [2], [20]. Small sample sizes decrease testing capacity and generalization of study results to different samples, especially those with increased variability. Nonetheless, unlike [4] which has a larger sample size (n = 99), it restricts the age range and does not include members of different ethnic backgrounds or from different US regions, so its conclusions cannot be quite generalizable.

One of the apparent deficiencies in this area is the lack of large-scale multi center studies. These studies should collect data from more people, including almost all aged, ethnicity, and comorbid patients, so as to generalize the results. Investigation with groups of subjects at various developmental stages would also helps to advance understanding of how specific biomarkers change over time and help improve the reliability of the ASD diagnostic criteria.

3.3 Early Diagnosis and ongoing monitoring

While there has been progress in detecting ASD at very early stages, proving that EEG biomarkers could identify ASD with 3 months old, the majority of other studies concerns children older than 1 year, or even adults; more importantly, none of the objective methods mentioned above are designed to monitor diagnostic performance over time. Many of the currently proposed models for detecting ASD in the early years fail to capture information on symptom development and decline over time, which is an important area of restriction for their application in clinical practice.

There is a pressing need for more research into predictive modeling that focuses on the early stages of ASD and the potential for tracking developmental changes as children grow. By leveraging EEG and multi modal data, it is possible to identify key biomarkers that can aid in early-stage diagnosis, starting from infancy. Long-term, longitudinal studies will be essential for understanding how these biomarkers evolve, and whether early intervention based on these predictions can lead to improved outcomes for children with ASD.

3.4 Normalization of Feature Extraction and Signal Processing Techniques

A key gap in the current research on ASD diagnosis lies in the variety of feature extraction techniques used across studies. For instance, some studies use wavelet transform [20], while others employ microstate analysis [3] or extract textural features from spectrograms [11], [5]. This lack of consensus on which features are most effective in distinguishing ASD from neuro typical development creates significant challenges in comparing results across studies and replicating findings.

To overcome this gap, there is an urgent need for standardized protocols in feature extraction and signal processing. Identifying a set of robust, reproducible EEG features that consistently correlate with ASD across different studies and populations would be a major advancement. The development of automated preprocessing pipelines would increase the reproducibility of results and help establish common benchmarks in the field, ensuring more reliable comparisons between studies.

3.5 Clinical Testing and Real-World Usage

Many studies report high accuracy rates in controlled environments, yet the real-world applicability of these diagnostic models remains underexplored. For example, while [11] reports an impressive 99.15% accuracy with deep learning models, this result is based on relatively clean, preprocessed data in a controlled setting. Few studies [5], have validated these models in clinical settings or with more diverse, real world data, highlighting a critical gap in the clinical relevance of these findings.

There is an urgent need for clinical validation of these diagnostic models in real-world settings. Future studies must test models in hospitals, clinics, and community health environments to assess their ability to handle the inherent variability and noise present in real-world data. Additionally, these models should be designed to be clinically interpretable, actionable, and easy to integrate into existing diagnostic workflows, to ensure that they can be effectively used by healthcare professionals.

3.6 Scalable Solutions and Real-Time implementation

A significant gap exists in the scalability and real time implementation of diagnostic models, as many studies focus primarily on classification accuracy and feature extraction without considering the practical deployment of these models in clinical practice. For example [5] employs a support vector machine classifier, but such models may not be easily scalable for widespread use or adaptable to real-time diagnostic systems, which are essential for early detection.

Research must focus on developing scalable, real-time diagnostic tools capable of processing large volumes of data from various sources, such as multiple sensors or clinics. These systems should provide fast, accurate results in a way that is both practical and cost-effective. Furthermore, the development of portable diagnostic systems, such as wearable EEG devices or mobile applications, could offer significant advantages in terms of accessibility and immediate utility for clinicians and families, facilitating earlier intervention and more frequent monitoring.

3.7 Durability and Explainability of ML Models

While DL models [11] often demonstrate high classification accuracy, a major gap remains in their interpretability, which is a critical issue for their adoption in clinical settings. Clinicians require models that not only perform well but also provide transparent, interpretable results that can be used to understand the reasoning behind a diagnosis. For instance, [3] and [7] use complex algorithms like DL and Naive Bayes, but they do not clearly identify which features are most important for making a diagnosis.

To address this gap, future research should prioritize the development of interpretable AI models. Techniques such as explainable deep learning or feature importance analysis should be integrated into these models to help clinicians understand how decisions are made. By improving the transparency of model outputs, trust in these systems can be fostered, making it easier to integrate them into clinical practice and ensuring they are used appropriately to inform diagnoses and treatment decisions.

4 Existing Datasets

The datasets existing from the reviewed papers concern EEG recording from both ASD and neuro typ-ical controls across varied age range. EEG is multi-channel data that required preprocessing and filtering off artifacts for better analysis to be carried out. The sample size has been variable, from as low as 30 par-ticipants to several hundred participants, but often with the ASD sample and control sample matched. Other studies use other inputs that include eye tracking, facial expressions in addition to the EEG signals. The 6-month and 12-month EEG dataset in [19] included 54 subjects (40 non-ASD, 14 ASD) and achieved 100% accuracy for both age groups, using SVM with correlation for feature selection. In [1] , the EEG, eye fixation, and facial expression dataset (80 subjects) achieved 87.5% accuracy using a hybrid fusion model combining random forest and weighted Naïve Bayes.

The EEG microstate dataset in [3] analyzed six template maps to compare ASD and control groups, but no sample size or accuracy was specified. In [2], the EEG signals dataset (20 samples) achieved 80% accuracy using a 6-layer CNN. In [7], the EEG and eye-tracking (ET) dataset (sample size unspecified) achieved 95.56% accuracy using the MMSDAE model. EEG spectrogram data in [8] achieved 95.25% accu-racy using a linear (SVM) with ten-fold cross-validation. The KSA dataset in [21] achieved 99.7% accuracy using DWT, Shannon entropy, and an ANN classifier with 10-fold cross-validation. The EEG spectrogram dataset in [5] also achieved 95.25% accuracy using SVM and TCENTRIST features. The EEG signals dataset in [4] achieved 95% accuracy with SVM for classifying ASD. Many datasets are publicly available, though some are proprietary. Preprocessing methods, such as filtering and segmentation, transform EG data into formats suitable for machine learning models. While these datasets are valuable, there is still a need for larger, more diverse datasets to improve model accuracy and clinical use.

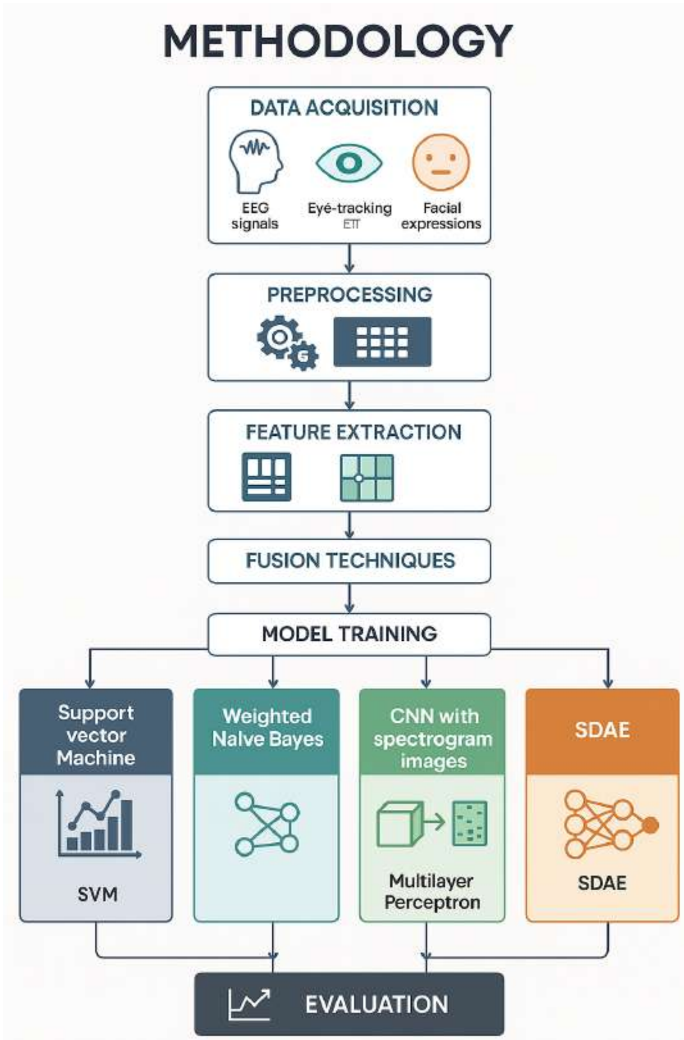

Figure 1. Methodology Architecture

5 Methodology

Research using EEG data to differentiate between ASD and non-ASD children has relied on different methods of ML, of which SVM has been widely used. For example, one of the recently published papers identified a number of papers that examined EEG data obtained from children at the age of 6 and 12 months, and employed SVM to classify ASD with high accuracy taking into consideration the age-dependent features of EEG spectra [19]. In another approach, a weighted Naive Bayes algorithm was used to improve the performance of detecting ASD by integrating EEG data with the eye movement and facial expressions data. This fusion of multimodal led to 87.5% accuracy. New hybrid approaches that merge EEG with other signs, for instance, eye-tracking or facial expressions enhanced the diagnostic cor-rectness too [1]. One such study combined these data sources using a weighted Naive Bayes algorithm, achieving significant accuracy improvements. Other study Microstate analysis is performed using only Cartool [3].1 illustrated the methodology flow of work.

In the present study, various feature representations have been proposed to classify EEG features with high precision, and in particular, multilayer perceptron networks. Certain works created spectrogram im-ages of EEG signals, the identification of which was made using the CNN, and the accuracy reached 99.15%. The use of multiple modalities as EEG and eye tracking has been found to provide better perfor-mance in ASD classification than using each modality separately, this is by the use of a two-step feature learning and fusion based on a deep learning method known as SDAE. In the first step, two SDAE models are proposed for feature extraction for EEG and ET modality, respectively. Subsequently, a third SDAE model in the second step is proposed for multimodal fusion of the learned EEG and ET features in a concatenated manner [7]. Table 5. below shows the comparision of the methods used for ASD detection.

Table 5. Performance Comparison of EEG-Based Methods for ASD Detection

Ref.	Name of Methods	Results	Attribute
[19]	Correlation + SVM	100% for both ages	SVM classifier with Pearson correlation
[1]	SVM, KNN, Weighted Naïve Bayes for hybrid fusion	87.50% (Hybrid Fusion), 83.75% (EEG with RF)	Hybrid fusion combining behavioral and physiological features using weighted Naïve Bayes
[3]	Mann Whitney Wilcoxon Tests for Comparison	82.43%	Global Explained Variance (GEV) and temporal parameters
[2]	CNN	80%	6-layer Convolutional Neural Network (CNN)
[7]	MMSDAE	95.56%	4-layer Stacked Denoising Autoencoder (SDAE) models
[11]	SVM	95.25%	Linear SVM with k-fold cross-validation
[20]	DWT + Shannon Entropy, ANN	99.7%	10-fold cross-validation with Discrete Wavelet Transform and entropy-based features
[5]	SVM	95.25%	SVM with 10-fold cross-validation using TCENTRIST features

6 Results Discussion of the studies

Different studies have shown the efforts of different machine learning algorithms to classify ASD from EEG data. Each study has its unique value as well as methods employed. These studies have shown nearly perfect results when predicting the SVM accuracy for the control group of infants at 6 and 12 months. High results were attained in other SVM applications too. With increase in dataset volumes, the accuracy grew lower which indicates the competency of SVMs but highlights the SVM's boundaries at large datasets due to its high complexity and variability. When employing extra data streams, the Bayesian network approach worked exceedingly well (EEG + eye movement + facial expression) which added novel forms, behavioral data enhanced diagnostic accuracy to 87.5% with the combination of data streams. But the integration of these signals separately provided ample improvements. Deep learning methods have adopted the mapping of the EEG spectrogram images using convolution neural networks for analysis in which the accuracy estimates range from 99.15%. These methods greatly outperformed the simpler machine learning techniques, proving their unmatched prowess for recognizing complex patterns in EEG.

It highlights the contribution of deep learning in this methodology for the diagnosis of ASD from EEG data. Another new multi modal approach that achieved high accuracy for ASD classification with competition included EEG and eye tracking data in Stacked De-noising Auto encoders (SDAE). These methods utilize the additive diagnostic value of different physiological modalities to increase the accuracy of the diagnosis when different data modalities are used. Algorithms for feature extraction based on DWT and entropy that separate the signal into sub bands like differential asymmetry index, left IV (NI), have recently come to the fore. These signal processing methods together with ANNs provided a good accuracy and specificity which means that advanced signal processing techniques may be useful in tasks of differentiation between ASD and non ASD patients. The last part to be analyzed were methods that are based on pictures of spectrograms, where the EEG data is visually represented after it has been subjected to short-time Fourier transform. This can also be applied with other machine learning algorithms for diagnostics features. In addition, it provided a classification with high accuracy and a new way of interpreting the data which has the potential to improve the diagnosis of patients with ASD. So, the results demonstrated that deep learning and multimodal approaches are more accurate compared to other conventional techniques. It is worth noting that there are still some pending issues and many hurdles are still to be crossed. The variability in quality and size of the datasets used will certainly impact the models' performance. In terms of practicality, it is clear that in well-structured clinical environments where instantaneous accurate decisions are critical, the high number of data types and the requirements of deep learning will pose challenges in actual practice. These results give, however, do provide some level of comfort for the possibility of more advanced meta learning paradigms even in the face of obstacles. Table 6. shows the some pros and cons according to the investigation of the studies discussed in previous sections.

Table 6. Pros and Cons of different discussed studies.

Ref.	Pros	Cons
[19]	100% early ASD detection accuracy for infants at particular ages was attained.	Less accuracy, indicating a less reliable model, in a sample with a larger variety of behaviors.
[1]	Uses a multimodal approach that integrates facial expressions, EEG, and ocular fixation to achieve greater accuracy.	Requires a lot of data processing and sophisticated technology, which might be difficult in real-world applications.
[3]	Microstate study shows important brain dynamics differences between ASD and neurotypicals.	Generalizability to larger ASD populations is limited by the small sample size.
[2]	Clear insights into categorization accuracy and overall model efficacy are provided by confusion matrix analysis.	Limited data on the variety of datasets and a lack of cross-demographic validation.
[7]	Comprehensive ASD detection with high accuracy (95.56%) using multimodal EEG and eye-tracking data fusion.	Requires sophisticated sensors, which limits its applicability in non-specialized settings.
[11]	Surpasses earlier techniques for ASD identification by achieving high accuracy with CNN-based deep learning.	ML models have comparatively reduced specificity, suggesting the possibility of false positives in bigger datasets.
[20]	Uses DWT and entropy functions with ANN to provide excellent classification accuracy (up to 99.8%).	Complex data processing limits real-time application.
[5]	Time-frequency spectrograms with high sensitivity (97.07%) and accuracy (95.25%) enhance the categorization of ASD.	The 16-child sample size may limit the generalizability to bigger groups.
[4]	Early intervention is supported by nonlinear EEG analysis, which successfully identifies early indicators of ASD.	The decline in predictive accuracy over time raises the possibility that existing techniques need to be improved for wider use.

7 Conclusion

Integrating ML into EEG analysis could dramatically impact therapeutic stratification and prognostic clas-sification for a broad spectrum of children suffering with ASD. From all the studies reviewed, we demon-strate that combining ML/DL techniques like SVM and deep learning with our stage 1 review EEG data can achieve high accuracy, particularly with the use of multimodal methods. Nevertheless, hurdles remain like the need for larger and more diverse data sets, better clinical application of deep learning techniques, and increased model interpretability. These obstacles are critical for developing practical implementation of the research into clinical practice. Continued refinement of the method will expand its acceptability to other populations. Ultimately, overcoming these barriers will enhance the preciseness and timeliness of ASD diagnosis and treatment, which are crucial for positive health results.

Author Contributions

Saad Akbar: Project Supervision, Conceptualization, Methodology **Shahid Munir Shah**: Data curation, Writing draft preparation. **Rubaishe Tahir**: Visualization, Investigation. **Muhammad Mujtaba Khan**: Validation, Software, Editing.: **Nuha Aamir**: Writing- Reviewing and Editing.

Compliance with Ethical Standards

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Liao, H. Duan, and G. Wang, "Application of Machine Learning Techniques to Detect the Children with Autism Spectrum Disorder," J. Healthc. Eng., vol. 2022, pp. 1–10, Mar. 2022, doi: 10.1155/2022/9340027.

[2] N. A. Ali, "Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm," IAES Int. J. Artif. Intell. IJ-AI, vol. 9, no. 1, p. 91, Mar. 2020, doi: 10.11591/ijai.v9.i1.pp91-99.

[3] D. F. D'Croz-Baron, M. Baker, C. M. Michel, and T. Karp, "EEG Microstates Analysis in Young Adults With Autism Spectrum Disorder During Resting-State," Front. Hum. Neurosci., vol. 13, p. 173, Jun. 2019, doi: 10.3389/fnhum.2019.00173.

[4] W. J. Bosl, H. Tager-Flusberg, and C. A. Nelson, "EEG Analytics for Early Detection of Autism Spectrum Disorder: A data-driven approach," Sci. Rep., vol. 8, no. 1, p. 6828, May 2018, doi: 10.1038/s41598-018-24318-x.

[5] M. N. A. Tawhid, S. Siuly, and H. Wang, "Diagnosis of autism spectrum disorder from EEG using a time–frequency spectrogram image-based approach," Electron. Lett., vol. 56, no. 25, pp. 1372–1375, Dec. 2020, doi: 10.1049/el.2020.2646.

[6] M. Quaak, L. van de Mortel, R. M. Thomas, and G. van Wingen, "Deep learning applications for the classification of psychiatric disorders using neuroimaging data: Systematic review and meta-analysis," NeuroImage Clin., vol. 30, p. 102584, Jan. 2021, doi: 10.1016/j.nicl.2021.102584.

[7] J. Han, G. Jiang, G. Ouyang, and X. Li, "A Multimodal Approach for Identifying Autism Spectrum Disorders in Children," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30, pp. 2003–2011, 2022, doi: 10.1109/TNSRE.2022.3192431.

[8] "Automated ASD detection using hybrid deep lightweight features extracted from EEG signals - ScienceDirect." Accessed: Sep. 26, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0010482521003425

[9] "Detection of Abnormalities for Diagnosing of Children with Autism Disorders Using of Quantitative Electroencephalography Analysis | Journal of Medical Systems." Accessed: Sep. 26, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s10916-010-9560-6

[10] "EEG complexity as a biomarker for autism spectrum disorder risk | BMC Medicine | Full Text." Accessed: Sep. 26, 2024. [Online]. Available: https://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-9-18

[11] Md. N. A. Tawhid, S. Siuly, H. Wang, F. Whittaker, K. Wang, and Y. Zhang, "A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from EEG," PLOS ONE, vol. 16, no. 6, p. e0253094, Jun. 2021, doi: 10.1371/journal.pone.0253094.

[12] "Nonlinear analysis of the sleep EEG in children with pervasive developmental disorder - PubMed." Accessed: Oct. 29, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18766147/

[13] X. Sun et al., “Prevalence of autism in mainland China, Hong Kong and Taiwan: a systematic review and meta-analysis,” Mol. Autism, vol. 4, no. 1, p. 7, Apr. 2013, doi: 10.1186/2040-2392-4-7.

[14] L. E. Mash et al., “Atypical Relationships Between Spontaneous EEG and fMRI Activity in Autism,” Brain Connect., vol. 10, no. 1, pp. 18–28, Feb. 2020, doi: 10.1089/brain.2019.0693.

[15] B. Cociu et al., “Multimodal Functional and Structural Brain Connectivity Analysis in Autism: A Preliminary Integrated Approach with EEG, fMRI and DTI,” IEEE Trans. Cogn. Dev. Syst., vol. PP, pp. 1–1, Mar. 2017, doi: 10.1109/TCDS.2017.2680408.

[16] J. C. Vasquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Noth, “Multimodal Assessment of Parkinson’s Disease: A Deep Learning Approach,” IEEE J. Biomed. Health Inform., vol. 23, no. 4, pp. 1618–1630, Jul. 2019, doi: 10.1109/JBHI.2018.2866873.

[17] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, “Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer’s Disease,” IEEE J. Biomed. Health Inform., vol. 22, no. 1, pp. 173–183, Jan. 2018, doi: 10.1109/JBHI.2017.2655720.

[18] H. Dibeklioglu, Z. Hammal, and J. F. Cohn, “Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding,” IEEE J. Biomed. Health Inform., vol. 22, no. 2, pp. 525–536, Mar. 2018, doi: 10.1109/JBHI.2017.2676878.

[19] F. C. Peck, L. J. Gabard-Durnam, C. L. Wilkinson, W. Bosl, H. Tager-Flusberg, and C. A. Nelson, “Prediction of autism spectrum disorder diagnosis using nonlinear measures of language-related EEG at 6 and 12 months,” J. Neurodev. Disord., vol. 13, no. 1, p. 57, Dec. 2021, doi: 10.1186/s11689-021-09405-x.

[20] R. Djemal, K. AlSharabi, S. Ibrahim, and A. Alsuwailem, “EEG-Based Computer Aided Diagnosis of Autism Spectrum Disorder Using Wavelet, Entropy, and ANN,” BioMed Res. Int., vol. 2017, no. 1, p. 9816591, 2017, doi: 10.1155/2017/9816591.

[21] “Autism.” Accessed: Sep. 26, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders