Machine Learning with Neuroimaging Data to Identify Autism Spectrum Disorder: A Systematic Review and Meta-Analysis.

Running Title: ML with Neuroimaging data in ASD classification

Authors: Da-Yea Song[1*], Constantin-Cristian Topriceanu[2*], Denis C. Ilie-Ablachim[3], Maria Kinali[4], Sotirios Bisdas[5,6]

*The two authors share co-first authorship.

Author Affiliations:

1. Institute of Neurology, University College London, London, UK
2. University College London Medical School, University College London, London, UK
3. Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Bucharest, Romania
4. Department of Paediatric Neurology, The Portland Hospital, London, UK
5. Lysholm Department of Neuroradiology, The National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Foundation Trust, London, United Kingdom
6. Department of Brain Repair & Rehabilitation, Queen Square Institute of Neurology, University College London, London, UK

Corresponding author:

Sotirios Bisdas

Department of Neuroradiology, 8-11 Queen Square, Box 65, London WC1N 3BG, United Kingdom

Tel : 0044 20 344 83148

Email : s.bisdas@ucl.ac.uk

ABSTRACT

Purpose: Autistic spectrum disorder (ASD) is diagnosed through observation or interview assessments, which is time-consuming, subjective, and with questionable validity and reliability. We aim to evaluate the role of machine learning (ML) with neuroimaging data to provide a reliable classification of ASD.

Methods: A systematic search of PubMed, Scopus, and Embase was conducted to identify relevant publications. Quality Assessment of Diagnostic Accuracy Studies-2 was used to assess studies' quality. A bivariate random-effects model meta-analysis was performed to evaluate the pooled sensitivity, specificity, and diagnostic performance through the hierarchical summary receiver operating characteristic (HSROC) curve. Meta-regression was also implemented.

Results: 44 studies (total of 5697 ASD and 6013 typically developing individuals) were included in the quantitative analysis. The pooled sensitivity for differentiating diagnostic groups was 86.25 95% confidence interval [CI] (81.24, 90.08), while the pooled specificity was 83.31 95% CI (78.12, 87.48**)** with a combined area under the HSROC (AUC) of 0.889. Higgins I^2 (>90%) and Cochran's Q ($p<0.0001$) suggest a high degree of heterogeneity. In the bivariate model meta-regression, a higher pooled specificity was observed in studies not using a brain atlas (90.91 95% CI [80.67, 96.00], p=**0.032**). Greater pooled sensitivity was seen in studies recruiting both males and female (89.04 95% CI [83.84, 92.72**],** p=**0.021**), and combining imaging modalities (94.12 95% [85.43, 97.76], p=**0.036**).

Conclusions: ML with neuroimaging data is an exciting prospect in detecting individuals with ASD but still requires more studies to optimize and improve reliability for usage in clinical practices.

Keywords: autism spectrum disorder; machine learning; neuroimaging; systematic review and meta-analysis

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by impairments in social communication with restricted or repetitive patterns of behaviors, interests, or activities and hyper- or hyposensitivity to sensory stimuli [1]. Having emerged as a public health issue [2], ASD is one of the fastest-growing developmental disabilities. Previous studies have reported on the presence of early ASD symptoms in 12-month infants [3], though abnormalities in development may become apparent as early as six months of life [4]. However, as the average age for diagnosis is around three to five years of age, there is a wide gap in children not receiving the appropriate resources [5]. The timing of ASD detection is crucial as findings emphasize the importance of early comprehensive interventions for the improved long-term outcome of children diagnosed with ASD [6].

ASD is typically considered a complex childhood condition with no single etiology. The current gold standard of diagnosing ASD is strictly based on lengthy behavioral observations by trained experts or caregiver interviews [7]. Despite efforts to standardize the assessment tools, the reliability and validity of the results are questionable when considering the administrators' subjectivity in ratings due to differences in training and background expertise [8][9]. Moreover, only 8% of pediatric specialists are capable of administering routine evaluations [10], and lacking easy access to such tools due to geographic or cultural adaptability is another critical limitation [11]. As a result, there is a high demand for a more objective approach that could provide quick and accurate detection of ASD.

For this reason, the search for an ASD-specific biomarker using neuroimaging data has been actively ongoing. Many studies have looked at anatomical differences in volume, cortical thickness, and surface areas in regions of interest [12]. Moving forward, more recent studies have looked at diffusion tensor imaging (DTI) [13] and functional MRI (fMRI) imaging [14]. Nonetheless, due to inconsistent findings between papers and lack of replication studies, a reliable biomarker for ASD remains elusive [15]. Reasons for this include previous studies having small sample sizes or recruiting participants that do not represent the generalized population. Additionally, with high heterogeneity in individuals with ASD, neural abnormalities related to the spectrum are likely to result from widespread connectivity networks rather than a single brain

region [16]. Therefore, a comprehensive procedure in analyzing a vast amount of neuroimaging data with large datasets is needed to overcome these limitations.

One promising and innovative approach is using machine learning (ML) with neuroimaging data. Being a subfield within artificial intelligence (AI), ML analyzes the input data patterns by extracting features and constructing the best fitting algorithm to make sense of the dataset with or without pre-existing knowledge [17]. Once classifiers are trained, the algorithm can be used on a new set of data and produce outcome labels [18]. Building complex models by processing large amounts of information, ML can surpass human performance in recognition of symptoms, early diagnosis, and prediction of prognosis while at the same time reducing human error [19]. Similarly, with the recent establishment of large neuroimaging database repositories for ASD, a growing number of studies are trying to combine ML with neural biomarkers to establish an objective and data-driven method in identifying individuals with ASD. If successful, it will reduce subjectivity and increase the diagnosis process's reproducibility by contributing to the current understanding of ASD etiology [20].

The study aimed to explore whether ML with neuroimaging data is reliable enough to distinguish individuals with ASD from their neurotypically developing (TD) peers through a systematic review and quantify its classification performance through a random-effects bivariate sensitivity-specificity meta-analysis. The literature search has identified a similar meta-analysis [21], but the novelty of our manuscript stems from the comprehensive meta-regression, which aimed to characterize ML protocol considerations in improving the sensitivity and specificity in the bivariate model.

METHODS AND MATERIALS

Search Process

The review was conducted to fulfill the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) criteria on published peer-reviewed journal articles [22]. Two reviewers (DYS and CT) independently conducted all the steps, and the inter-observer reliability was quantified using Cohen's kappa (κ=0.877, $p<0.001$). In cases of discrepancy, a senior reviewer (SB) provided input for resolution.

The review question was: “Can ML learning with neuroimaging data be employed to distinguish ASD from non-ASD individuals?”. As we aimed to evaluate the prospect of using ML with neuroimaging data to classify ASD, we included in this review studies which enrolled participants of all age-groups.

Thus, a systematic search of Embase, PubMed, and Scopus was used to identify relevant publications published between 01/01/2010 and 15/09/2020. Search items were defined using the PICO (Patient/Intervention/Comparator/Outcomes) framework: (P)= (“autism spectrum disorder”, “autism”, “ASD”, “Asperger’s syndrome”); (I)=(“MRI”, “MR”, “DTI”, “DWI”,“CT”, “PET”, “PET-CT”, “SPECT”, “EEG”); (C)=(“machine learning”, “artificial intelligence”, “deep learning”); (O)= (“diagnosis”, “screening”, “identification”, “classification”). The PICO framework categories were combined using “AND”, while we grouped the variations within categories via “OR”. Reference lists of included articles were also reviewed to identify further eligible publications.

Inclusion and exclusion criteria

Inclusion criteria were peer-reviewed, English-written manuscripts available online through electronic indexing fulfilling the following criteria: (1) ML algorithms with neuroimaging data were used as an index test to distinguish ASD from TD individuals, (2) ASD was diagnosed using internationally accepted criteria (DSM-V or ICD 10) or gold-standard diagnostic evaluations (ADOS, ADI-R) as the reference standard, (3) participants received both the index and the reference standard tests with blinding of the assessors. Any studies were excluded if they 1) did not receive a final diagnosis (e.g., considered high-risk of ASD based on positive family history); 2) participants with ASD were grouped with other developmental disorders (e.g., individuals with ASD and individuals with ADHD considered together as the ‘clinical’ group); 3) used ML only to predict symptom severity; 4) combined neuroimaging data with other behavioral or clinical information; or 5) any non-original research (i.e., reviews, lecture notes, book chapters, and conference abstracts).

Data Extraction

Reviewers (DYS, CT) independently extracted information based on participant characteristics (number of individuals within each diagnostic group, age, sex, method of diagnostic

confirmation), neuroimaging data (modality, use of a database, pre-processing software, atlas), ML (algorithms, cross-validation), and accuracy results (sensitivity, specificity, positive predictive value, negative predictive value). In addition to screening supplementary materials for data, raw cell values were used to calculate psychometric properties when not provided in the studies.

Quality Assessments

Quality of the systematic review was assured by limiting the risk of bias and applicability concerns in line with the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) questionnaire [23]. Patient selection, conduct, and interpretation of ASD status classification based on ML algorithms and clinical diagnostic tests were analyzed to meet the review question and avoid bias introduction.

Statistical Methods

Studies reporting enough variables to allow the calculation of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were included in the quantitative analysis. Statistical analyses were performed using R version-3.6.2 (packages “meta” and “mada”) with $p<0.05$ considered to be of significance.

The univariate random-effects model was employed to calculate the pooled specificity and sensitivity according to the diagnostic test accuracy meta-analysis methodology [24]. As the specificity and sensitivity are inter-dependent, a bivariate random-effects model meta-analysis was used to generate a hierarchical summary receiver operating characteristic (HSROC) curve with its 95% confidence region [25]. We also calculated the area under the ROC (AUC) curve.

Furthermore, we measured the pooled diagnostic odds ratio (DOR), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and their corresponding 95% confidence intervals (CI). DOR represents the odds of having a positive index test in ASD compared to healthy control participants **(Supplementary Equation 1**). PLR/NLR describes how many times more/less likely is the index test results to be positive/negative in ASD vs TD individuals (**Supplementary Equations 2 and 3**).

Inter-study heterogeneity was determined via Cochran’s Q and Higgins I^2 statistics. Cochran’s Q test with a $p<0.05$ or $I^2>50\%$ was interpreted as indicating the presence of heterogeneity [26]. Publication bias was assessed via the visual inspection of contour-enhanced

Funnel and using Egger's test. Asymmetrical funnel plots or Egger's test $p<0.05$ were interpreted as indicating the possibility of publication bias.

We performed the best set meta-analysis, which included the dataset corresponding to each study's maximal accuracy. As many studies reported multiple analyses by altering various parameters, a complete case meta-analysis that included all the analyses reported were also performed. Meta-regression was employed to explore the contributors of heterogeneity in the bivariate model with the following variables used as a moderator: (1) age group of participants (toddlers defined as <2 years, children and adolescents defined as <19 years, adults >20 years or lifetime in cases where studies included participants across age groups); (2) sex of participants (only male included vs both males and females included); (3) diagnostic test for ASD; (4) participant origin (recruited at a research center/hospital vs data from a publicly available database); (5) imaging modality; (6) atlas template (EEG studies were removed from this analysis); (7) ML algorithm; (8) cross-validation (CV) method or equivalent. All meta-regression moderators were treated as categorical variables. Only the best set meta-analysis was used to conduct the meta-regression as the complete case meta-analysis had unequal contributions from the studies.

As small-sized studies suffer from over-fitting, which often leads to higher sensitivity and specificity, we replicated the above meta-analysis while only including studies with over 100 participants.

RESULTS

Search Results

Database searchers revealed 231 articles, of which 84 duplicates were identified and removed. However, we carefully screened non-original research papers and added 26 new studies. Upon screening the titles and abstracts against our research question, 73 manuscripts were thoroughly assessed. The qualitative analysis considered 65 studies after applying the inclusion and exclusion criteria. As some studies lack the data required for the meta-analysis, 44 studies were included in the quantitative analysis with a total of 5697 ASD and 6013 TD individuals [27-70]. **Figure 1** presents a visual representation of the PRISMA flow chart. Characteristics of the included participants are presented in **Table 1**, while **Table 2** summarizes the study details.

Qualitative Assessment

In general, the 44 studies included in the qualitative analysis evaluating the use of ML with neuroimaging data to distinguish ASD from TD individuals were of good methodological quality (**Figure 2**). According to QUADAS-2, the risk of bias was assessed across four domains: patient selection, index test, reference standard, and flow and timing. Regarding patient selection, 26.15% were considered high-risk because several studies only included male participants or chose neuroimaging data that followed specific protocols without providing sufficient justification, and 32.31% were labelled as unclear because they lacked the information to assess the domain. Regarding flow and timing, 9.23% were regarded as unclear. While there were no concerns for the index test, the reference standard domain resulted in 9.23% of the studies being labelled as unclear. The applicability concerns were at low risk across three domains: patient selection, index test, and reference standard.

Meta-analysis

Table 3 summarizes the meta-analysis results. There was a high degree of heterogeneity (I^2=95.90%, p<0.0001) between the studies. When attempting to differentiate ASD from TD participants, the pooled sensitivity was 86.25 95% CI (81.24, 90.08) (**Figure 3**), while the specificity was 83.31 95% CI (78.12, 87.48) (**Figure 4**). The pooled DOR was 19.66 (95% CI 13.40, 28.84), while the AUC was 0.889 (**Figure 5**). The inspection of the color-enhanced funnel plots (**Supplementary Figures 1 and 2)** and the Egger test (p<0.001) are in keeping with the possibility of publication bias When considering only the studies that had over 100 participants, both the sensitivity of 83.23 95% CI (76.79, 88.16) and the specificity of 78.90 95% CI (70.85, 85.19) are significantly lower. The high degree of heterogeneity and the possibility of publication bias remain in this analysis as well.

In the complete case analysis (**Supplementary Table 1**), a high degree of heterogeneity (I^2=96.80, p<0.0001) and the possibility of publication bias (Egger test p<0.0001) were also noted. However, ML's diagnostic performance in differentiating ASD from TD participants was lower: DOR 5.40 (4.87, 5.98) and AUC 0.723.

Meta-regression

Meta-regression results are presented in **Table 4.** A higher pooled sensitivity in the bivariate model was observed in studies recruiting both males and females (89.04 95% CI [83.84, 92.72**],** p=**0.021**), combining multiple imaging modalities (94.12 95% [85.43, 97.76], p=**0.036**) on a backbone of anatomical MRI, DTI or fMRI or using EEG (99.85% 95% CI [37.18, 1.00], p=**0.025**). A higher pooled specificity in the bivariate model was seen in studies not using a brain atlas (90.91 95% CI [80.67, 96.00], p=**0.032**). The bivariate meta-regression found no other moderator variable to be associated with the study heterogeneity.

DISCUSSION

We reviewed published studies on neuroimaging data with ML algorithms as an objective method to detect individuals with ASD. Based on the 44 studies included in the quantitative analyses, the pooled sensitivity and specificity were 86.25 95% CI (81.24, 90.08) and 83.31 95% CI (78.12, 87.48), respectively. However, when considering only large studies which are less prone to overfitting, the sensitivity of 83.23 95% CI [76.79, 88.16] and specificity of 78.90 95% CI (70.85, 85.19) were lower, but they are more likely to reflect the performance in real-life settings. While there is no standard protocol in processing neuroimaging data or implementing ML algorithms, our meta-regression analysis provides insights into improving classification performance. As such, a higher pooled sensitivity was associated with including both males and females, combining structural and functional neuroimaging modalities, or using EEG, while a higher specificity was obtained in studies who did not use a brain atlas.

First, studies that included both male and female participants obtained a better sensitivity in distinguishing ASD from TD individuals. Most epidemiological studies report a 4:1 male:female ratio in ASD [71][72]. Because of the lower prevalence of ASD in females, 11 studies included only male participants. Focusing only on males may bias the results and limit the prospect of generalizing the findings. Furthermore, recent studies highlighted the possibility that females are being underdiagnosed as they are probably exhibiting different clinical features [73][74]. Therefore, including both males and females enables addressing the core neuroimaging features of ASD rather than reporting sex-specific traits.

ASD has been associated with both different structural and functional characteristics, both histologically and on neuroimaging, compared to TD individuals [75]. Thus, it is not surprising to see studies that employed a combination of anatomical MRI plus DTI, or fMRI obtained a higher diagnostic sensitivity in the bivariate model. Hence, we urge that future effort should concentrate on combining various neuroimaging modalities. Similarly, EEG has emerged as having a better sensitivity performance in ASD classification in the meta-regression. It has been postulated that children with ASD have a different brain network topology characterized by decreased long-range coherence (especially between the frontal and occipital lobe) but increased short range connectivity between the frontal and parietal/temporal lobes [76]. Although it seems like a promising lead to follow, none of the included studies had more than 100 participants raising concerns about overfitting.

In large, ML algorithms can be divided into supervised, semi-supervised, unsupervised, and reinforcement learning [77]. Most of the studies used supervised algorithms, with the most common choices being support vector machines (SVM), neural networks (NN), and decision trees (DT). However, supervised learning has certain limitations, including overtraining and overfitting [78]. Thus, to maintain a high classification performance when exposed to new input data, the supervised learning algorithm needs to be constantly re-trained. Additionally, unlike unsupervised/semi-supervised learning, supervised learning can't infer new information as it only specializes in the training dataset. The latter is of paramount importance, given the complexity of ASD. As such, cutting-edge supervised methods such as NN demonstrated similar performance to primitive ones such as Naïve Bayes and SVM. Studies employing semi-supervised approaches such as granular NN [69] or stacked deep auto-encoders [29] reported sensitivities and specificities over 95%, but had small sample sizes. Thus, there is a demand for research into semi-supervised ML algorithms for ASD classification.

Atlas selection affects how the boundaries will be labelled and pre-defines the number of features that will be provided to the ML algorithm. Surprisingly, we observed a better specificity in studies that did not report using a brain atlas. However, a recent review [79] highlighted the limitations of the existing brain atlases and the consequences of their inappropriate use. First of all, most atlases stem from small cohorts which are unlikely to be representative of the general population. For example, many atlases are based on the Talairach and Tournoux atlas, derived from a single cadaveric brain of an elderly Caucasian female. In addition, most atlases lack the

appropriate clinical data measurements to support their claim of representing a normal brain. There is a lack of an age-specific atlas for infants, children, or adolescents, and such use of an age-inappropriate atlas can be problematic. Using the ICBM-DTI-81 atlas, which is based on 18-59 years old individuals on data acquired on 8–12-year-old children, a study [47] obtained a poor specificity of less than 25%. In addition, if the selected features cover a wide area, there is a higher chance that any subtle differences present in individuals with ASD might be overseen. Thus, we recommend that future studies be cautious on its selection for an appropriate brain atlas.

Cross-validation is a technique commonly used to evaluate the performance of a predictive ML algorithm. In meta-regression, the choice of cross-validation method did not affect the sensitivity or specificity of the classification. Given the complexity of ASD, data subsamples may not always be representative or have the same probability density function as the general dataset. Thus, cross-validation may not always guarantee good performance.

As the definitive diagnosis based solely on clinical assessment is difficult, there is a high demand for ML research with neuroimaging data as a diagnostic tool. Our results suggest that using semi-supervised ML algorithms based on combined anatomical with functional neuroimaging data could provide a strong foundation for future research. However, several questions need to be investigated before ML and neuroimaging can be used in clinical settings. A potential future direction would be to explore how the neural biomarkers specifically contribute to ASD-related traits. One way of doing this would be to see whether the identified biomarkers effectively distinguish individuals with ASD and other developmental disorders. Additionally, as ASD covers a broad spectrum with varying symptom manifestation, we urge future studies to investigate how symptom severity influences the results. Another interesting result is that our meta-regression did not find any significant difference in sensitivity or specificity between studies which included toddlers, children and adolescents, adults, or participants of all ages. Such results suggest a similar classification performance regardless of age. However, as the median age for ASD diagnosis is four years old, we urge future studies using ML with neuroimaging data to replicate the findings in the relevant pediatric population to ensure a smoother transition to clinical practice.

The use of ML with neuroimaging data to classify neuro-psychiatric disorders is gaining popularity. Similar classification performances (i.e., sensitivity >80% and specificity >80%) were obtained in recent meta-analyses for Alzheimer's disease [80] and schizophrenia [81], but lower

values were observed for bipolar disorder [82]. While for schizophrenia, functional imaging yielded better sensitivities, this was not the case for ASD as we found that combined modalities are superior. Conversely, although combining multiple ML algorithms improved the classification accuracy in Alzheimer's disease, this was not observed in the three studies which attempted to do so for ASD. Theoretically, a hybrid system combining deep learning approaches such as DNN for feature extraction and a traditional ML algorithm such as SVM for classification can yield better results. Traditional ML algorithms are unable to perform feature extraction as they have been designed for classification purposes based on well-defined optimal features. On the other hand, deep learning can automatically identify optimal features from the data, but to achieve a high classification performance it requires massive amounts of neuroimaging data which may not always be clinically feasible. In addition, the extracted data may be better suited for an algorithm which is specialized to perform classifications.

The high degree of heterogeneity (I^2 >95%) is concerning and probably reflects the lack of protocol standardization. We observed high inter-study variability in the choice of imaging modality, the protocol for the imaging acquisition, data pre-processing, the ML algorithm, the post-processing protocol, and the cross-validation method. Reporting ML protocols in a transparent and reproducible way, while difficult to standardize, will help promote higher generalizability. In addition, image protocol standardization is feasible and should be explored before larger-scale studies in clinical or research settings.

This systematic review includes a high number of high-impact publications which have demonstrated the potential of using ML with neuroimaging data to predict ASD status. However, many manuscripts have shown either a lack of transparency, replicability, ethics and/or effectiveness (TREE). These concerns have been extensively reported in the AI/ML literature [83] and addressing TREE issues is of vital importance to prevent wasting valuable research [84]. Transparent, responsible, and ethical use of AI is even more important in ASD research as this area involves individuals with neurodevelopmental disorders. Firstly, to avoid deceptive claims, researchers should ensure that the data collected is representative of the population concerned and that the ML/AI algorithm is always compared to the current gold standard [85]. Secondly, to ensure replicability, the AI/ML methodology should be made available to other researchers to promote external validation of the findings. Thirdly, to benefit the general population, the AI

pipeline should not widen pre-existing healthcare and social inequalities but should be cost-effectively implemented with easily interpretable outputs in clinical settings [86].

Limitations of this systematic review with meta-analysis include the high heterogeneity between studies and the possibility of publication bias for some analyses. However, our meta-regression highlighted a few sources of heterogeneity: sex of the recruited participants, pre-processing software, imaging modality and cross-validation method. Another limitation was the exclusion of studies that lacked sufficient quantitative data. For studies that possibly reported in a biased, we only included important results. Lastly, all studies excluded participants if head motion impacted the image quality, this might have impacted the results as it can introduce selection bias by limiting individuals with greater symptoms (the latter causing excessive repetitive movements); though we should acknowledge that severe motion artefacts cannot be reliably retrospectively corrected. Although our findings suggest that combining anatomical and functional neuroimaging data yields better results, we recognize that it can be challenging to perform multiple imaging modalities in young children. Lastly, the ML literature suffers from data leakage, which has not been considered within this review.

CONCLUSION

ML using neuroimaging data is a promising prospect in ASD classification. Given the high sensitivity and specificity achieved in the studies, our results provide insight as a potential robust approach to aid in clinical settings, but further studies are required to test its reliability in the relevant age groups. Findings need to be interpreted with caution as there is high degree of heterogeneity between studies, which stems mostly from differences in study design, image acquisition and ML protocol. Image protocol standardization and transparent ML protocol reporting will be crucial in generalizing results and establishing a foundation for application in real-world practice.

ACKNOWLEDGMENTS

This study is funded by the National Institute for Health Research UCLH Biomedical Research Centre (BRC399/NS/RB/101410).

REFERENCES

[1] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: Diagnositc and Statistical Manual of Mental Disorders, Fifth Edition. Arlington, VA: American Psychiatric Association, 2013.

[2] Craig JN, Laura Kresch C. Autism: An Emerging Public Health Problem. Public health reports (1974). 2003;118:393-9.

[3] Ozonoff S, Iosif A-M, Baguio F, Cook IC, Hill MM, Hutman T, Rogers SJ, Rozga A, Sangha S, Sigman M, Steinfeld MB, Young GS. A Prospective Study of the Emergence of Early Behavioral Signs of Autism. Journal of the American Academy of Child and Adolescent Psychiatry. 2010;49:256-66.e2.

[4] Zwaigenbaum L, Bryson S, Rogers T, Roberts W, Brian J, Szatmari P. Behavioral manifestations of autism in the first year of life. International journal of developmental neuroscience. 2005;23:143-52.

[5] Mandell DS. Factors Associated With Age of Diagnosis Among Children With Autism Spectrum Disorders. Pediatrics (Evanston). 2005;116:1480-6.

[6] Stahmer AC, Collings NM, Palinkas LA. Early Intervention Practices for Children With Autism: Descriptions From Community Providers. Focus on autism and other developmental disabilities. 2016;20:66-79.

[7] Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. Eur Child Adolesc Psychiatry. 2013;22:329-40.

[8] Randall M, Albein-Urios N, Brignell A, Gulenc A, Hennel S, Coates C, Symeonides C, Hiscock H, Marraffa C, Silove N, Bayl V, Woolfenden S, Williams K. Diagnosing autism: Australian paediatric research network surveys: Diagnosing autism: an APRN survey. Journal of paediatrics and child health. 2016;52:11-7.

[9] Taylor LJ, Eapen V, Maybery M, Midford S, Paynter J, Quarmby L, Smith T, Williams K, Whitehouse AJO. Brief Report: An Exploratory Study of the Diagnostic Reliability for Autism Spectrum Disorder. J Autism Dev Disord. 2017;47:1551-8.

[10] Dosreis S, Weiner CL, Johnson L, Newschaffer CJ. Autism Spectrum Disorder Screening and Management Practices Among General Pediatric Providers. Journal of developmental and behavioral pediatrics. 2006;27:S88-S94.

[11] Antezana L, Scarpa A, Valdespino A, Albright J, Richey JA. Rural Trends in Diagnosis and Services for Autism Spectrum Disorder. Frontiers in psychology. 2017;8:590-.

[12] Geschwind DH. Advances in Autism. Annual review of medicine. 2009;60:367-80.

[13] Di X, Azeez A, Li X, Haque E, Biswal BB. Disrupted focal white matter integrity in autism spectrum disorder: A voxel-based meta-analysis of diffusion tensor imaging studies. Progress in neuro-psychopharmacology & biological psychiatry. 2018;82:242-8.

[14] Tyszka JM, Kennedy DP, Paul LK, Adolphs R. Largely Typical Patterns of Resting-State Functional Connectivity in High-Functioning Adults with Autism. Cerebral cortex (New York, NY 1991). 2013;24:1894-905.

[15] Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Molecular psychiatry. 2012;17:1174-9.

[16] McAlonan GM. Mapping the brain in autism. A voxel-based MRI study of volumetric differences and intercorrelations in autism. Brain (London, England : 1878). 2004;128:268-76.

[17] Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PloS one. 2019;14:e0224365-e.

[18] Lanka P, Rangaprakash D, Dretsch MN, Katz JS, Denney TS, Deshpande G. Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets. Brain imaging and behavior. 2019.

[19] Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial Intelligence Transforms the Future of Health Care. The American journal of medicine. 2019;132:795-801.

[20] Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science (American Association for the Advancement of Science). 2015;349:255-60.

[20] Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science (American Association for the Advancement of Science). 2015;349:255-60.

[21] Moon SJ, Hwang J, Kana R, Torous J, Kim JW. Accuracy of Machine Learning Algorithms for the Diagnosis of Autism Spectrum Disorder: Systematic Review and Meta-Analysis of Brain Magnetic Resonance Imaging Studies. JMIR mental health. 2019;6:e14108.

[22] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of Clinical Epidemiology. 2009;62:e1-34.

[23] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155:529.

[24] Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. Statistical Methods of Meta-Analysis. Korean journal of radiology. 2015;16:1188-96.

[25] Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of clinical epidemiology. 2005;58:982-90.

[26] Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549-d.

[27] Emerson RW, Adams C, Nishino T, Hazlett HC, Wolff JJ, Zwaigenbaum L, Constantino JN, Shen MD, Swanson MR, Elison JT, Kandala S, Estes AM, Botteron KN, Collins L, Dager SR, Evans AC, Gerig G, Gu H, McKinstry RC, Paterson S, Schultz RT, Styner M, Schlaggar BL, Pruett JR, Piven J. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. Science translational medicine. 2017;9:eaag2882.\

[28] Shen MD, Nordahl CW, Li DD, Lee A, Angkustsiri K, Emerson RW, Rogers SJ, Ozonoff S, Amaral DG. Extra-axial cerebrospinal fluid in high-risk and normal-risk children with autism aged 2–4 years: a case-control study. The Lancet Psychiatry. 2018;5:895-904.

[29] Xiao X, Fang H, Wu J, Xiao C, Xiao T, Qian L, Liang F, Xiao Z, Chu KK, Ke X. Diagnostic model generated by MRI-derived brain features in toddlers with autism spectrum disorder. Autism research. 2017;10:620-30.

[30] Abraham A, Milham MP, Di Martino A, Craddock RC, Samaras D, Thirion B, Varoquaux G. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. NeuroImage (Orlando, Fla). 2017;147:736-45

[31] Akhavan Aghdam M, Sharifi A, Pedram MM. Combination of rs-fMRI and sMRI Data to Discriminate Autism Spectrum Disorders in Young Children Using Deep Belief Network. Journal of digital imaging. 2018;31:895-903.

[32] Bajestani GS, Behrooz M, Khani AG, Nouri-Baygi M, Mollaei A. Diagnosis of autism spectrum disorder based on complex network features. Computer methods and programs in biomedicine. 2019;177:277-83.

[33] Bosl WJ, Loddenkemper T, Nelson CA. Nonlinear EEG biomarker profiles for autism and absence epilepsy. Neuropsychiatric electrophysiology. 2017;3.

[34] Brahim A, Farrugia N. Graph Fourier transform of fMRI temporal signals based on an averaged structural connectome for the classification of neuroimaging. Artificial intelligence in medicine. 2020;106:101870.

[35] Chen T, Chen Y, Yuan M, Gerstein M, Li T, Liang H, Froehlich T, Lu L. The Development of a Practical Artificial Intelligence Tool for Diagnosing and Evaluating Autism Spectrum Disorder: Multicenter Study. JMIR medical informatics. 2020;8:e15767.

[36] Dekhil O, Hajjdiab H, Shalaby A, Ali MT, Ayinde B, Switala A, Elshamekh A, Ghazal M, Keynton R, Barnes G, El-Baz A. Using resting state functional MRI to build a personalized autism diagnosis system. PloS one. 2018;13:e0206351.

[37] Duchesnay E, Cachia A, Boddaert N, Chabane N, Mangin J-F, Martinot J-L, Brunelle F, Zilbovicius M. Feature selection and classification of imbalanced datasets: application to PET images of children with autistic spectrum disorders. NeuroImage (Orlando, Fla). 2011;57:1003.

[38] Eill A, Jahedi A, Gao Y, Kohli JS, Fong CH, Solders S, Carper RA, Valafar F, Bailey BA, Müller R-A. Functional Connectivities Are More Informative Than Anatomical Variables in Diagnostic Classification of Autism. Brain connectivity. 2019;9:64-612.

[39] Gori I, Giuliano A, Muratori F, Saviozzi I, Oliva P, Tancredi R, Cosenza A, Tosetti M, Calderoni S, Retico A. Gray Matter Alterations in Young Children with Autism Spectrum Disorders: Comparing Morphometry at the Voxel and Regional Level: Gray Matter Alterations in ASD Children. Journal of neuroimaging. 2015;25:866-74.

[40] Grossi E, Olivieri C, Buscema M. Diagnosis of autism through EEG processed by advanced computational algorithms: A pilot study. Computer methods and programs in biomedicine. 2017;142:73-9.

[41] Grossi E, Buscema M, Della Torre F, Swatzyna RJ. The “MS-ROM/IFAST” Model, a Novel Parallel Nonlinear EEG Analysis Technique, Distinguishes ASD Subjects From Children Affected With Other Neuropsychiatric Disorders With High Degree of Accuracy. Clinical EEG and neuroscience. 2019;50:319-31.

[42] Iidaka T. Resting state functional magnetic resonance imaging and neural network classified autism and control. Cortex. 2015;63:55-67.

[43] Ingalhalikar M, Parker D, Bloy L, Roberts TPL, Verma R. Diffusion based abnormality markers of pathology: Toward learned diagnostic prediction of ASD. NeuroImage (Orlando, Fla). 2011;57:918-27.
[44] Irimia A, Lei X, Torgerson CM, Jacokes ZJ, Abe S, Van Horn JD. Support Vector Machines, Multidimensional Scaling and Magnetic Resonance Imaging Reveal Structural Brain Abnormalities Associated With the Interaction Between Autism Spectrum Disorder and Sex. Frontiers in computational neuroscience. 2018;12:93.
[45] Jiao Y, Chen R, Ke X, Chu K, Lu Z, Herskovits EH. Predictive models of autism spectrum disorder based on brain regional cortical thickness. NeuroImage (Orlando, Fla). 2010;50:589-99.
[46] Kam T-E, Suk H-I, Lee S-W. Multiple functional networks modeling for autism spectrum disorder diagnosis: Multiple Functional Networks Modeling for ASD Diagnosis. Human brain mapping. 2017;38:5804-21.
[47] Payabvash S, Palacios EM, Owen JP, Wang MB, Tavassoli T, Gerdes M, Brandes-Aitken A, Cuneo D, Marco EJ, Mukherjee P. White Matter Connectome Edge Density in Children with Autism Spectrum Disorders: Potential Imaging Biomarkers Using Machine-Learning Models. Brain connectivity. 2019;9:29-220.
[48] Pham T-H, Vicnesh J, Wei JKE, Oh SL, Arunkumar N, Abdulhay EW, Ciaccio EJ, Acharya UR. Autism Spectrum Disorder Diagnostic System Using HOS Bispectrum with EEG Signals. International journal of environmental research and public health. 2020;17:971.
[49] Schirmer MD, Venkataraman A, Rekik I, Kim M, Mostofsky SH, Nebel MB, Rosch K, Seymour K, Crocetti D, Irzan H, Hütel M, Ourselin S, Marlow N, Melbourne A, Levchenko E, Zhou S, Kunda M, Lu H, Dvornek NC, Zhuang J, Pinto G, Samal S, Zhang J, Bernal-Rusiel JL, Pienaar R, Chung AW. Neuropsychiatric disease classification using functional connectomics - results of the connectomics in neuroimaging transfer learning challenge. Medical image analysis. 2021;70:101972-.
[50] Spera G, Retico A, Bosco P, Ferrari E, Palumbo L, Oliva P, Muratori F, Calderoni S. Evaluation of Altered Functional Connections in Male Children With Autism Spectrum Disorders on Multiple-Site Data Optimized With Machine Learning. Frontiers in psychiatry. 2019;10.
[51] Xiao Z, Wu J, Wang C, Jia N, Yang X. Computer-aided diagnosis of school-aged children with ASD using full frequency bands and enhanced SAE: A multi-institution study. Experimental and therapeutic medicine. 2019;17:4055-63.

[52] Zhang F, Savadjiev P, Cai W, Song Y, Rathi Y, Tunç B, Parker D, Kapur T, Schultz RT, Makris N, Verma R, O'Donnell LJ. Whole brain white matter connectivity analysis using machine learning: An application to autism. NeuroImage (Orlando, Fla). 2018;172:826-37.
[53] Ecker C, Marquand A, Mourão-Miranda J, Johnston P, Daly EM, Brammer MJ, Maltezos S, Murphy CM, Robertson D, Williams SC, Murphy DGM. Describing the brain in autism in five dimensions--magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. The Journal of neuroscience. 2010;30:10612-23.
[54] Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, Brammer MJ, Murphy C, Murphy DG. Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. NeuroImage (Orlando, Fla). 2010;49:44-56.
[55] Yahata N, Morimoto J, Hashimoto R, Lisi G, Shibata K, Kawakubo Y, Kuwabara H, Kuroda M, Yamada T, Megumi F, Imamizu H, Náñez Sr JE, Takahashi H, Okamoto Y, Kasai K, Kato N, Sasaki Y, Watanabe T, Kawato M. A small number of abnormal brain connections predicts adult autism spectrum disorder. Nature communications. 2016;7:11254.
[56] Yassin W, Nakatani H, Zhu Y, Kojima M, Owada K, Kuwabara H, Gonoi W, Aoki Y, Takao H, Natsubori T, Iwashiro N, Kasai K, Kano Y, Abe O, Yamasue H, Koike S. Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis. Translational psychiatry. 2020;10:1-11.
[57] Chen CP, Keown CL, Jahedi A, Nair A, Pflieger ME, Bailey BA, Müller R-A. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. NeuroImage clinical. 2015;8:238-45.
[58] Deshpande G, Libero LE, Sreenivasan KR, Deshpande HD, Kana RK. Identification of neural connectivity signatures of autism using machine learning. Frontiers in human neuroscience. 2013;7:670.
[59] Eslami T, Mirjalili V, Fong A, Laird AR, Saeed F. ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data. Frontiers in neuroinformatics. 2019;13:70.
[60] Fu Y, Zhang J, Li Y, Shi J, Zou Y, Guo H, Li Y, Yao Z, Wang Y, Hu B. A novel pipeline leveraging surface-based features of small subcortical structures to classify individuals with autism

spectrum disorder. Progress in neuro-psychopharmacology & biological psychiatry. 2021;104:109989-.

[61] Ghiassian S, Greiner R, Jin P, Brown MRG. Using Functional or Structural Magnetic Resonance Images and Personal Characteristic Data to Identify ADHD and Autism. PloS one. 2016;11:e0166934.

[62] Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage clinical. 2018;17:16-23.

[63] Huang F, Tan E-L, Yang P, Huang S, Ou-Yang L, Cao J, Wang T, Lei B. Self-weighted adaptive structure learning for ASD diagnosis via multi-template multi-center representation. Medical image analysis. 2020;63:101662.

[64] Kassraian-Fard P, Matthis C, Balsters JH, Maathuis MH, Wenderoth N. Promises, Pitfalls, and Basic Guidelines for Applying Machine Learning Classifiers to Psychiatric Imaging Data, with Autism as an Example. Frontiers in psychiatry. 2016;7:177.

[65] Kazeminejad A, Sotero RC. Topological Properties of Resting-State fMRI Functional Networks Improve Machine Learning-Based Autism Classification. Frontiers in neuroscience. 2018;12:1018.

[66] Li H, Parikh NA, He L. A Novel Transfer Learning Approach to Enhance Deep Neural Network Classification of Brain Functional Connectomes. Frontiers in neuroscience. 2018;12:491.

[67] Plitt M, Barnes KA, Martin A. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. NeuroImage clinical. 2015;7:359-66.

[68] Rakić M, Cabezas M, Kushibar K, Oliver A, Lladó X. Improving the detection of autism spectrum disorder by combining structural and functional MRI information. NeuroImage clinical. 2020;25:102181-.

[69] Tomasiello S. A granular functional network classifier for brain diseases analysis. Computer methods in biomechanics and biomedical engineering. 2019;8:1-7.

[70] Zu C, Gao Y, Munsell B, Kim M, Peng Z, Cohen JR, Zhang D, Wu G. Identifying disease-related subnetwork connectome biomarkers by sparse hypergraph learning. Brain Imaging and Behavior. 2018;13:879-92.

[71] Fombonne E. Epidemiology of Pervasive Developmental Disorders. Pediatric research. 2009;65:591-8.

[72] Halladay AK, Bishop S, Constantino JN, Daniels AM, Koenig K, Palmer K, Messinger D, Pelphrey K, Sanders SJ, Singer AT, Taylor JL, Szatmari P. Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. Molecular autism. 2015;6:36.

[73] Carter AS, Black DO, Tewani S, Connolly CE, Kadlec MB, Tager-Flusberg H. Sex Differences in Toddlers with Autism Spectrum Disorders. J Autism Dev Disord. 2007;37:86-97.

[74] Ros-Demarize R, Bradley C, Kanne SM, Warren Z, Boan A, Lajonchere C, Park J, Carpenter LA. ASD symptoms in toddlers and preschoolers: An examination of sex differences. Autism research. 2019;13:157-66.

[75] Stigler KA, McDonald BC, Anand A, Saykin AJ, McDougle CJ. Structural and functional magnetic resonance imaging of autism spectrum disorders. Brain research. 2011;1380:146-61.

[76] Barttfeld P, Wicker B, Cukier S, Navarta S, Lew S, Sigman M. A big-world network in ASD: Dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections. Neuropsychologia. 2011;49:254-63.

[77] Koprowski R, Foster KR. Machine learning and medicine: book review and commentary. Biomedical engineering online. 2018;17:17-0.

[78] Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC medical informatics and decision making. 2019;19:281-16.

[79] Dickie DA, Shenkin SD, Anblagan D, Lee J, Blesa Cabez M, Rodriguez D, Boardman JP, Waldman A, Job DE, Wardlaw JM. Whole Brain Magnetic Resonance Image Atlases: A Systematic Review of Existing Atlases and Caveats for Use in Population Imaging. Frontiers in neuroinformatics. 2017;11:1.

[80] Jo T, Nho K, Saykin AJ. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. Frontiers in aging neuroscience. 2019;11:220-.

[81] Kambeitz J, Kambeitz-Ilankovic L, Leucht S, Wood S, Davatzikos C, Malchow B, Falkai P, Koutsouleris N. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of

multivariate pattern recognition studies. Neuropsychopharmacology (New York, NY). 2015;40:1742-51.
[82] Librenza-Garcia D, Kotzian BJ, Yang J, Mwangi B, Cao B, Pereira Lima LN, Bermudez MB, Boeira MV, Kapczinski F, Passos IC. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. Neuroscience and biobehavioral reviews. 2017;80:538-54.
[83] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC medicine. 2019;17:195-.
[84] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. The Lancet (British edition). 2014;383:267-76.
[85] Collins GS, Moons KGM. Comparing risk prediction models. BMJ : British Medical Journal. 2012;344:e3186-e.
[86] Vollmer S, Mateen BA, Bohner G, Kiraly FJ, Ghani R, Jonsson P et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;369:368.

Figure 1. **PRISMA flow chart of study selection process**

ASD, autism spectrum disorder; *PRISMA,* Preferred Reporting Items for Systematic Reviews and Meta-Analysis; ML, machine learning.

Figure 2. QUADAS-2 Questionnaire : Quality Assessment Results

According to QUADAS-2, risk of bias is assessed across four domains (patient selection, index test, reference standard, and flow and timing), while applicability concerns are assessed in only the first three domains.

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2.

Figure 3. Forest plot for the univariate random-effects model sensitivity meta-analysis of the studies using machine learning to distinguish Autistic Spectrum Disorder (ASD) from Typically Developing (TP) individuals

Figure 4. Forest plot for the univariate random-effects model specificity meta-analysis of the studies using machine learning to distinguish Autistic Spectrum Disorder (ASD) from Typically Developing (TD) individuals

Figure 5 Hierarchical summary receiver operating characteristic (HSROC) curve with its 95% confidence region of the diagnostic performance of machine learning in differentiating Autistic Spectrum Disorder from Typical development

Youden's J Index represents the optimum cut-off between sensitivity and specificity.

ASD, autistic spectrum disorder; CI, confidence interval; HSROC, hierarchical summary receiver operating characteristic; ML, machine learning.

Table 1. Characteristics of the included participants

Study	N (ASD/TD)	Age	Sex	Diagnosis	Participant Source
Toddlers					
Emerson et al (2017) [27]	11 / 48	6m; 24m	Both	DSM-IV; ADOS; ADI-R	IBIS
Shen et al (2018) [28]	159 / 77	2-4 years	Both	ADOS; ADI-R	Recruited
Xiao et al (2017) [29]	46 / 39	18-37 months	Both	DSM-IV; ADOS; ADI-R	Recruited
Children and adolescents					
Abraham et al (2017) [30]	403 / 468	9-18 years	Both	ADOS; ADI-R	ABIDE
Akhavan et al (2018) [31]	116/69	5-10 years	Both	ADOS; ADI-R	ABIDE
Bajestani et al (2019) [32]	30 / 30	4-8 years	Both	Unclear	Unclear
Bosl et al (2017) [33]	44 / 47	3-12 years	Both	ADOS	Recruited
Brahim et al (2020) [34]	403 / 468	< 20 years	Both	ADOS; ADI-R	ABIDE
Chen et al (2020) [35]	119 / 131	6-18 years	Both	ADOS; ADI-R	ABIDE
Dekhil et al (2018) [36]	123 / 160	9-15 years	Both	ADOS; ADI-R	NDAR
Duchesnay et al (2011) [37]	45 / 13	5-15 years	Both	DSM-IV; ADI-R	Recruited
Eill et al (2019) [38]	46 /47	7-18 years	Both	ADOS; ADI-R	Recruited
Gori et al (2015) [39]	21 / 20	2-5 years	Males only	DSM-IV	Recruited
Grossi et al (2017)[40]	15 / 10	7-14 years	Both	DSM-V; ADOS	Recruited
Grossi et al (2019) [41]	20 / 20	4-14 years	Both	DSM-V	Unclear
Iidaka et al (2015) [42]	312 / 328	< 20 years	Both	DSM-IV; ADOS; ADI-R	ABIDE
Ingalhalikar et al (2011) [43]	45 / 30	7-13 years	Both	Unclear	Unclear
Irimia et al (2018) [44]	110 / 83	9-15 years	Both	DSM-V; ADOS; ADI-R	Recruited
Jiao et al (2010) [45]	22 / 16	6-15 years	Both	DSM-IV; ADI-R	Recruited
Kam et al (2017) [46]	119 / 144	< 20 years	Both	ADOS; ADI-R	ABIDE
Payabvash et al (2019) [47]	14 / 33	8-12 years	Males only	ADOS; ADI-R	Recruited
Pham et al (2020) [48]	40 / 37	4-13 years	Both	Unclear	Recruited
Schirmer et al (2021) [49]	25 / 25	8-12 years	Both	ADOS; ADI-R	Recruited
Spera et al (2019) [50]	102 / 88	6-13 years	Males only	ADOS	ABIDE
Xiao et al (2019) [51]	117 / 81	5-12 years	Both	ADOS; ADI-R	ABIDE
Zhang et al (2018) [52]	70 / 79	8-13 years	Males only	Unclear	Recruited

Adults					
Ecker et al (2010) (1) [53]	20 / 20	20-68 years	Males only	ICD-10; ADOS; ADI-R	Recruited
Ecker et al (2010) (2) [54]	22 / 22	18-42 years	Males only	ICD-10; ADOS; ADI-R	Recruited
Yahata et al (2016) [55]	74 / 107 *	18-42 years	Both	DSM-IV; ADOS; ADI-R	Recruited
Yassin et al (2020) [56]	45 / 125	20-60 years	Males only	DSM-IV; ADOS; ADI-R	Recruited
Lifetime					
Chen et al (2015) [57]	126 / 126	6-35 years	Both	ADOS; ADI-R	ABIDE
Desphande et al (2013) [58]	15 / 15	16-34 years	Both	ADOS; ADI-R	Recruited
Eslami et al (2019) [59]	505 / 530	9-34 years	Both	ADOS; ADI-R	ABIDE
Fu et al (2021) [60]	364 / 381	6-34 years	Males only	ADOS; ADI-R	ABIDE
Ghiassian et al (2016) [61]	538 / 573	8-25 years	Both	ADOS; ADI-R	ABIDE
Heinsfeld et al (2018) [62]	505 / 530	7-45 years	Both	ADOS	ABIDE
Huang et al (2020) [63]	159 / 197	8-21 years	Both	ADOS; ADI-R	ABIDE
Kassraian-Fard et al (2016) [64]	77 / 77	< 40 years	Males only	ADOS; ADI-R	ABIDE
Kazeminejad et al (2018) [65]	374 / 442 *	5-65 years	Both	ADOS; ADI-R	ABIDE
Li et al (2018) [66]	149 / 161 *	10-33 years	Both	ADOS; ADI-R	ABIDE
Plitt et al (2015) [67]	148 / 148	11-23 years	Males only	ADOS; ADI-R	ABIDE
Rakic et al (2020) [68]	368 / 449	5-64 years	Both	ADOS; ADI-R	ABIDE
Tomasiello et al (2019) [69]	78 / 104	9-20 years	Both	ADOS; ADI-R	ABIDE
Zu et al (2018) [70]	45 / 47	Unclear	Both	ADOS; ADI-R	ABIDE

ABIDE, Autism Brain Imaging Data Exchange; ADOS, Autism Diagnostic Observation Schedule; ADI-R, Autism Diagnostic Interview-Revised; ASD, autism spectrum disorder; DSM, Diagnostic and Statistical Manual of Mental Disorders; IBIS, Infant Brain Imaging Study; NDAR, National Database for Autism Research; TD, typically developing.

Table 2. Characteristics of the included studies

Study	Imaging Modality	Pre-processing Software	Atlas	Machine Learning Algorithm	Cross-validation Method
Toddlers					
Emerson et al (2017) [27]	rs-fMRI	N/A	Self-generated	SVM	Nested LOO
Shen et al (2018) [28]	aMRI	N/A	N/A	Balance-Boosted DT	OOB
Xiao et al (2017) [29]	aMRI	FreeSurfer	Desikan-Killiany	RF; Naive Bayes;SVM	5-fold; OOB
Children and adolescents					
Abraham et al (2017) [30]	rs-fMRI	PCP	MSDL; Yeo; HO; ICA; K-Means; Ward	SVM	10-fold; LSO
Akhavan et al (2018) [31]	aMRI; rs-fMRI	SPM	AAL	Neural network	10-fold
Bajestani et al (2019) [32]	EEG	N/A	N/A	kNN	LOO
Bosl et al (2017) [33]	EEG	NetStation	N/A	SVM	10-fold
Brahim et al (2020) [34]	rs-fMRI	PCP	Glasser	LR; L-SVM; RBF-SVM	Intra-Site
Chen et al (2020) [35]	rs-fMRI	SPM	AAL	Naïve Bayes	10-fold
Dekhil et al (2018) [36]	rs-fMRI	FSL	Parietal Cortex; TPJ; Neubert Ventral Frontal; Sallet Dorsal Frontal	SVM	2, 4, 10-fold; LSO
Duchesnay et al (2011) [37]	PET	N/A	N/A	SVM; Lasso LR; LDA	LOO; 10-fold; aPena
Eill et al (2019) [38]	aMRI; DTI; fcMRI	AFNI; FreeSurfer; FSL	Power; HO	CRF	OOB
Gori et al (2015) [39]	aMRI	SPM; FreeSurfer	Desikan-Killiany	SVM	LPO
Grossi et al (2017)[40]	EEG	N/A	N/A	LR; Naïve Bayes; k-CM; kNN; Sn; RF; SMO	LOO
Grossi et al (2019) [41]	EEG	SystemPlus Evolution	N/A	Back propagation; kNN; Sn; k-CM	N/A
Iidaka et al (2015) [42]	rs-fMRI	SPM; DPARSF	AAL	PNN	LOO; 2, 10, 50-fold

Ingalhalikar et al (2011) [43]	DTI	N/A	EVE	SVM	LOO
Irimia et al (2018) [44]	aMRI; DTI	LONI; TrackVis; FreeSurfer	Destrieux	SVM	10-fold
Jiao et al (2010) [45]	aMRI	FreeSurfer	Self-generated	SVM; MLP; FT; LMT	10-fold
Kam et al (2017) [46]	rs-fMRI	N/A	AAL	SVM; RFE-SVM; G-SVM; DRBM	10-fold
Payabvash et al (2019) [47]	DTI	FSL	ICBM-DTI-81	Naïve Bayes; RF; L-SVM; SVM-Polynomial Kernel; NN	1000-fold
Pham et al (2020) [48]	EEG	N/A	N/A	LDA; QDA; SVM; kNN;RBF-SVM; PNN	10-fold
Schirmer et al (2021) [49]	rs-fMRI	AFNI; SPM	AAL; CC200; HO	SVM	5-fold
Spera et al (2019) [50]	rs-fMRI	PCP	AAL; CC200; HO	SVM	LSO
Xiao et al (2019) [51]	rs-fMRI	FSL; MELODIC	N/A	Stacked Auto-Encoders	11, 33, 66, 99, 198-fold
Zhang et al (2018) [52]	DTI	Slicer	Self-generated	SVM	10-fold
Adults					
Ecker et al (2010) (1) [53]	aMRI	FreeSurfer	N/A	SVM	LTO
Ecker et al (2010) (2) [54]	aMRI	SPM	N/A	RFE-SVM	LOO
Yahata et al (2016) [55]	rs-fMRI	SPM	AAL; Brainvisa Sulci	L1-SCCA; SLR	LOO
Yassin et al (2020) [56]	aMRI	FreeSurfer	N/A	LR; SVM; DT	10-fold
Lifetime					
Chen et al (2015) [57]	rs-fMRI	FSL	Surface Based Atlas Power	PSO-SVM; RFE-SVM; RF	LOO; OOB
Desphande et al (2013) [58]	DTI; rs-fMRI	SPM	N/A	SVM	10-fold
Eslami et al (2019) [59]	rs-fMRI	PCP	CC200; AAL; Talariach and Tournoux	Autoencoder; Single Layer Perceptron; SVM; RF	10-fold
Fu et al (2021) [60]	aMRI	FSL	N/A	Boost learning DT	10-fold
Ghiassian et al (2016) [61]	aMRI; rs-fMRI	SPM	HO; Bangor Cerebellar	SVM	5-fold

Heinsfeld et al (2018) [62]	rs-fMRI	PCP	CC200	SVM; RF; DNN	10-fold
Huang et al (2020) [63]	rs-fMRI	DPARSF	AAL; CC200; Dosenbach 160	SVC; SASL; LR; Binary; kNN;	10-fold
Kassraian-Fard et al (2016) [64]	rs-fMRI	SPM	CC200	LR; Lasso LR; SVM; PNN; LDA; Naïve Bayes	Nested 10-fold
Kazeminejad et al (2018) [65]	rs-fMRI	AFNI; ANTS; FSL	AAL	SVM	10-fold
Li et al (2018) [66]	rs-fMRI	N/A	AAL	SVM; DNN; DTL-NN	5-fold
Plitt et al (2015) [67]	rs-fMRI	AFNI	DiMartino; Power; Destrieux	RF; kNN; L-SVM; RBF-SVM; Naïve Bayes; LDA; LR; L1-LR; L2L-R; EN-LR	LOO; Stratified 10-fold; Stratified 3-fold
Rakic et al (2020) [68]	aMRI; rs-fMRI	N/A	AAL; CC200; Destrieux	Stacked Auto-Encoders; Multi-layer Perceptron	10-fold
Tomasiello et al (2019) [69]	rs-fMRI	N/A	N/A	Granular Functional Network	5-fold
Zu et al (2018) [70]	rs-fMRI	N/A	AAL	SVM; STM	10-fold

3D, three dimensional; AAL, automated anatomical labelling; AFNI: Analysis of Functional Neuroimages; aMRI, anatomical magnetic resonance imaging; CC, Craddock; CRF, conditional random field; DNN, deep neural network; DPARSF: Data Processing Assistant for Resting-State fMRI; DRBM, discriminative restricted Boltzmann machine; DT, decision network; DTI, diffusion tensor imaging; DTL-NN, deep transfer learning neural network; EEG, electroencephalography; ENLR, elastic-net-regularized logistic regression; FSL, FMRIB software library; FT, functional trees; G-SVM, graph theory-based SVM; HO, Harvard-Oxford; ICA, independent component analysis; ICBM, International Consortium of Brain Mapping; k-CM, k-contractive map; kNN, k-nearest neighbors; L1-LR: L1-regularized LR; L1-SCCA, L1-norm regularized sparse canonical correlation analysis; L2-LR: L2-regularized LR; LDA, latent discriminant analysis; LMT, logistic model tree; LOO, leave one out; LPO, leave one participant out; LR, logistic regression; LSO, leave site out; L-SVM, linear SVM; LTO, leave two out; MSDL, multi-subject dictionary learning; MLP, multilayer perceptron; MRI, magnetic resonance imaging; OOB, out-of-bag; PET, positron emission tomography; PNN, probabilistic neural network; PSO SVM, particle swarm optimization; QDA, quadratic discriminant analysis; SVM; RBF-SVM: radial basis functional kernel SVM; RF, random

forest; rs-fMRI, resting state functional MRI; SASL, salient-adaptive sparsity learning; SLR, sparse LR; SMO, sequential minimal optimization; Sn, sine net neural networks; SPM, statistical parametric mapping; SVC, support vector classified; SVM, support vector machine; STM, support tensor machine; TPJ, tempo-parietal junction.

Table 3: Diagnostic performance of machine learning in differentiating Autistic Spectrum Disorder from Typical development: Meta-analysis results in the best set

Number of studies	Number of Participants		Parameter	Heterogeneity		Effect size		Egger's test
	ASD	TD		I^2	p-value	Pooled Estimate (95% CI)	AUC	p-value
44	5697	6013	Sensitivity	95.90 %	**<0.0001**	86.25 (81.24, 90.08)	0.889	**<0.0001**
			Specificity	95.90%	**<0.0001**	83.31 (78.12, 87.48)		
			Diagnostic Odds Ratio	47.01%	**0.001**	19.66 (13.40, 28.84)		
			Positive Likelihood Ratio	42.20%	**0.002**	3.76 (3.15, 4.48)		
			Negative Likelihood Ratio	43.83%	**0.001**	0.24 (0.20, 0.29)		
23*	5066	5416	Sensitivity	96.7%	**<0.0001**	83.23 (76.79, 88.16)	0.871	**<0.0001**
			Specificity	97.7%	**<0.0001**	78.90 (70.85, 85.19)		
			Diagnostic Odds Ratio	56.87%	**<0.0001**	14.66 (9.18, 3.15)		
			Positive Likelihood Ratio	55.38%	**0.001**	3.29 (2.67, 4.06)		
			Negative Likelihood Ratio	53.33%	**0.001**	0.25 (0.20, 0.32)		

* Represents the number of studies having at least 100 participants.
Significant p-values are highlighted in bold.

AUC, under the hierarchical summary receiver operating characteristic curve; CI, confidence interval.

Other abbreviations as in **Table 1**.

Table 4. Meta-regression results of the best set analysis.

Category	Sub-category	Number of studies	Pooled Sensitivity (95% CI)	Pooled Specificity (95% CI)	Bivariate Model p-value for Sensitivity	Bivariate Model p-value for FPR
Age-group	Toddler	3	83.33 [77.76; 87.73]	91.25 [46.18; 99.22]	ref	ref
	Children and Adolescents	23	89.61 [81.43; 94.43]	83.42 [74.99; 89.41]	0.680	0.975
	Adults	4	77.85 [67.30; 85.72]	86.70 [81.30; 90.72]	0.718	0.482
	Lifetime	13	83.33 [77.76; 87.73]	80.24 [70.60; 87.29]	0.941	1.000
Sex	Only males	10	74.33 [65.48; 81.55]	80.31 [73.58; 85.66]	ref	ref
	Both males and females	34	89.04 [83.84; 92.72]	84.18 [77.68; 89.05]	**0.021**	0.707
Diagnosis	ADOS/ADI-R	27	84.02 [76.33; 89.55]	80.18 [72.34; 86.23]	ref	ref
	DSM	2	94.28 [47.26; 99.67]	94.08 [17.11; 99.92]	0.485	0.798
	DSM AND ADOS/ADI-R	9	89.83 [82.00; 94.48]	88.25 [85.65; 90.44]	0.308	0.112
	ICD and ADOS/ADI-R	2	83.33 [68.95; 91.84]	88.10 [74.41; 94.96]	0.988	0.409
Participant origin	Recruited	18	90.48 [80.85; 95.53]	85.14 [78.86; 89.79]	ref	ref
	Database	23	83.18 [76.85; 88.05]	80.98 [72.04; 87.56]	0.485	0.289
Imaging modality	aMRI	8	80.83 [77.74; 83.58]	80.53 [73.71; 85.92]	ref	ref
	rs-fMRI	22	82.24 [74.95; 87.75]	81.83 [73.02; 88.23]	0.944	0.895
	Multiple	5	94.12 [85.43; 97.76]	81.30 [58.97; 92.94]	**0.036**	0.957

	EEG	5	99.85 [37.18; 1.00]	93.57 [84.09; 97.56]	**0.025**	0.105
	DTI	3	63.73 [29.89; 87.87]	85.98 [67.92; 94.67]	0.264	0.897
	PET	1	91.11 [78.59; 96.62]	76.92 [47.85; 92.37]	0.456	0.809
Atlas	AAL	7	83.09 [72.11; 90.33]	73.75 [57.24; 85.50]	ref	ref
	CC200	3	70.26 [65.89; 74.29]	66.79 [60.90; 72.21]	0.206	0.584
	Desikan-Killiany	2	80.60 [69.39; 88.39]	76.27 [63.83; 85.41]	0.798	0.945
	Destrieux	2	90.53 [60.38; 98.36]	84.96 [71.22; 92.80]	0.492	0.274
	EVE	1	73.33 [58.68; 84.19]	83.33 [65.68; 92.89]	0.563	0.593
	Glasser	1	53.35 [48.46; 58.17]	69.44 [65.12; 73.45]	0.114	0.851
	ICBM-DTI-81	1	21.43 [7.07; 49.43]	96.97 [81.39; 99.57]	**0.011**	0.090
	MSDL	1	90.82 [87.58; 93.28]	61.32 [56.83; 65.63]	0.401	0.561
	Multiple	9	83.70 [74.71; 89.93]	79.79 [69.57; 87.20]	0.930	0.466
	None	9	91.08 [81.22; 96.02]	90.91 [80.67; 96.00]	0.280	**0.032**
	Self-generated	3	86.41 [78.35; 91.78]	91.22 [48.56; 99.13]	0.650	0.556
ML algorithm	Linear Discriminant Analysis	1	91.11 [78.59; 96.62]	76.92 [47.85; 92.37]	ref	ref
	Support Vector Machine	18	86.04 [78.27; 91.34]	81.57 [73.16; 87.78]	0.636	0.847
	Neural Network	11	90.64 [78.37; 96.28]	88.67 [71.58; 96.05]	0.826	0.633
	Tree Based Algorithms	8	87.48 [69.86; 95.47]	85.80 [77.74; 91.27]	0.625	0.665
	Logistic Regression	2	72.02 [65.28; 77.90]	81.24 [74.46; 86.54]	0.311	0.767
	Naive Baayes	1	73.11 [64.45; 80.30]	58.02 [49.41; 66.16]	0.417	0.579

	Multiple	3	80.95 [68.14; 89.41]	78.04 [70.88; 83.85]	0.541	0.818
Validation method	Intra-site CV	1	53.35 [48.46; 58.17]	69.44 [65.12; 73.45]	ref	ref
	k-fold CV	26	86.77 [78.72; 92.09]	82.02 [73.83; 88.07]	0.105	0.565
	Leave p-out CV	11	83.05 [77.20; 87.64]	85.29 [78.45; 90.23]	0.148	0.399
	Out-of-bag error	4	87.88 [80.99; 92.51]	83.98 [71.00; 91.83]	0.077	0.431

Reference was chosen by the statistical program.

FPR, false positive rate; ref, reference; Other abbreviations as in **Table 1** and **Table 2**.